A Combined Strategy for Wind Speed Forecasting Using Data Preprocessing and Weight Coefficients Optimization Calculation

TONGJI GUO\textsuperscript{1}, LIFANG ZHANG\textsuperscript{2}, ZHENKUN LIU\textsuperscript{2}, AND JIANZHOU WANG\textsuperscript{2}

\textsuperscript{1}School of Statistics, Jiangxi University of Finance and Economics, Nanchang 330013, China
\textsuperscript{2}School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China

Corresponding author: Lifang Zhang (zhanglf_1106@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71671029, in part by the Major Program of National Social Science Foundation of China under Grant 17ZDA094.

ABSTRACT Wind speed forecasting is an essential procedure in electric grid dispatching. Short-term wind speed forecasting can be a great challenge and an intractable issue in increasing wind energy output and guaranteeing power safety. The current wind speed forecasting models are based on a single model, which is generally an artificial intelligence model or a statistical forecasting method. However, these models cannot perform well in all cases. An effective combined model is proposed in this paper, and this model includes four parts: weight coefficient optimization calculation based on the nonpositive constraint combination theory, singular-spectrum analysis, combined forecasting and discussion of results. The developed model can decrease the negative influences of the component models and maximize the advantages of each component model. To evaluate the forecasting accuracy of our proposed model, ten minutes of wind speed data from Shandong Peninsula, China, were used as test cases. It is clearly demonstrated that the developed combined strategy outperforms the individual forecasting methods in terms of forecast performance and stability.

INDEX TERMS Combined model, data preprocessing approach, nonpositive constraint combination theory, weight coefficient optimization.

I. INTRODUCTION

The background and motivation, literature review, contributions of our work and organization of this paper are detailed and introduced in this section.

A. BACKGROUND AND MOTIVATION

Due to the increasing amount of industrial and economic activity, people are becoming increasingly dependent on energy. However, traditional fossil fuels can no longer meet the demands of human pursuit of sustainable economic development; thus, people are changing their focus to renewable energy [1]. Wind power, with the advantages of being renewable and pollution-free, has been universally considered to be one of the most important renewable energy sources [2]. According to the statistics, until 2018, the global cumulatively installed wind capacity was approximately 591.55 GW; 209533 MW was contributed by China, whose proportion of cumulatively installed wind capacity increases annually [3].

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

However, as the most important factor of wind power, wind speed, is fluctuating and uncertain, wind power generation may be seriously affected by variable and uncontrollable wind speed conditions [4]. Thus, it is significant to establish a wind speed prediction system for capturing the random and erratic characteristic of wind speed as well as possible.

B. LITERATURE REVIEW

Recently, with the goal of efficient forecasting, many scientists have presented important forecasting strategies, which can be divided into five categories: i) physical methods, ii) statistical methods, iii) artificial intelligence methods, and iv) combined methods [5]. More specifically, the physical methods use physical considerations (e.g., temperature, topography, and roughness) to achieve wind speed forecasting [6]. For example, Di et al. [7] aimed to increase wind speed forecasting precision by optimizing the parameters of numerical weather prediction (NWP) based on a surrogate modeling-based optimization (ASMO) approach. To further emphasize the significance of parameter selection, elaborate
physical interpretations were given to illustrate the necessity of using the best parameters for the improvement of wind speed forecasting. Zhao et al. [8] realized that there were inevitable errors in numerical technologies, thus, for calculating and assessing the inherent errors, the residuals between forecasting results and realistic series were investigated, and a nonlinear and nonparametric method was presented to correct the raw forecasting results. The experimental results indicated that the error correction of numerical methods could greatly enhance prediction accuracy. However, compared to statistical methods and artificial neural networks (ANNs), physical strategies are not reliable for short-term prediction because they always need more computational time and too many physical considerations [9], [10]. The statistical forecasting methods use statistical equations to predict wind speed. Compared with other predicting methods, statistical strategies are easier to model, and they always exhibit satisfactory real-time performance [11]. In a retrospective of previous studies, Aasim et al. [12] proposed a novel repeated wavelet transform (WT) system based autoregressive integrated moving average model (ARIMA) aiming to enhance wind speed forecasting capacity in a very short period. By comparing it with other relative models at different time scales, the proposed model was shown to be the most effective for very short-term forecasting. Pearre and Swan [13] developed two forecasting methods; one method was based on a statistical correction, and the other method relied on interpolating correction topographies and instantaneous forecast errors. Before forecasting, classification in terms of site-specific corrections was conducted according to wind speed and direction, which effectively enhanced forecasting performance. With the improvement of forecasting technologies, artificial intelligence methods are being developed and shown to represent an improvement over statistical models with regard to wind speed forecasting [14]–[16]. Artificial intelligence strategies mainly include support vector machines (SVMs) [17], fuzzy logic methods [18], and ANNs [19]. Among these distinguished methods, ANNs have wider applications in wind speed forecasting, in which forward and recurrent networks are key components. In the first component, back-propagation neural network (BPNN) plays a dominant role due to its ability to approximate nonlinear functions. In recurrent networks, the Elman neural network (ENN) performs well in terms of sensitivity to the original series. This good performance is mainly due to its feedback connection, which always enhances its dynamic properties and nonlinear mapping capacities [20]. Support vector machines have also developed rapidly over the past few years; their forecasting capacity has been shown to be satisfactory. For instance, it was demonstrated that an SVM was superior to an ANN model and an adaptive neuro-fuzzy inference system model [21]. Although intelligent techniques are adept in capturing the nonlinear characteristics of time series, there are still intrinsic challenges, considering the randomness and instability of raw wind speed data [22]. Therefore, increasing numbers of studies have adopted data preprocessing technology to extract the main features of time series and filter noise to remarkably improve wind speed forecasting performance and accuracy.

Data preprocessing technology is adopted to filter out high frequency noise in the original data, which will greatly enhance the precision of the predicting method. Wavelet decomposition (WD) is a widely utilized technique for denoising nonlinear and unstable time sequence data. Liu et al. [23] put forward a method including WD, a genetic algorithm (GA) and an SVM. In this method, WD is used to decompose the wind speed data. As a frequently used wind data processing technique, empirical mode decomposition (EMD) uses the Hilbert-Huang transform to sift the data into a series of intrinsic mode functions (IMFs). Liu et al. [24] integrated EMD and an ANN to predict wind speed. It was verified that the effectiveness of the EMD-ANN method is superior to that of an individual ANN. Considering of the drawback of EMD, such as model mixing, EEMD was proposed. In addition, there are many improved methods, such as fast ensemble empirical mode decomposition (FEEMD) and wavelet packet decomposition (WPD). Singular spectrum analysis (SSA) is a new denoising technology; it is efficient in decomposing original data into independent and interpretable components, such as trend, cyclic, oscillatory and noise components [25]. SSA is currently widely used in different fields, particularly in the wind speed forecasting, pollution, financial, and economic fields. For instance, Gao et al. [26] developed a novel hybrid forecasting strategy integrating SSA, a firefly algorithm and an artificial neural network. The developed model provides satisfactory forecasting performance compared with a single artificial neural network. Liu et al. [27] applied SSA to decompose raw data into several components. With the trend component forecasted by convolutional neural network (CNN) and the detail components forecasted by SVR, the developed hybrid model was shown to be better than compared methods in enhancing wind speed prediction capacity.

The overall disadvantages of the mentioned approaches are summarized below:

1. Relative to other forecasting methods, a physical method is not the best choice for short-term wind speed prediction. Moreover, substantial physical information is needed, which increases the consumption of time and materials.
2. Wind speed series are nonlinear and unstable, indicating that a forecasting model should perform under the assumption of nonlinearity. However, the assumption of statistical modeling is that the original series is linear; thus, it is difficult to greatly improve forecasting accuracy.
3. The artificial intelligence method is an outstanding improvement over other compared models. However, there are still intrinsic drawbacks including overfitting and falling easily into a local optimum.
4. To filter out the noise in volatility sequences, many data preprocessing approaches are taken into account. However, we should not ignore the disadvantages of the previous...
methods, such as the model mixing in EMD and the residual noise in EEMD.

To find more high-accuracy prediction models, a great deal of studies have been devoted to combined forecasting methods that integrate the abovementioned individual forecasting models. A combined forecasting method was developed in 1969 by Liu et al. [28], which in general, sets the weights of each component model in accordance with its past prediction accuracy [29]. Combined methods can overcome the limitations of individual models and achieve satisfactory accuracy, and they are considered an advanced approach to perfect forecasting abilities. For weight coefficient optimization calculation in a combined model, recently, many studies have aimed to find a way to determine the weights of the component models to achieve high accuracy. For the traditional combination method, which operates on the basis of the minimum sum of the prediction error square, depending on whether the combination methods have different or similar forecasting accuracies, the individual model that provides the best performance is assigned a weight of 1, and the rest are assigned 0. Considering these drawbacks of the conventional method, a novel combination strategy using nonpositive constraints and artificial intelligence algorithms with a weight-determined combination method (AI-CM-NNTC) was developed by Xiao et al. [30]. The AI-CM-NNTC combined method was verified to be superior to previous combination strategies and component models in prediction performance. For example, Jiang and Liu [31] proposed a new combined system employing variable-weight combination theory for wind speed prediction. An ELM optimized by the multi-objective grasshopper optimization algorithm was applied to combine the predicted values of all hybrid models. The experimental results demonstrated the validity of the proposed combined strategy. In addition, other AI-CM-NNTC combined methods, including an ELM optimized by a multi-objective salp swarm algorithm (MSSA) [32] and a least-squares support vector machine (LSSVM) optimized by a multi-objective ant lion algorithm (MOALO) [33] have also been employed for wind speed prediction.

C. CONTRIBUTIONS

To achieve better forecasting results in our study, a combined model was developed, which used the cuckoo search (CS) algorithm to optimize the weight coefficients of each component model based on nonpositive constraint theory. For the reasons that an individual forecasting model cannot always supply stable and high accuracy in any situation and that artificial intelligence models always have high precision with regard to wind speed forecasting, the combined model integrates SVM, Elman, BP optimized by the firefly algorithm (FA-BP) and BP based on particle swarm optimization and the genetic algorithm (GA-PSO-BP) into the developed system. Meanwhile, SSA is adopted to preprocess the original data, and ultimately, we employ the CS to optimize the weight of each component model based on nonpositive constraints. The experiments demonstrate that the proposed combined model is always more effective for wind speed forecasting because it is stable and has high accuracy.

The primary contribution of our research for wind speed forecasting is illustrated as follows:

1. Two individual ANNs based on a data preprocessing strategy and two hybrid systems based on a data pretreatment strategy and different optimization algorithms are employed to establish the proposed combined model, which guarantees that the advantages of each component model are maximized. SSA-SVM, SSA-Elman, SSA-FA-BP, and SSA-GA-PSO-BP are simultaneously employed to represent the nonlinear characteristics of the original wind speed series and to improve forecasting performance.

2. Three different optimization algorithms are used in our study: one algorithm is used to provide the optimal weight for each forecasting component, and the other two algorithms are used to optimize the weights and thresholds of the benchmark neural network in hybrid component models. The cuckoo optimization algorithm can be adopted to determine the optimal weight due to its excellent optimization ability and simple operation process. GA-PSO is developed to find the global optimal solution and improve the optimization efficiency of the hybrid component model.

3. Considering that the parameters of the data preprocessing technology can influence forecasting performance to a great extent, a comparison experiment with the combined model and the component models is conducted to determine the appropriate range of the correlation coefficient and the number of components $r$. This approach of defining the $r$ value can be applied to the forecasting process well.

D. ORGANIZATION

The main content of this paper can be summarized as follows. Section II gives the definition of the SSA methodology. Section III introduces the component forecasting models (SVM, Elman, FA-BP and GA-PSO-BP). The combined model theory is proposed in Section IV. The simulation process and forecasting results are analyzed in Section V and Section VI, respectively. Finally, Section VII draws conclusions.

II. SINGULAR SPECTRUM ANALYSIS METHODOLOGY

SSA is helpful in decomposing raw time series into a few IMFs, which contain the “clean” sequences, oscillations and noise based on singular value decomposition [34], [35]. The specific process is demonstrated below.

A. FIRST STAGE: DECOMPOSITION

Step 1: Transform the original time series $X = (x_1, x_2, \ldots, x_N)$ of length $N$ into a trajectory matrix. Subsequently, the trajectory matrix can be defined by Eq. (1).

$$T = (X_1 X_2 \cdots X_K) = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_1 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix} (1)$$

.$$T = (X_1 X_2 \cdots X_K) = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_1 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix} (1)$$

.$$T = (X_1 X_2 \cdots X_K) = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_1 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix} (1)$$
The matrix $\mathbf{T}$ represents a Hankel matrix whose dimensionality is $L \times K$, $K = N - L + 1$, and all the $x_{ij}$ on the anti-diagonals where $i + j$ is a constant ($i = 1, 2, \ldots, L$, $j = 1, 2, \ldots, K$), are equal.

For example, if we assume that $N = 10$ and $L = 5$, then $K = 6$, and the trajectory of the matrix is shown in the following equation:

$$
\mathbf{T} = \begin{pmatrix}
6 & 12 & 18 & 24 & 30 & 36 \\
12 & 18 & 24 & 30 & 36 & 42 \\
18 & 24 & 30 & 36 & 42 & 48 \\
24 & 30 & 36 & 42 & 48 & 54 \\
30 & 36 & 42 & 48 & 54 & 60
\end{pmatrix}
$$

where the dimensionality of this Hankel matrix is $L \times K(5 \times 6)$, and $x_{ij} = x_{i-1,j+1}$. Specific, when $i = 2$ and $j = 3$, $x_{2,3} = x_{1,4}$.

Step 2: The eigenvalues of $\mathbf{T} \mathbf{T}^T$ are computed and sorted in decreasing order by SVD, that is, $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L \geq 0$. The homologous eigenvectors can be represented by $(U_1, \cdots, U_L)$. Subsequently, SVD can be written as Eq. (3),

$$
\mathbf{T} = \mathbf{T}_1 + \mathbf{T}_2 + \cdots + \mathbf{T}_d,
$$

where $T_i = \sqrt{\lambda_i} U_i V_i^T (i = 1, 2, \ldots, d), d$ is the rank of $\mathbf{T}$, and $V_i = \mathbf{T}_i U_i / \sqrt{\lambda_i}$ are the principal components.

### B. SECOND STAGE: RECONSTRUCTION

The reconstruction stage is subdivided into two parts: grouping and averaging.

1) GROUPING

This part divides $\{1, 2, \ldots, d\}$ into $r$ disjoint subsets $S_1, S_2, S_3, \ldots, S_r$; the resultant trajectory matrices can be computed for each subset $S_1, S_2, S_3, \ldots, S_r$, and Eq. (3) can be stated as $\mathbf{T} = \mathbf{T}_{S_1} + \mathbf{T}_{S_2} + \cdots + \mathbf{T}_{S_r}$.

2) AVERAGING

The last step is transforming the grouped matrices $T_S$ into a novel sequence whose length is $N$. Define $\mathbf{Y}_1 = (y_1, y_2, \cdots, y_N)$ as the converted one-dimensional sequence; the elements in $\mathbf{Y}_1$ are given in Eq. (4).

$$
y_k = \begin{cases}
1 & \text{for } 1 \leq k < L \\
\frac{1}{L} \sum_{j=1}^{L} y_{j,k-j+1} & \text{for } L \leq k \leq L^* \\
\frac{1}{L^*} \sum_{j=1}^{L^*} y_{j,k-j+1} & \text{for } L^* < k \leq K^* \\
\frac{1}{N-K^*+1} \sum_{j=k-K^*+1} \sum_{j=k-K^*+1} y_{j,k-j+1} & \text{for } K^* \leq k \leq N
\end{cases}
$$

where $L^* = \min(L, K)$, and $K^* = \max(L, K)$. If $L < K$, $y_0 = \min(L, K)$, and $K^* = \max(L, K)$. If $L \geq K$, $y_{0,k-j+1} = y_{j,k-j+1}$ if $L \geq K$, $y_{0,k-j+1} = y_{k-j+1,j}$. The operational process of SSA is summarized in Fig. 1, A.

### C. PARAMETER SELECTION

There are two significant parameters that play a vital role: the window length $L$ and the number of components $r$. The selection of these two parameters directly influences the decomposition of the time series and, in turn, influences the forecasting precision.

- The window length determines the dimension of Hankel matrix. The $L$ value is in the range $N/3 \leq L \leq N/2$.
- The number of components $r$ defines the number of decomposed series employed for reconstruction. The general criterion for $r$ value selection is based on the contribution of $T_i$ to the trajectory matrix of $\mathbf{T}$, and the sum of component contributions is at least a preset threshold, for instance, 90% above [36, 37]. In our experiments, we use historical data to determine $r$ and then apply them to the final wind speed forecasting; the detailed method will be discussed in Section V.

### III. WIND SPEED FORECASTING MODELS

There are numerous ANNs that have been adopted to wind speed prediction, including several classical forecasting models and several forecasting models based on stochastic heuristic optimization algorithms. In our paper, Elman, SVM, FA-BP, and GA-PSO-BP are chosen as component predicting strategies to build the developed combined model.

#### A. ELMAN NEURAL NETWORK (ENN)

The ENN is a partial recurrent neural network introduced by Elman in 1990 [38]. An ENN consists of four layers: input, hidden, output and feedback layers. The trait of time-delay memory increases the sensitivity of the ENN to historical data, a local feedback network internally enhances the capacity of the ENN to address dynamic information, and these together give ENNs an advantage over static neural networks in modeling hydraulic systems [39, 40].

#### B. SUPPORT VECTOR MODEL (SVM)

The SVM was first developed by Vapnik in 1979 [41] on the basis of structured risk minimization and statistical machine learning theory. For nonlinear regression of SVM, the fundamental ideas of changing the data to a high-dimensional feature space and transforming the nonlinear regression issue into linear regression issues are realized by a kernel function.

#### C. FORECASTING MODELS BASED ON A BPNN

BPNN is a multilayer mapping network with numerous applications. However, the shortcoming of BPNN is that the network cannot forecast precisely, due to the training result easily falling into a local optimum. For this reason, many stochastic heuristic optimization algorithms, such as CS, the FA and the PSO algorithm, have been used for optimizing BPNN. These optimization algorithms have the ability to overcome the shortcomings in BP of easily falling into local optimization, missing the globally optimal solution and slowing convergence. Li et al. optimized the initial weights...
and bias in BPNN [42] with an adaptive GA. In addition, some improved algorithms or hybrid algorithms have been proposed to address the weaknesses of the basic algorithm; for instance, [43] proposed improved cuckoo search (ICS) by changing the method of adjusting $p_a$ and $\alpha$. The experimental results verified the validity of the developed algorithm. Chen et al. [44] used a hybrid algorithm that combines PSO with an artificial fish swarm algorithm for neural network training. The efficiency of their proposed hybrid algorithm is quite good. In this paper, we used the Firefly algorithm and hybrid GA-PSO algorithm to optimize the initial parameters in BP.

**D. BACK PROPAGATION NEURAL NETWORK BASED ON THE FIREFLY ALGORITHM**

The firefly algorithm (FA) was initially proposed by Yang [45] for solving optimization problems; it originates from the flashing behavior of fireflies, in which the fireflies are attracted to each other by flashing lights. The FA can be used to address the most difficult optimization problems.

**E. BACK PROPAGATION NEURAL NETWORK OPTIMIZED BY A HYBRID GA-PSO ALGORITHM**

PSO was suggested by Kennedy and Eberhart in 1995 [46]. The flowchart of PSO is presented in Fig. 1, B.

However, the diversity of PSO is insufficient, so it is difficult to find a global optimum with a small population. Nevertheless, the selection, crossover and mutation in GA can compensate for this deficiency. The genetic algorithm, developed by Holland [47], is a heuristic and stochastic optimization algorithm using evolutionary and genetic theories. The genetic algorithm can quickly converge to achieve optimization by genetic operations. Fig. 1, C, shows the flowchart of GA. Our hybrid algorithm using GA-PSO is developed in this paper; the GA-PSO algorithm can use the advantages of the component algorithms to improve optimization efficiency.

The steps of the developed optimization algorithm can be summarized as:

Step 1: There are $N$ particles in each generation, and their fitness values are assessed and ranked. $N/2$ particles are generated by performing PSO steps to obtain the best particles, and the $N/2$ offspring are used in the next generation.
FIGURE 2. The structure of combined theory based on cuckoo search and non-positive constraint theory. (A. The procedure of determining the weight coefficient. B. The pseudocode of CS algorithm.)

Step 2: The remaining $N/2$ particles are excluded. In addition, in the GA process, $N/2$ individuals are created by crossover and mutation between the genetic variants and the surviving $N/2$ PSO members.

Step 3: A novel population is formed with $N$ particles whose fitness values will be assessed and ranked once more.

A rudimentary GA-PSO algorithm is outlined in Algorithm 1.

IV. THE COMBINED FORECASTING MODEL

From a conceptual perspective, weighting-based combined methods can be defined as combining several independent forecasting models and providing them with an appropriate value that represents the effectiveness of each component. The general prediction process is shown in Fig. 2, A. The process of these approaches can be summarized as follows. (a) Examine the measured and forecasting values. (b) Set weight coefficients for each individual component based on its effectiveness in the fitting process. (c) Finally, compute the weighted predicting consequences by:

$$x_i = x_i + \beta_0 e^{-\gamma r^2} (x_j - x_i) + \alpha \left( rand - \frac{1}{2} \right)$$ (5)

where $\hat{F}_{t+h}$ is the final predicted consequence for $h$ steps ahead at time $t$; $\hat{F}_{i,t+h | t}$ is the predicted consequence of the $i$th component at $h$ steps ahead at time $t$; $M$ is the number of individual strategies; and $w_i$ is the weight coefficient of the $i$th individual model, which satisfies $w_1 + w_2 + \cdots + w_M = 1$.

For weight coefficient calculation, the traditional combination method based on the minimum sum of the squared forecast error has been shown to be inefficient; therefore, this paper develops a combined model using non-positive constraint theory and CS to decide the weights of combined models.

A. COMBINED MODEL THEORY

Definition 1: The conventional combined approach assigns a weight coefficient to every combined model based on
Algorithm 1 GA-PSO

Parameters:
- \( \text{Gen}_{\text{max}} \) — maximum iteration times
- \( t \) — current iteration
- \( k \) — current iteration of PSO
- \( K_{\text{max}} \) — maximum iteration times for PSO
- \( N \) — population size
- \( F_i \) — fitness function
- \( x_p \) — the optimal fitness values and solution

1: /* Assign parameters for GA and PSO. */
2: /* Randomly create the population of \( N \). */
3: FOR EACH \( i: 1 \leq i \leq N \) DO
4: /* Assess the fitness function \( F_i \). */
5: /* Rank the population of \( N \). */
6: /* Recognize the optimal \( N/2 \) particles and get rid of the rest particles. */
7: END FOR
8: WHILE \( (g < \text{Gen}_{\text{max}}) \) DO
9: IF \( (k < K_{\text{max}}) \) DO
10: /* Get the novel velocities \( V(t+1) \) and positions \( X(t+1) \). */
11: \[
V_{ij}(t+1) = w V_{ij}(t) + C_1 r_1^i (P_{\text{best}} - x_{ij}(t)) + C_2 r_2^i (G_{\text{best}} - x_{ij}(t))
\]
12: \( x_{ij}(t+1) = x_{ij}(t) + V_{ij}(t+1) \)
13: /* Update \( P_{\text{best}} \) and \( G_{\text{best}} \). */
14: /* Calculate the corresponding fitness function \( F_i \). */
15: /* Let \( k = k + 1 \). */
16: IF /* Recognize the optimal \( N/2 \) particles and get rid of the rest particles. */
18: /* Select the best \( N/2 \) members from \( X(i+1) \) to \( X_k(i+1) \). */
19: /* Conduct crossover process to form \( X_c(i+1) \). */
20: /* Dynamic mutation to create \( N/2 \) members \( X_m(i+1) \). */
21: /* Form the new generation of \( N \) members from PSO and members from GA. */
22: /* Assess the fitness function \( F_i \). */
23: /* Rank the new generation of \( N \) members. */
24: /* Let \( t = t + 1 \). */
25: END WHILE
26: RETURN \( x_p \)

minimizing sum of the squared error (SSE):

\[
\min J = L^T E L = \sum_{t=1}^{T} \sum_{j=1}^{m} \sum_{i=1}^{m} \left( R^T L \right)_{i,j} = 1 \quad L \leq 0 \quad (6)
\]

where \( L = (l_1, l_2, \ldots, l_m)^T \) represents the weight vector, \( R = (1,1,\ldots,1)^T \) represents a unit vector and \( E_{ij} = e_i e_j^T \), where \( e_i = (e_{1i}, e_{2i}, \ldots, e_{Ni}) \). \( E = (E_{ij})_{m \times m} \) represents the error information matrix.

Definition 2: The combined approach, on the basis of the nonnegative constraint theory and the CS algorithm weight-determined combination method, is given as follows:

\[
\min J = L^T EL = \sum_{t=1}^{T} \sum_{j=1}^{m} \sum_{i=1}^{m} l_{ij} e_i e_j \quad \left\{ \begin{array}{l}
R^T L \approx 1 \\
-3 \leq L \leq 3 
\end{array} \right. \quad (7)
\]

The weight vector has a maximum value of 1 and a minimum of 0, and the sum of the weights is approximately 1. The value of the weight vector is in the range \([-3, 3]\).

B. WEIGHT COEFFICIENT OPTIMIZATION CALCULATION

Four classical wind speed forecasting models are involved in the combined model; hence, the detailed process of the weight coefficient optimization calculation is carried out as follows:

a) Collect the single model forecasts \( y_1, y_2, y_3 \) and \( y_4 \), which represent the forecasts of SSA-Elman, SSA-SVM, SSA-FA-BPNN and SSA-GAPSO-BPNN, respectively.

b) Design the parameters for the CS algorithm and randomly initialize the population of \( 4 \) nests \( x_i \) \( (i = 1, 2, \ldots, n) \).

The objective function is based on minimizing the SSE.

c) According to the Lévy flight behavior of the nest-seeking of cuckoo, let \( X_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})^T \) be the current solution for the \( i \)th cuckoo; then, the new solution \( x_i^{t+1} \) is generated as follows:

\[
x_i^{t+1} = x_i^t + \alpha \oplus L(\lambda), \quad i = 1, 2, \ldots, N \quad (8)
\]

d) Keep the best solution \( X = (x_1, x_2, x_3, x_4)^T \), which represents the weight coefficients. Thus, the combined forecasting result is:

\[
Y = y_1 x_1 + y_2 x_2 + y_3 x_3 + y_4 x_4 \quad (9)
\]

Fig.2 demonstrates the construction of the proposed combination strategy. The rudimentary CS algorithm is outlined as the pseudocode in Fig.2. B. Table 1 presents the parameters of the CS.

V. SIMULATION PROCESS AND ANALYSIS

For this study, to further investigate the stability and credibility of the developed strategy, wind speed data from three different wind turbines in Penglai, China, are used to demonstrate the performance of the developed model.

A. DATA COLLECTION

Penglai city lies in northeast China, in Shandong Province, and it is adjacent to the Yellow Sea and Bohai Sea. The location of the research sites is presented in Fig.3. A. The city enjoys a monsoon climate owing to distinct continental climate features; its annual average wind speed is 5.2 m/s and it is not affected by tropical storms, typhoons or other extreme weather.
TABLE 1. The experiment parameters of CS.

| Experimental parameters | Default value |
|-------------------------|---------------|
| Stopping criteria       |               |
| Maximum iteration       | 100           |
| Population size         | 20            |
| Maximum iteration times | 100           |
| $P_a$                   | 0.25          |
| $\alpha$                | 1             |

This table gives the significant parameters included in CS algorithm. The cuckoo search is a helpful tool to optimize the weights of combined strategy. Its advantage lies in its excellent optimization ability and easy operation process.

FIGURE 3. The geographical location, data characteristic, and original wind speed time series of Penglai area in China.

With the aim of fully verifying the proposed combined model in terms of stability and universality, in the selection of wind speed data, we choose data with different turbines from different months, which contributes to showing the combination’s applicability to different months and turbines.
TABLE 2. Assessment indicators for prediction effectiveness.

| Metric | Definition | Equation |
|--------|------------|----------|
| MAE    | The average absolute forecast error. | \[ MAE = \frac{1}{N} \sum_{n=1}^{N} |y_n - \hat{y}_n| \] |
| MSE    | The average of the forecasting error squares. | \[ MSE = \frac{1}{N} \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 \] |
| MAPE   | The average of absolute error. | \[ MAPE = \frac{1}{N} \sum_{n=1}^{N} \frac{|y_n - \hat{y}_n|}{y_n} \times 100\% \] |

This table presents the evaluation indicators in terms of forecasting accuracy in our paper. \( y_n \) represents the real value and \( \hat{y}_n \) is the forecasting value. It can be seen from the expression that the involved three criteria are all contrary Indicators, which means the smaller the MAE, MSE, MAPE values, the better the forecasting performance.

TABLE 3. The parameters of Elman.

| Experimental parameters       | Default value |
|--------------------------------|---------------|
| the input nodes number        | 8             |
| the hidden nodes number       | 17            |
| the output nodes number       | 1             |
| net.trainparam.epochs         | 1000          |
| net.trainparam.show           | 20            |

This table demonstrates the involved parameters in Elman neural network. With the characteristic of handling dynamic information, Elman is obviously over a static neural network in modeling hydraulic system. The objective function is \[ E(k) = \frac{1}{2} \sum (y(k) - y(k))^2 (y(k) - y(k)) \], where \( y(k) \) represents the output of \( k \)th step.

Fig.3, B shows a descriptive statistic of the study data according to four metrics. Cases 1-3 are the data from June 2011 of three wind turbines, Case 4 is the data from March 2011 of the second wind turbine and Case 5 is the data from October 2011 of the third wind turbine. From the four metrics and the line chart, we can clearly see that Case 4 and Case 5 appear to have different trends. Fig.3, C shows the original wind speed time series in five cases.

The total available wind speed data include 1748 samples, of which 1508 samples (approximately 251.3 hours) are training data, and 240 samples (40 hours) are test data. A rolling forecasting strategy is applied to predict the wind speed over the next 40 hours (240 test points). To forecast more accurately, 8 previous ten-minute data are employed to predict wind speed in the next step; that is, the input datasets are \( \{x(t - 8), x(t - 7), x(t - 6), x(t - 5), x(t - 4), x(t - 3), x(t - 2), x(t - 1)\} \) and the output datasets are \( \{x(t)\}, t = 9, 10, \ldots, 1748 \). \( x \) denotes the wind speed time series, and the real value is continuously employed to update the latest value. This principle applies to all models presented in this paper. The rolling process can be seen in Fig.4, part 3.

B. ASSESSMENT INDICATORS FOR PREDICTION EFFECTIVENESS

To assess the forecasting ability of our developed combined strategy, three performance evaluation indices must be introduced to comprehensively measure the proposed model. The three indices adopted in this study are the mean absolute error (MAE), mean square error (MSE) [48] and mean absolute percent error (MAPE) [49]. The detailed definitions of the indicators are given in Table 2.

C. EXPERIMENT PREPARATION

Before the experiment, the parameters in the model and optimization algorithms must be defined. For the Elman neural network, the numbers of input layer and output layer neurons are set to 8 and 1, respectively, and the number of hidden layer neurons is set to 9. The training function is trainlm, the hidden layer function is tansig, and the output layer uses purelin as the output function; the details are given in Table 3.

For the FA-BP model and GA-PSO-BP model, based on Kolmogorov’s theorem that \( 2i + 1 \) hidden neurons can map any function for \( i \) inputs, the number of nodes in the hidden layer is set to 17. Table 4 shows the parameters of FA.

The number of variables for the GA-PSO is eight. The details are given in Table 5.

D. DATA PREPROCESSING AND SIMULATION

To eliminate instability and to guarantee that the forecasting performance is reliable and does not depend on randomly initialized parameters, we repeated the experiments for each model ten times and then averaged the values for the three indices.
FIGURE 4. The Schematic of the developed combined strategy including data preprocessing process, rolling prediction process and combined process based on cuckoo search.
The experiments were conducted in MATLAB R2014b on Windows 7 with a 3.30 GHz Intel(R) Core (TM) i5-4590 64-bit CPU with 8 GB of RAM.

1) DATA PREPROCESSING IN OUR DEVELOPED MODEL
As the two most important parameters in the SSA algorithm, the values of $L$ and $r$ directly affect the accuracy of forecasting. The general rule for $L$ value selection is that it is in the range $N/3 \leq L \leq N/2$, where $N$ denotes the length of sequences, so in this paper, the window length $L$ is set to 700. Because the wind speed is elusive and changeable, we should not only keep the main characteristics and the integrity of information but also filter out noise in denoising. Therefore, in order to find an excellent $r$ value, we define $r$ by studying historical data to apply them to the final forecasting. The idea is to use BP to forecast the previous wind speed by adjusting the $r$ value in five cases and recording the corresponding MAPE values. Finally, we compare the MAPE values and determine the $r$ value. Using a BP neural network to find the $r$ value is done for two reasons: one reason is that in this paper, the combination of models has two component models based on a BP neural network; the second reason is that four of the component models are artificial intelligence models, so using a BP neural network defines a more representative $r$ value.

First, the historical data (1748 data) in the five cases are divided into two parts: one part is adopted to training, and the other part is applied to forecasting, and the number of nodes in each layer and other parameters in the BP neural network are consistent with the above. Second, we let $r = 50$, adjust the $r$ value at 10-min intervals, and record the forecasting accuracy and compare precision under the different $r$ value. Finally, we obtain a superb $r$ value. By means of experiments, $r$ was set to 220, 220, 220, 240, and 210. The process of denoising is shown in Fig. 4, part 2.

2) SIMULATION PROCESS OF OUR DEVELOPED MODEL
To demonstrate the prediction capacity of our proposed model, multiple experiments are performed to compare the accuracy of all forecasting consequences. The experimental process consists of the following steps:

(a) Use SSA to decompose and reconstruct the raw data, where the $r$ value is chosen by studying the historical data (train data). We conduct this step with the aim of enhancing the forecasting precision of our proposed model.

(b) Conduct four individual-method forecasts with reconstructed wind speed time series, and collect the forecast results (for all cases).

(c) Conduct four individual-method forecasts with original wind speed time series, and collect the forecast results (for all cases).

(d) Calculate the weight coefficients, and combine all models based on the proposed combined theory and CS.

(e) Make a comparison in terms of prediction performance for the involved models, which include the proposed combined model, the SSA-combined model, SVM, SSA-SVM, Elman, SSA-Elman, FA-BP, SSA-FA-BP, GA-PSO-BP, SSA-GA-PSO-BP, and individual BP.

(f) Based on experimental studies, the proposed model achieves wind speed forecasting computation.

Fig. 4 is the schematic of our developed combined strategy.

---

### TABLE 4. The experiment parameters of FA.

| Experimental parameters | Default value |
|-------------------------|---------------|
| Stopping criteria       |               |
| Maximum number of iteration | 100        |
| $\alpha$                | 0.25          |
| $\beta$                 | 1             |
| $\gamma$                | 1             |
| Population size         | 20            |

This table reveals the experiment parameters of FA. FA also has the property of simple operation and easy implementation. It must be noticed that when the iteration number reaches maximum, we will stop the algorithm process. And in our paper, the maximum iteration is 100.

### TABLE 5. The parameters of GA-PSO.

| Experimental parameters | Default value |
|-------------------------|---------------|
| Population size         | 20            |
| Maximum iteration times | 100           |
| Maximum iteration times of PSO | 20 |
| $C_1$                   | 1.49445       |
| $C_2$                   | 1.49445       |
| $w$                     | 0.7298        |
| GA cross rate           | 0.75          |
| GA mutation rate        | 0.15          |

This table presents the parameters of the developed hybrid model GA-PSO. Although PSO has some advantages in optimization process, the diversity of PSO is insufficient so it is hard to find global optimum with the small population, nevertheless, the selection, crossover and mutation in GA can make up the abovementioned deficiency. Thus, the combination of the two algorithms can enhance forecasting ability to a great extent.
TABLE 6. The forecasting results of BP, FA-BP, GA-PSO-BP in five cases.

| Model       | Index             | Dataset         |
|-------------|-------------------|-----------------|
|             | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 |
| BP          |        |        |        |        |        |
| MAE (m/s)   | 0.4393 | 0.406  | 0.4468 | 0.6103 | 0.4508 |
| MAPE (%)    | 6.71%  | 6.83%  | 7.60%  | 7.74%  | 7.65%  |
| MSE (m/s)   | 0.3705 | 0.3202 | 0.3455 | 0.5992 | 0.3543 |
| FA-BP       |        |        |        |        |        |
| MAE (m/s)   | 0.4412 | 0.4001 | 0.4385 | 0.6097 | 0.4415 |
| MAPE (%)    | 6.65%  | 6.74%  | 7.46%  | 7.61%  | 7.49%  |
| MSE (m/s)   | 0.3707 | 0.3281 | 0.3557 | 0.5689 | 0.3421 |
| GA-PSO-BP   |        |        |        |        |        |
| MAE (m/s)   | 0.4424 | 0.3995 | 0.4357 | 0.6007 | 0.4396 |
| MAPE (%)    | 6.66%  | 6.75%  | 7.44%  | 7.59%  | 7.48%  |
| MSE (m/s)   | 0.3787 | 0.3199 | 0.3359 | 0.5731 | 0.3428 |

This table demonstrates the forecasting results of single forecasting models, including BP, FA-BP, and GA-PSO-BP. Among the three indicators, MAPE is the most valuable standard. Its definition is: $$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$$, where $y_i$ is the real value, and $\hat{y}_i$ is the forecast value.

TABLE 7. The forecasting results of SSA-BP, SSA-FA-BP, and SSA-GAPSO-BP.

| Model       | Index             | Dataset         |
|-------------|-------------------|-----------------|
|             | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 |
| SSA-BP      |        |        |        |        |        |
| MAE (m/s)   | 0.2145 | 0.252  | 0.2328 | 0.271  | 0.2399 |
| MAPE (%)    | 3.28%  | 4.17%  | 4.01%  | 3.42%  | 4.22%  |
| MSE (m/s)   | 0.086  | 0.1086 | 0.0936 | 0.1175 | 0.0966 |
| SSA-FA-BP   |        |        |        |        |        |
| MAE (m/s)   | 0.2132 | 0.2494 | 0.2318 | 0.2703 | 0.2379 |
| MAPE (%)    | 3.26%  | 4.14%  | 3.99%  | 3.41%  | 4.16%  |
| MSE (m/s)   | 0.0851 | 0.1056 | 0.0928 | 0.1155 | 0.0958 |
| SSA-GAPSO-BP|        |        |        |        |        |
| MAE (m/s)   | 0.2128 | 0.2503 | 0.2318 | 0.2684 | 0.2383 |
| MAPE (%)    | 3.25%  | 4.15%  | 3.99%  | 3.39%  | 4.16%  |
| MSE (m/s)   | 0.0847 | 0.107  | 0.0928 | 0.1149 | 0.0958 |

This table gives the forecasting results of models with SSA data preprocessing technology. The parameters of single models are the same as those in Table 6. The forecasting performance in this table far surpass that in the previous table; that is, the data preprocessing approach is beneficial to the improvement of forecasting accuracy.

VI. EXPERIMENTAL ANALYSIS AND COMPARISON

In this section, to test the proposed model globally, four simulations are carried out to demonstrate the effectiveness of the developed model.

A. EXPERIMENT I: ANALYSIS AND COMPARISON OF INDIVIDUAL MODELS

The optimization algorithm can determine the optimal parameters for BP to enhance its global search performance; thus, hybrid algorithms based on FA and GA-PSO are used. Table 6 presents the forecasting results of BP, FA-BP, and GA-PSO-BP in the five cases.

As we can see from Table 6, the performance of individual BP is always stable; it follows that the set of experimental parameters in the BP model is reasonable. On the other hand, the values of the MSE, MAE and MAPE of the hybrid approaches are superior to those of BP. For instance, in Case 1, the MAPE of BP is 6.71%; however, the MAPE of FA-BP and GA-PSO-BP are 6.65% and 6.66%, respectively.

The results of simulation for SSA-BP, SSA-FA-BP, and SSA-GAPSO-BP are listed in Table 7. Fig. 5 clearly lists the MAPE values obtained from each involved model in every case.

Table 7 and Fig. 5 show that:
1. The SSA can achieve better performance, and all the models have more satisfactory forecasting precisions compared to Table 6, thereby proving that the $r$ value obtained in this paper is very efficient.
2. The overall performance of SSA-BP is not completely satisfactory, and higher volatility appears when SSA-BP and algorithm optimization are combined.
3. BP already has forecasting values after the SSA technique, so the optimization algorithm can only rarely optimize the result. It is clear that, compared with the optimization algorithm, SSA contributes more to improving the forecasting accuracy.

Table 8 presents the forecasting results of Elman, SSA-Elman, SVM and SSA-SVM in two different situations.

In the five cases, the value of MAPE is reduced by 1.42%, 2.87%, 2.03%, 2.26% and 2.39% for Elman and 2.56%, 2.52%, 2.64%, 3.74% and 2.70% for the SVM. Obviously, SSA greatly reduces the forecasting errors, which also proves...
that data preprocessing technology is vital for wind speed forecasting.

**B. EXPERIMENT II: ANALYSIS AND COMPARISON OF THE PROPOSED MODEL AND BENCHMARK MODELS**

Tables 9 and 10 establish the forecasting results of the combined model and benchmark models in the five cases.

Tables 9 and 10 indicate that the most satisfactory prediction results are always obtained from the proposed model; the proposed combined model has clear significant improvement, in contrast with the performance of the benchmark methods. Specifically, the average MAPE values of the developed model in Case 1 are 0.06%, 0.05%, 2.49%, and 0.86% superior to those of SSA-FA-BP, SSA-GA-PSO-BP, SSA-Elman, and SSA-SVM, respectively.

In addition, for wind speed forecasting, the data preprocessing technique is extremely necessary. For example, the average MAPE value of the proposed strategy in Case 1 is 3.35% lower than that of the weighting-based combined model without SSA.

**C. EXPERIMENT III: COMPARISON OF DIFFERENT WIND TURBINES AND MONTHS**

To fully test the proposed combination model for stability and universality, Cases 1-3 provide measurement data at the same time for different turbines, and Cases 4 and 5 are from different months. We can see from Table 10 that the combined method using SSA and the weight coefficient optimization calculation shows a higher accuracy standard. For all turbines and months, under the five sets

---

**FIGURE 5.** The Forecasting Results of BP, SSA-BP, FA-BP, SSA-FA-BP, GA-PSO-BP, SSA-GA-PSO-BP in five cases of Experiment I.

**TABLE 8.** The forecasting results of Elman, SSA-Elman, SVM and SSA-SVM.

| Model       | Index   | Dataset                                                                 |
|-------------|---------|-------------------------------------------------------------------------|
|             | MAE(m/s) | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 |
| Elman       | 0.4697  | 0.4304 | 0.4587 | 0.6056 | 0.4656 |
| MAPE (%)    | 7.11%   | 7.17%  | 7.75%  | 7.74%  | 7.83%  |
| MSE(m/s)    | 0.4139  | 0.3697 | 0.3666 | 0.5854 | 0.3746 |
| SSA-Elman   | 0.3726  | 0.3033 | 0.3276 | 0.4473 | 0.3171 |
| MAPE (%)    | 5.69%   | 5.30%  | 5.72%  | 5.48%  | 5.44%  |
| MSE(m/s)    | 0.2299  | 0.1533 | 0.159  | 0.3257 | 0.1614 |
| SVM         | 0.4351  | 0.3936 | 0.4299 | 0.596  | 0.43   |
| MAPE (%)    | 6.62%   | 6.64%  | 7.37%  | 7.57%  | 7.37%  |
| MSE(m/s)    | 0.3575  | 0.3129 | 0.3337 | 0.5598 | 0.3337 |
| SSA-SVM     | 0.259   | 0.2428 | 0.2765 | 0.3082 | 0.2683 |
| MAPE (%)    | 4.06%   | 4.12%  | 4.73%  | 3.83%  | 4.67%  |
| MSE(m/s)    | 0.1133  | 0.1017 | 0.1228 | 0.1512 | 0.1173 |

This table shows the forecasting results of ENN, SSA-ENN, SVM, and SSA-SVM. Similar to the situation in Tables 6 and 7, the forecasting capacity of all the models is compared based on the three criteria in the five cases, which directly demonstrates the forecasting capacity of the individual models and models using the denoising technique.
of data, the MAPE values are always approximately 3%, and the performance of the combined model is always stable.

**D. EXPERIMENT IV: TEST WITH R VALUES**

To determine whether the r value obtained by studying the historical data is appropriate for the final test set and to identify a reasonable range of the correlation coefficient as well as the sum of component contributions for future wind speed forecasting in Penglai, we carried out Experiment IV. The experimental results are shown in Tables 11-15. The MAPE (BP) represents the performances of FA-BP and GA-PSO-BP. The correlation coefficient denotes the degree of relevance between the reconstructed time series and the true time series. The MAPE is the mean absolute percent error between the reconstructed time sequences and the true time sequences.

Table 11 shows when the r value is 50, the forecast errors of all the models are high; with the increase in r, the forecasting accuracy becomes more desirable, but when r exceeds 220, the forecasting accuracy becomes undesirable. Thus, we can see that the proposed approach of defining r is effective and reasonable. It is noteworthy that when r is small, the reconstructed time series is a poor description of the original series, which leads to poor forecasting performance; thus, in order to save time in the experiment, we set r = 170 as the initial value in the succeeding cases.

This table demonstrates the forecasting results of models for comparison in Experiment II. It must be mentioned that data preprocessing technology is not included in this comparison. And the benchmark models have been discussed in the previous experiment.

### TABLE 9. The forecasting results of the proposed model without SSA.

| Dataset | Index | Model | FA-BP  | GA-PSO-BP | Elman | SVM | Combined |
|---------|-------|-------|--------|-----------|-------|-----|----------|
| Case 1  | MAE(m/s)| 0.4412 | 0.4424 | 0.4697 | 0.4351 | 0.4281 |
|         | MAPE (%)| 6.65% | 6.66% | 7.11% | 6.62% | 6.55% |
|         | MSE(m/s)| 0.3707 | 0.3780 | 0.4139 | 0.3753 | 0.3534 |
| Case 2  | MAE(m/s)| 0.4001 | 0.3995 | 0.4304 | 0.3936 | 0.3923 |
|         | MAPE (%)| 6.74% | 6.75% | 7.17% | 6.64% | 6.53% |
|         | MSE(m/s)| 0.3281 | 0.3199 | 0.3697 | 0.3129 | 0.3204 |
| Case 3  | MAE(m/s)| 0.4385 | 0.4537 | 0.4587 | 0.4299 | 0.4307 |
|         | MAPE (%)| 7.46% | 7.44% | 7.75% | 7.37% | 7.34% |
|         | MSE(m/s)| 0.3357 | 0.3339 | 0.3666 | 0.3337 | 0.3401 |
| Case 4  | MAE(m/s)| 0.5689 | 0.5731 | 0.5854 | 0.5598 | 0.5583 |
|         | MAPE (%)| 7.61% | 7.59% | 7.74% | 7.57% | 7.48% |
|         | MSE(m/s)| 0.5689 | 0.5731 | 0.5854 | 0.5598 | 0.5583 |
| Case 5  | MAE(m/s)| 0.4415 | 0.4396 | 0.4656 | 0.4300 | 0.4295 |
|         | MAPE (%)| 7.49% | 7.48% | 7.83% | 7.37% | 7.33% |
|         | MSE(m/s)| 0.3421 | 0.3428 | 0.3746 | 0.3337 | 0.3363 |

This table demonstrates the forecasting results of models for comparison in Experiment II. It must be mentioned that data preprocessing technology is not included in this comparison. And the benchmark models have been discussed in the previous experiment.

### TABLE 10. The forecasting results of the proposed model based on SSA.

| Dataset | Index | Model | SSA-BP | SSA-GA-PSO-BP | SSA-Elman | SSA-SVM | Combined |
|---------|-------|-------|--------|--------------|-----------|---------|----------|
| Case 1  | MAE(m/s)| 0.2132 | 0.2128 | 0.3700 | 0.2590 | 0.2114 |
|         | MAPE (%)| 3.26% | 3.25% | 5.69% | 4.06% | 3.20% |
|         | MSE(m/s)| 0.0851 | 0.0847 | 0.2299 | 0.1133 | 0.0855 |
| Case 2  | MAE(m/s)| 0.2494 | 0.2503 | 0.3033 | 0.2428 | 0.2385 |
|         | MAPE (%)| 4.14% | 4.15% | 5.30% | 4.12% | 4.01% |
|         | MSE(m/s)| 0.1056 | 0.1070 | 0.1533 | 0.1017 | 0.0977 |
| Case 3  | MAE(m/s)| 0.2320 | 0.2309 | 0.3276 | 0.2765 | 0.2311 |
|         | MAPE (%)| 3.99% | 3.99% | 5.72% | 4.73% | 3.98% |
|         | MSE(m/s)| 0.0934 | 0.092 | 0.159 | 0.1228 | 0.0924 |
| Case 4  | MAE(m/s)| 0.2703 | 0.2684 | 0.4719 | 0.3105 | 0.2578 |
|         | MAPE (%)| 3.32% | 3.31% | 5.48% | 3.83% | 3.12% |
|         | MSE(m/s)| 0.1155 | 0.1149 | 0.3696 | 0.1510 | 0.1103 |
| Case 5  | MAE(m/s)| 0.2379 | 0.2383 | 0.3176 | 0.2620 | 0.2384 |
|         | MAPE (%)| 3.88% | 3.88% | 5.44% | 4.67% | 3.84% |
|         | MSE(m/s)| 0.0958 | 0.0958 | 0.1669 | 0.1133 | 0.0964 |
TABLE 11. Test results of the single model on different \( r \) for Case 1.

| \( r \) value | Correlation coefficient | MAPE | MAPE(BP) | MAPE(Elman) | MAPE(SVM) |
|--------------|-------------------------|------|----------|-------------|-----------|
| 50           | 0.9708                  | 9.29%| 10.01%   | 8.35%       | 8.19%     |
| 60           | 0.9755                  | 8.34%| 9.95%    | 7.65%       | 7.58%     |
| 70           | 0.9794                  | 7.61%| 6.68%    | 7.33%       | 6.80%     |
| 80           | 0.9822                  | 7.14%| 6.61%    | 6.96%       | 6.67%     |
| 90           | 0.9838                  | 6.75%| 6.37%    | 6.73%       | 6.39%     |
| 100          | 0.9855                  | 6.29%| 5.93%    | 6.38%       | 5.94%     |
| 110          | 0.9869                  | 6.00%| 5.67%    | 6.20%       | 5.64%     |
| 120          | 0.9884                  | 5.70%| 5.39%    | 6.18%       | 5.46%     |
| 130          | 0.9896                  | 5.41%| 5.32%    | 6.20%       | 5.48%     |
| 140          | 0.9906                  | 5.14%| 5.01%    | 6.13%       | 5.63%     |
| 150          | 0.9917                  | 4.81%| 4.78%    | 6.17%       | 5.08%     |
| 160          | 0.9923                  | 4.59%| 4.71%    | 6.07%       | 5.09%     |
| 170          | 0.993                   | 4.35%| 4.29%    | 6.15%       | 4.69%     |
| 180          | 0.9936                  | 4.21%| 4.11%    | 6.05%       | 4.65%     |
| 190          | 0.9941                  | 4.03%| 3.95%    | 5.97%       | 4.79%     |
| 200          | 0.9946                  | 3.88%| 3.66%    | 5.79%       | 4.57%     |
| 210          | 0.9950                  | 3.73%| 3.45%    | 5.75%       | 4.46%     |
| 220          | **0.9954**              | **3.57%**| **3.28%**| **5.69%**   | **4.06%** |
| 230          | 0.9957                  | 3.42%| 3.34%    | 5.99%       | 4.07%     |
| 240          | 0.9960                  | 3.29%| 3.35%    | 6.01%       | 4.09%     |
| 250          | 0.9963                  | 3.20%| 3.35%    | 6.03%       | 4.06%     |
| 260          | 0.9965                  | 3.12%| 3.61%    | 6.11%       | 4.17%     |

This table demonstrates the test results of the single model on different \( r \) in Case 1. Correlation coefficient denotes the relevance degree between reconstructed time series and true time series. The definition is \( p = \frac{\sum_{i=1}^{m} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{m} (y_i - \bar{y})^2 \sum_{i=1}^{m} (\hat{y}_i - \bar{\hat{y}})^2}} \), where \( y_i \) represents the true time sequence and \( \hat{y}_i \) is the reconstructed time sequence, \( \bar{y} \) and \( \bar{\hat{y}} \) are corresponding mean values. MAPE is the mean absolute percent error.

TABLE 12. The test results of the single model on different \( r \) for Case 2.

| \( r \) value | Correlation coefficient | MAPE | MAPE(BP) | MAPE(Elman) | MAPE(SVM) |
|--------------|-------------------------|------|----------|-------------|-----------|
| 170          | 0.9922                  | 4.60%| 4.84%    | 6.04%       | 4.93%     |
| 180          | 0.9928                  | 4.42%| 4.60%    | 5.87%       | 4.78%     |
| 190          | 0.9932                  | 4.26%| 4.40%    | 5.57%       | 4.54%     |
| 200          | 0.9937                  | 4.10%| 4.46%    | 5.36%       | 4.55%     |
| 210          | 0.9942                  | 3.93%| 4.30%    | 5.32%       | 4.39%     |
| 220          | **0.9946**              | **3.75%**| **4.12%**| **5.30%**   | **4.12%** |
| 230          | 0.9950                  | 3.61%| 4.56%    | 5.41%       | 4.30%     |
| 240          | 0.9954                  | 3.49%| 4.88%    | 6.06%       | 5.14%     |
| 250          | 0.9957                  | 3.38%| 5.25%    | 6.41%       | 5.37%     |
| 260          | 0.996                  | 3.25%| 5.28%    | 6.56%       | 5.31%     |

This table demonstrates the test results of the single model on different \( r \) in Case 2. Correlation coefficient denotes the relevance degree between reconstructed time series and true time series. This definition is the same with that in Table 11. MAPE is the mean absolute percent error. The bold fonts represent the best results.

For Cases 2 and 3, as Tables 12 and 13 shown, when \( r = 220 \), all the models have the best forecasting accuracy, and the trends presented in Tables 12-13 are in accordance with those presented in Table 11.

Table 14 indicates that all the models provide the best forecasting accuracy when \( r = 240 \). However, for the SVM, when \( r = 190, 200, \) and 240, the MAPE values of the SVM are all 3.83%, and the forecasting results of the SVM show a different trend.

The test results of the single model for different \( r \) in Case 5 are shown in Table 15. When \( r = 210 \), the forecasting accuracy of BP and Elman are the best across the experimental cases, SVM also obtain a higher accuracy. This situation shows that the method of defining \( r \) has some limitations; it cannot obtain the best \( r \) value for the entire model, but it can apply an appropriate \( r \) value to reduce forecasting error.

In addition, through the series of experiments and Fig. 6, it can be concluded that:

(a) When the \( r \) value is small, the forecasting accuracy is undesirable because the reconstructed time series is a poor description of the original series. With the increase in \( r \), the forecasting accuracy improves or is sustained.
TABLE 13. The test results of the single model on different $r$ for Case 3.

| $r$ value | Correlation coefficient | MAPE | MAPE(BP) | MAPE(Elman) | MAPE(SVM) |
|-----------|-------------------------|------|----------|-------------|-----------|
| $r=170$   | 0.9918                  | 4.46%| 4.77%    | 6.04%       | 5.08%     |
| $r=180$   | 0.9924                  | 4.27%| 4.47%    | 5.88%       | 4.82%     |
| $r=190$   | 0.9931                  | 4.09%| 4.37%    | 5.86%       | 4.61%     |
| $r=200$   | 0.9936                  | 3.95%| 4.20%    | 5.81%       | 4.39%     |
| $r=210$   | 0.9939                  | 3.81%| 4.13%    | 5.75%       | 4.67%     |
| $r=220$   | 0.9944                  | 3.69%| 4.01%    | 5.72%       | 4.73%     |
| $r=230$   | 0.9948                  | 3.55%| 4.00%    | 5.73%       | 4.91%     |
| $r=240$   | 0.9952                  | 3.41%| 5.56%    | 6.01%       | 5.38%     |
| $r=250$   | 0.9954                  | 3.35%| 6.47%    | 6.39%       | 6.08%     |
| $r=260$   | 0.9958                  | 3.22%| 6.57%    | 6.45%       | 6.15%     |

This table demonstrates the test results of the single model on different $r$ in Case 3. Correlation coefficient denotes the relevance degree between reconstructed time series and true time series. This definition is the same with that in Table 11. MAPE is the mean absolute percent error. The bold fonts represent the best results.

TABLE 14. The test results of the single model on different $r$ for Case 4.

| $r$ value | Correlation coefficient | MAPE | MAPE(BP) | MAPE(Elman) | MAPE(SVM) |
|-----------|-------------------------|------|----------|-------------|-----------|
| $r=170$   | 0.9911                  | 4.55%| 3.83%    | 6.06%       | 4.13%     |
| $r=180$   | 0.9918                  | 4.39%| 3.65%    | 5.86%       | 3.91%     |
| $r=190$   | 0.9924                  | 4.23%| 3.55%    | 5.84%       | 3.83%     |
| $r=200$   | 0.9931                  | 4.09%| 3.52%    | 5.84%       | 3.83%     |
| $r=210$   | 0.9936                  | 3.93%| 3.47%    | 5.83%       | 3.84%     |
| $r=220$   | 0.9940                  | 3.79%| 3.42%    | 5.82%       | 3.85%     |
| $r=230$   | 0.9943                  | 3.69%| 3.42%    | 5.80%       | 3.89%     |
| $r=240$   | 0.9947                  | 3.58%| 3.35%    | 5.48%       | 3.83%     |
| $r=250$   | 0.9951                  | 3.43%| 3.67%    | 6.05%       | 4.26%     |
| $r=260$   | 0.9954                  | 3.31%| 4.12%    | 7.49%       | 4.38%     |

This table demonstrates the test results of the single model on different $r$ in Case 4. Correlation coefficient denotes the relevance degree between reconstructed time series and true time series. This definition is the same with that in Table 11. MAPE is the mean absolute percent error. The bold fonts represent the best results.

TABLE 15. The test results of the single model on different $r$ for Case 5.

| $r$ value | Correlation coefficient | MAPE | MAPE(BP) | MAPE(Elman) | MAPE(SVM) |
|-----------|-------------------------|------|----------|-------------|-----------|
| $r=170$   | 0.9918                  | 4.46%| 4.77%    | 6.21%       | 5.08%     |
| $r=180$   | 0.9924                  | 4.27%| 4.47%    | 6.18%       | 4.82%     |
| $r=190$   | 0.9931                  | 4.09%| 4.23%    | 5.93%       | 4.61%     |
| $r=200$   | 0.9936                  | 3.95%| 4.00%    | 5.80%       | 4.39%     |
| $r=210$   | **0.9939**              | **3.81%**| **3.89%**| **5.44%** | **4.67%** |
| $r=220$   | 0.9944                  | 3.69%| 3.99%    | 6.21%       | 4.73%     |
| $r=230$   | 0.9948                  | 3.55%| 4.81%    | 6.29%       | 4.91%     |

This table demonstrates the test results of the single model on different $r$ in Case 5. Correlation coefficient denotes the relevance degree between reconstructed time series and true time series. This definition is the same with that in Table 11. MAPE is the mean absolute percent error. The bold fonts represent the best results.

When $r$ exceeds a certain value, the forecasting accuracy becomes close to the forecasting accuracy defined by the original time series data because the reconstructed time series data is almost the same as the original time series data.

(b) The method of defining the $r$ value proposed in this paper can apply to the forecasting process well, even if it cannot apply the best $r$ value for every model.

For wind speed forecasting in the Penglai area, the appropriate range of the correlation coefficient between the original series and the reconstructed series is approximately 99.5%, and the sum of the $r$ component contribution is at least 99%; these results can provide a reference for the selection of $r$ for the wind farm in Penglai.

E. EXPERIMENT V: THE DETERMINATION OF FORECASTING SUB-MODELS

In this section, to verify the effectiveness of the selected forecasting sub-models of the combined model, we conducted...
two simulation experiments in terms of the forecasting performance of several individual forecasting models (ARIMA, BP, SVM, ELM, Elman, and a general regression neural network (GRNN)) and some hybrid models using different optimization algorithms (GA-BP, PSO-BP, FA-BP, and GA-PSO-BP). To enhance the reliability of the forecasting results, we also repeated the simulation for each model ten times and then took the mean values of the three evaluation criteria as the final criteria values. The detailed forecasting results are shown in Tables 16 and 17. From these tables, we can observe that:

(1) When we compare the forecasting accuracy of the six individual forecasting models, we can find that the SVM can provide the best forecasting performance in Cases 1 to 5, followed by BP and Elman. The forecasting capacity of ARIMA, ELM, and GRNN are inferior to SVM, BP and Elman in all cases. For instance, the mean MAPE values of SVM, BP, Elman, ARIMA, ELM, and GRNN are 7.11%, 7.31%, 7.52%, 7.96%, 8.47%, and 8.94%, respectively. Thus, the SVM and Elman, as two effective individual forecasting models, are used to construct our developed combined model.

(2) Moreover, to overcome the drawbacks of BP (i.e., it easily falls into local optimization, it misses the globally optimal solution and it converges slowly) and further enhance the forecasting performance of sub-models, the weights and thresholds of the BP neural network are optimized by...
optimization algorithms, including GA, PSO, FA, and GA-PSO. The comparative results of hybrid models integrating the BP neural network and these four optimization algorithms are presented in Table 17, from which we can observe that optimization algorithms can enhance the forecasting capacity of a BP neural network to some extent, with lower MAE, MAPE, and MSE values. Furthermore, the optimization performance of FA and GA-PSO is better than that of GA and PSO. Thus, the hybrid FA-BP and GA-PSO-BP are also adopted as sub-models to build the final combined model.

In brief, the SVM, Elman, FA-BP, and GA-PSO-BP are used as benchmark forecasting models to structure the combined model.

**F. EXPERIMENT VI: COMPUTATION TIME OF THE TRAINING MODEL**

In this section, we used 1508 samples as training data and calculated the running time for training the proposed combined strategy and all compared models in all datasets. The computation results are presented in Table 18, from which we can obtain the following results: (1) The running time of training individual models, such as BP, ELM, and Elman, is obviously shorter than that of the hybrid models and our proposed combined model. (2) When the optimization algorithms and data preprocessing technologies are added to build a hybrid model, the computation time increases to some extent. For example, the mean computation time of FA-BP, GA-PSO-BP, SSA-FA-BP, SSA-PSO-BP, SSA-SVM, and SSA-Elman are 4.5691 s, 5.6742 s, 11.4632 s, 14.1280 s, 11.0476 s, and 11.2613 s, respectively, which are longer than those of the individual forecasting models. (3) The runtime of training the developed combined model is the longest among all involved models, with values of 19.8912 s, 20.2926 s, 20.4187 s, 20.5094 s, and 19.5521 s from Case 1 to Case 5, respectively. The computation time is within an acceptable range.

**G. TESTS WITH DIFFERENT CAPACITY**

To further demonstrate the effectiveness of our developed combined model in wind speed forecasting with different capacity, we conduct a discussion in this section, which considers the original wind speed data with lengths of 874, 1748, 3496, and 5244. The detailed results are presented in Table 19. From the table, we can observe that the proposed combined model can be used in wind speed forecasting with different capacity, because the evaluation criteria values in the five cases vary within reasonable limits. For example, the MAPE values are within the range of 3%∼5%. Moreover, when the data length is 1748, the forecasting capacity of...
TABLE 19. The forecasting performance of the combined model in wind speed forecasting with different capacity.

| Capacity | Model | Index | Case 1 | Case 2 | Case 3 | Case 4 | Case 5 |
|----------|-------|-------|--------|--------|--------|--------|--------|
| 874      | The combined model | MAE(m/s) | 0.2137 | 0.2651 | 0.2418 | 0.2791 | 0.2497 |
|          |       | MAPE (%) | 3.42%  | 4.21%  | 4.09%  | 3.25%  | 3.96%  |
|          |       | MSE(m/s) | 0.0960 | 0.1096 | 0.0933 | 0.1199 | 0.1003 |
| 1748     | The combined model | MAE(m/s) | 0.2114 | 0.2385 | 0.2311 | 0.2578 | 0.2384 |
|          |       | MAPE (%) | 3.20%  | 4.01%  | 3.98%  | 3.12%  | 3.84%  |
|          |       | MSE(m/s) | 0.0855 | 0.0977 | 0.0924 | 0.1103 | 0.0964 |
| 3496     | The combined model | MAE(m/s) | 0.2125 | 0.2394 | 0.2498 | 0.2773 | 0.2478 |
|          |       | MAPE (%) | 3.33%  | 4.12%  | 4.27%  | 3.21%  | 3.87%  |
|          |       | MSE(m/s) | 0.0908 | 0.1007 | 0.1012 | 0.1186 | 0.1001 |
| 5244     | The combined model | MAE(m/s) | 0.2123 | 0.2565 | 0.2488 | 0.2697 | 0.2395 |
|          |       | MAPE (%) | 3.26%  | 4.14%  | 4.24%  | 3.19%  | 3.85%  |
|          |       | MSE(m/s) | 0.0867 | 0.1028 | 0.1008 | 0.1160 | 0.0999 |

This table reveals the forecasting accuracy of our developed model in wind speed forecasting with different capacity in five cases. The length of wind speed time series is set to 874, 1748, 3496, and 5244, respectively. It is obvious that our proposed model has good forecasting performance in wind speed forecasting with different capacity.

Our developed combined model is best with MAPE values of 3.20%, 4.01%, 3.98%, 3.12%, and 3.84% from case 1 to case 5, respectively. Too few data samples, such as 874, may cause a high forecasting error due to inadequate training, while too many data samples can enhance forecasting accuracy but result in a much higher cost. Thus, in our paper, we adopted 1748 samples to verify the forecasting performance of our developed combined model. The simulation results demonstrate that our developed combined model can provide accurate and stable wind speed forecasting results.

VII. CONCLUSION

Anticipating the characteristic of wind speed with precision and accuracy is significant because of the increasingly crucial effect of wind energy in power systems. Wind speed data are intermittent and random, which increases forecasting difficulty to a great extent. Most previous studies have made efforts to develop a precise and stable forecasting model. However, many of these studies cannot achieve the desired accuracy across various sites and various wind speed time series. Motivated by recent progress in combined model forecasting, we developed a combined strategy based on data preprocessing and parameter optimization calculation, which is able to achieve precise forecasting for different turbines and months. In this paper, ten-minute wind speed series of three wind turbines in Penglai, China, were adopted as cases to validate the effectiveness and stability of the model.

We evaluated and compared the combined model with the component models and obtained the appropriate range of the correlation coefficient and the sum of the contribution of r components to define the r value. Based on empirical results and discussion, the following conclusions can be drawn.

(a) The forecasting capacity of our proposed model is best across different component forecasting models, with MAPE values of 3.20%, 4.01%, 3.98%, 3.12%, and 3.84% from case 1 to case 5, respectively, which improved 2.49%, 1.29%, 1.74%, 2.36%, and 1.60%, respectively, compared with the worst comparative model, SSA-Elman; (b) The SSA data pretreatment technique shows strong performance in dealing with erratic and fluctuating wind speed series, which contributes greatly to improving forecasting performance and accuracy. The MAPE values of the proposed combined model with and without SSA, are 6.55%, 6.53%, 7.34%, 7.48%, and 7.33%, and 3.20%, 4.01%, 3.98%, 3.12%, and 3.84%, from case 1 to case 5, respectively; (c) The method of defining the r value proposed in this paper also applies to the forecasting process. In addition, for wind speed forecasting in the Penglai area, the appropriate range of the correlation coefficient between the original series and the reconstructed series is approximately 99.5%, and the sum of the contribution of the r component is at least 99%. Our developed combined model can effectively integrate the advantages of different sub-models and further enhance the forecasting accuracy and stability. Therefore, the proposed model can be utilized in wind speed forecasting with great promise in the future and provide a reference for the management of large wind farms.

The main drawbacks of the proposed combined system can be attributed to two factors. One factor is that other factors affecting wind speed forecasting are not contained; the other factor is that only forecasting accuracy is considered in the fitness function of the optimization algorithm, ignoring the forecasting stability, which is also important in optimization problems. To overcome these drawbacks, more potential factors that may impact wind speed forecasting performance must be investigated. Moreover, to enhance optimization
ability, more novel and valid optimization algorithms with high universality must be explored in future research.

APPENDIX

LIST OF ABBREVIATIONS

| FA     | firefly algorithm |
| WD     | Wavelet decomposition |
| GA     | genetic algorithm |
| ANN    | artificial neural network |
| ENN    | Elman neural network |
| ICS    | improved cuckoo search |
| MSE    | mean square error |
| IMFs   | intrinsic mode functions |
| NWP    | numerical weather prediction |
| ASMO   | a surrogate modeling-based optimization |
| EEEMD  | ensemble empirical mode decomposition |
| ARIMA  | autoregressive integrated moving average model |
| FEEMD  | fast ensemble empirical mode decomposition |
| MOALO  | multi-objective ant lion algorithm |
| GRNN   | general regression neural network |
| MSSA   | multi-objective salp swarm algorithm |
| WT     | Wavelet Transform |
| PSO    | particle swarm optimization |
| CS     | cuckoo search |
| WPD    | wavelet packet decomposition |
| MAE    | mean absolute error |
| SSE    | error sum of square |
| SVM    | support vector machine |
| SVR    | support vector regression |
| ELM    | extreme learning machine |
| BPNN   | back propagation neural network |
| CNN    | Convolutional Neural Network |
| SSA    | singular spectrum analysis |
| EMD    | empirical model decomposition |
| MAPE   | mean absolute percentage error |
| LSSVM  | Least squares support vector machine |

REFERENCES

[1] T. Niu, J. Wang, H. Lu, and P. Du, “Uncertainty modeling for chaotic time series based on optimal multi-input multi-output architecture: Application to offshore wind speed,” *Energy Convers. Manage.*, vol. 156, pp. 597–617, Jan. 2018, doi: 10.1016/j.enconman.2017.11.071.

[2] S. Salcedo-Sanz, Á. M. Pérez-Bellido, E. G. Ortiz-García, A. Portilla-Figueras, L. Prieto, and F. Correoso, “Accurate short-term wind speed prediction by exploiting diversity in input data using banks of artificial neural networks,” *Neurocomputing*, vol. 72, nos. 4–6, pp. 1336–1341, Jan. 2009, doi: 10.1016/j.neucom.2008.09.010.

[3] Data on Global Onshore and Offshore Wind Power in 2018. Accessed: Jul. 30, 2019. [Online]. Available: http://news.bjx.com.cn/html/20190516/980725.shtml

[4] C. Tian and Y. Hao, “Point and interval forecasting for carbon price based on an improved analysis-forecast system,” *Appl. Math. Model.*, vol. 79, pp. 126–144, Mar. 2020, doi: 10.1016/j.apm.2019.10.022.

[5] Z. Liu, P. Jiang, L. Zhang, and X. Niu, “A combined forecasting model for time series: Application to short-term wind speed forecasting,” *Appl. Energy*, vol. 259, Feb. 2020, Art. no. 114137, doi: 10.1016/j.apenergy.2019.114137.

[6] Q. Zhou, C. Wang, and G. Zhang, “Hybrid forecasting system based on an optimal model selection strategy for different wind speed forecasting problems,” *Appl. Energy*, vol. 250, pp. 1559–1580, Sep. 2019, doi: 10.1016/j.apenergy.2019.05.016.

[7] Z. Di, J. Ao, Q. Duan, J. Wang, W. Gong, C. Shen, Y. Gan, and Z. Liu, “Improving WRF model turbine-height wind-speed forecasting using a surrogate- based automatic optimization method,” *Atmos. Res.*, vol. 226, pp. 1–16, Sep. 2019, doi: 10.1016/j.atmosres.2019.04.011.

[8] X. Zhao, J. Liu, D. Yu, and J. Chang, “One-day-ahead probabilistic wind speed forecast based on optimized numerical weather prediction data,” *Energy Convers. Manage.*, vol. 164, pp. 560–569, May 2018, doi: 10.1016/j.enconman.2018.03.030.

[9] P. Du, J. Wang, W. Yang, and T. Niu, “A novel hybrid model for short-term wind power forecasting,” *Appl. Soft Comput.*, vol. 80, pp. 93–106, Jul. 2019, doi: 10.1016/j.asoc.2019.03.035.

[10] R. Wang, J. Li, J. Wang, and C. Gao, “Research and application of a hybrid wind energy forecasting system based on data processing and an optimized extreme learning machine,” *Energies*, vol. 11, no. 7, p. 1712, Jul. 2018, doi: 10.3390/en11071712.

[11] Y. Hao and C. Tian, “The study and application of a novel hybrid system for air quality early-warning,” *Appl. Soft Comput.*, vol. 74, pp. 729–746, Jan. 2019, doi: 10.1016/j.asoc.2018.09.005.

[12] Assim, S. N. Singh, and A. Mohapatra, “Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting,” *Renew. Energy*, vol. 136, pp. 758–768, Jun. 2019, doi: 10.1016/j.renene.2019.01.031.

[13] N. S. Pearre and L. G. Swan, “Statistical approach for improved wind speed forecasting for wind power production,” *Sustain. Energy technol. Assessments*, vol. 27, pp. 180–191, Jun. 2018, doi: 10.1016/j.seta.2018.04.010.

[14] H. Li, J. Wang, R. Li, and H. Lu, “Novel analysis–forecast system based on multi-objective optimization for air quality index,” *J. Cleaner Prod.*, vol. 208, pp. 1365–1383, Jan. 2019, doi: 10.1016/j.jclepro.2018.10.129.

[15] H. Yang, Z. Zhu, C. Li, and R. Li, “A novel combined forecasting system for air pollutants concentration based on fuzzy theory and optimization of aggregation weight,” *Appl. Soft Comput.*, vol. 87, Feb. 2020, Art. no. 105972, doi: 10.1016/j.asoc.2019.105972.

[16] H. Bo, X. Niu, and J. Wang, “Wind speed forecasting system based on the variational mode decomposition strategy and immune selection multi-objective dragonfly optimization algorithm,” *IEEE Access*, vol. 7, pp. 178063–178081, 2019, doi: 10.1109/ACCESS.2019.2957062.

[17] X. Zhang, J. Wang, and Y. Gao, “A hybrid short-term electricity price forecasting framework: Cuckoo search-based feature selection with singular spectrum analysis and SVM,” *Energy Econ.*, vol. 81, pp. 899–913, Jun. 2019, doi: 10.1016/j.eneco.2019.05.026.

[18] H. Yang, Z. Jiang, and H. Lu, “A hybrid wind speed forecasting system based on a ‘decomposition and ensemble strategy and fuzzy time series’,” *Energies*, vol. 10, no. 9, p. 1422, Sep. 2017, doi: 10.3390/en10091422.

[19] J. Wang, T. Niu, H. Lu, W. Yang, and P. Du, “A novel framework of reservoir computing for deterministic and probabilistic wind power forecasting,” *IEEE Trans. Sustain. Energy*, vol. 11, no. 1, pp. 337–349, Jan. 2020, doi: 10.1109/tste.2019.2890875.

[20] C. Wu, J. Wang, X. Chen, P. Du, and W. Yang, “A novel hybrid system based on multi-objective optimization for wind speed forecasting,” *Renew. Energy*, vol. 146, pp. 149–165, Feb. 2020, doi: 10.1016/j.renene.2019.04.157.

[21] W. Yang, J. Wang, H. Lu, T. Niu, and P. Du, “Hybrid wind energy forecasting and analysis system based on divide and conquer scheme: A case study in China,” *J. Cleaner Prod.*, vol. 222, pp. 942–959, Jun. 2019, doi: 10.1016/j.jclepro.2019.03.036.

[22] Y. Hao and C. Tian, “A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting,” *Appl. Energy*, vol. 238, pp. 368–383, Mar. 2019, doi: 10.1016/j.apenergy.2019.01.063.

[23] D. Liu, D. Niu, H. Wang, and L. Fan, “Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm,” *Renew. Energy*, vol. 62, pp. 592–597, Feb. 2014, doi: 10.1016/j.renene.2013.08.011.

[24] H. Liu, C. Chen, H.-Q. Tian, and Y.-F. Li, “A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks,” *Renew. Energy*, vol. 48, pp. 545–556, Dec. 2012, doi: 10.1016/j.renene.2012.06.012.

[25] J. B. Elsner and A. A. Tsonis, *Singular Spectrum Analysis: A New Tool in Time Series Analysis*. Springer, 1996.
[26] Y. Gao, C. Qu, and K. Zhang, “A hybrid method based on singular spectrum analysis, firefly algorithm, and BP neural network for short-term wind speed forecasting,” Energies, vol. 9, no. 10, p. 757, Sep. 2016, doi: 10.3390/en9100757.

[27] H. Liu, X. Li, Y. Li, Z. Duan, and Y. Xu, “Smart wind speed deep learning based multi-step forecasting model using singular spectrum analysis, convolutional gated recurrent unit network and support vector regression,” Renew. Energy, vol. 143, pp. 842–854, Dec. 2019, doi: 10.1016/j.renene.2019.05.039.

[28] Liu, Jin, and Gao, “A new hybrid approach for short-term electric load forecasting applying support vector machine with ensemble empirical mode decomposition and whale optimization,” Energies, vol. 12, no. 8, p. 5290, Apr. 2019, doi: 10.3390/en12085290.

[29] P. Jang, C. Li, R. Li, and H. Yang, “An innovative hybrid air pollution early-warning system based on pollutants forecasting and Extentics evaluation,” Knowl.-Based Syst., vol. 164, pp. 174–192, Jan. 2019, doi: 10.1016/j.knosys.2018.10.036.

[30] L. Xiao, J. Wang, Y. Dong, and J. Wu, “Combined forecasting models for wind energy forecasting: A case study in China,” Renew. Sustain. Energy Rev., vol. 44, pp. 271–288, Apr. 2015, doi: 10.1016/j.rser.2014.12.012.

[31] P. Jiang and Z. Liu, “Variable weights combined model based on multi-objective optimization for short-term wind speed forecasting,” Appl. Soft Comput., vol. 82, Sep. 2019, Art. no. 105587, doi: 10.1016/j.asoc.2019.105587.

[32] J. Wang, Y. Gao, and X. Chen, “A novel hybrid interval prediction approach based on modified lower upper bound estimation in combination with multi-objective salp swarm algorithm for short-term load forecasting,” Energies, vol. 11, no. 6, p. 1561, Jun. 2018, doi: 10.3390/en11061561.

[33] R. Li and Y. Jin, “A wind speed interval prediction system based on multi-objective optimization for machine learning method,” Appl. Energy, vol. 228, pp. 2207–2220, Oct. 2018, doi: 10.1016/j.apenergy.2018.07.032.

[34] I. A. Moosa and J. J. Vaz, “Cointegration, error correction and exchange rate forecasting,” J. Int. Financial Mark. Inst. Money, vol. 44, pp. 21–34, Sep. 2016, doi: 10.1016/j.intfin.2016.04.007.

[35] X. Zhang, P. Liu, Y. Zhao, C. Deng, Z. Li, and M. Xiong, “Error correction-based forecasting of reservoir water levels: Improving accuracy over multiple lead times,” Environ. Model. Softw., vol. 104, pp. 27–39, Jun. 2018, doi: 10.1016/j.envsoft.2018.02.017.

[36] T. Qiao, J. Ren, C. Craigie, J. Zabalza, C. Maltin, and S. Marshall, “Singular spectrum analysis for improving hyperspectral imaging based beef eating quality evaluation,” Comput. Electron. Agricult., vol. 115, pp. 21–25, Jul. 2015, doi: 10.1016/j.compag.2015.05.007.

[37] C. M. Rocco, “Singular spectrum analysis and forecasting of failure time series,” Rel. Eng. Syst. Saf., vol. 114, pp. 126–136, Jun. 2013, doi: 10.1016/j.ress.2013.01.007.

[38] J. L. Elman, “Finding structure in time,” Cogn. Sci., vol. 14, no. 2, pp. 179–211, 1990, doi: 10.1080/03640219090900212.

[39] X. Chandra and M. Zhang, “Cooperative coevolution of Elman recurrent neural networks for chaotic time series prediction,” Neurocomputing, vol. 86, pp. 116–123, Jun. 2012, doi: 10.1016/j.neucom.2012.01.014.

[40] M. Cacciola, G. Megali, D. Pellicanó, and F. C. Morabito, “Elman neural networks for characterizing voids in welded strips: A study,” Neural Comput. Appl., vol. 21, no. 5, pp. 869–875, Jul. 2012, doi: 10.1007/s00521-011-0609-3.

[41] V. N. Vapnik, “An overview of statistical learning theory,” IEEE Trans. Neural Netw., vol. 10, no. 5, pp. 988–999, 1999, doi: 10.1109/72.788640.

[42] X. Li, Y. Liu, and W. Xin, “Wind speed prediction based on genetic neural network,” in Proc. 4th IEEE Conf. Ind. Electron. Appl., May 2009, pp. 2448–2451, doi: 10.1109/ICIEA.2009.5138642.

[43] E. Valian, S. Tavakoli, S. Mohanna, and A. Haghi, “Improved cuckoo search for reliability optimization problems,” Comput. Ind. Eng., vol. 64, no. 1, pp. 459–468, Jan. 2013, doi: 10.1016/j.cie.2012.07.011.

[44] X. Chen, J. Wang, D. Sun, and J. Liang, “A novel hybrid evolutionary algorithm based on PSO and AFSA for feedforward neural network training,” in Proc. 4th Int. Conf. Wireless Commun., Netw. Mobile Comput., Oct. 2008, doi: 10.1109/WiCom.2008.2518.

[45] X.-S. Yang, “Firefly algorithms for multimodal optimization,” in Stochastic Algorithms: Foundations and Applications (Lecture Notes in Computer Science: Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Berlin, Germany: Springer, 2009, doi: 10.1007/978-3-642-04944-6_14.

[46] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in Proc. IEEE Int. Conf. Neural Netw., vol. 4, pp. 1942–1948, 1995.

[47] J. H. Holland, “Genetic algorithms,” Sci. Amer., vol. 267, no. 1, pp. 66–72, Jul. 1992, doi: 10.1038/scientificamerican0792-66.

[48] L. Zhang, Y. Dong, and J. Wang, “Wind speed forecasting using a two-stage forecasting system with an error correcting and nonlinear ensemble strategy,” IEEE Access, vol. 7, pp. 176000–176023, 2019, doi: 10.1109/ACCESS.2019.2957174.

[49] X. Niu and J. Wang, “A combined model based on data preprocessing strategy and multi-objective optimization algorithm for short-term wind speed forecasting,” Appl. Energy, vol. 241, pp. 519–539, May 2019, doi: 10.1016/j.apenergy.2019.03.097.

TONGJI GUO is currently pursuing the Ph.D. degree with the Jiangxi University of Finance and Economics, Nanchang, China. His research interests are energy economic analysis and health statistics.

LIFANG ZHANG received the bachelor’s degree in economics from Qingdao University, Qingdao, China, in 2018. She is currently pursuing the master’s degree with the Dongbei University of Finance and Economics, Dalian, China. Her main subjects of interest concerns economic loss assessment of haze pollution in big data environment, wind speed forecasting, data mining, and machine learning, which tend to be multidisciplinary in approach. She has published two academic articles in SCI retrieval.

ZHENKUN LIU received the bachelor’s degree in science from Shandong Technology and Business University, Yantai, China, in 2018. He is currently pursuing the master’s degree with the Dongbei University of Finance and Economics, Dalian, China. His research interests include time series analysis, applied statistics, artificial intelligence, high-dimensional data analysis, energy prediction theory, and method. He has published two academic articles in SCI retrieval, which has been published in top international journals.

JIANZHOU WANG is currently working as a Professor at the Dongbei University of Finance and Economics, Dalian, China, and a Doctoral Supervisor of big data at the School of Engineering and Information Technology, University of Technology Sydney, Sydney. He has published more than 150 academic articles in his academic career, and in the past five years, he has published 100 articles in SSCI and SCI retrieval, most of which have been published in top international academic journals. His research interests include economic loss assessment of haze pollution in big data environment, big data mining, artificial intelligence prediction theory and method, energy economic analysis, financial mathematical model, and complexity analysis.

VOLUME 8, 2020