A Hybrid Short-Term Wind Power Prediction Model Combining Data Processing, Multiple Parameters Optimization and Multi-Intelligent Models Apportion Strategy

HUA YAN AND ZHENGPING WU
College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, China
Corresponding author: Zhengping Wu (huwzp@sina.com)

This work was supported by the National Natural Science Foundation of China under Grant 61871258.

ABSTRACT As the wind power dates have strong volatility and intermittency, it is difficult to meet the safety and stability of power system operation. To make full use of the characteristics of wind power data and further improve wind power forecasting performance, an innovative hybrid framework is proposed to response to the above challenges, of which the non-stationary weakening technique, sample entropy (SE)-based prediction model allocation, optimized reduced kernel extreme learning machine (RKELM) and a novel deep learning network are integrated befittingly. Initially, the original wind power sequence is decomposed into multiple sub-sequences by empirical wavelet transform (EWT) adaptively, after which SE is employed to analyse the high and low components of sub-sequences. Then, a simplified gate recurrent unit network (SGRU) to achieve the prediction of the component with low SE-value. Besides, for the components with high SE-values, the multiple parameters optimization is implemented for RKELM based on slime mould algorithm (SMA), which possesses superior convergence speed and precision. Ultimately, the final prediction results can be obtained by accumulating the predicted values of the multiple models. In the experiment phase, two datasets are adopted to verify the predictive ability of the proposed hybrid EWT-SMA-RKELM-SGRU model, where the sufficient results indicate that the proposed model has a superior performance.

INDEX TERMS Short-term wind power forecasting, prediction model allocation, slime mould algorithm, multiple parameters optimization, simplified gate recurrent unit network.

I. INTRODUCTION

With striking aggravate of environment pollution, the gradual depletion of traditional energy, and a goal for zero-carbon emissions of government, people has been an increasing attention to the renewable energy [1]. In recent decades, wind power, as one of comprehensive technology involving environmental protection and sustainable development, has become remarkably significant as renewable energy comparing with other traditional energy sources, such as coal, petroleum and natural gas [2]. Generally, wind power operation possessing abundant uncertainties, which conclude intermittent factors of varied topography, atmospheric pressure, sun illumination and random effects of wind speed as well, has a great challenge. For those mentioned above, a precise short-term wind power forecasting undoubtedly plays a crucial part in aspects of effective dispatch plan and optimal unit combination scheduling. In reality, owing to the fact that data-driven forecasting approach has the ability of excavating the dynamic behavior of wind power development trends, stability and accuracy of prediction results are the important indicators for short-term wind power forecasting. Consequently, the development for accurate wind power model appears remarkably importance by degrees, thereby promoting economic benefit for wind power plants effectively and conspicuously. It is worth noting that as most significant factor of wind power, wind speed seriously and actually influences over the quality and stability of power system, which reveals usually characteristics of extreme abruptness [3]. Researchers have been gradually adopting...
measures a lot on prediction of wind power, owing to providing sensible suggestions for power system management, wind power generator turbines control and energy negotiations [4].

Wind power data changing with various factors, leading to dynamic features and time-varying information, is generally gathered as time-series, which may include the main trend, periodic trend and quasi-periodic trend implicitly [5]. Therefore, volatility of the wind power series impact undoubtedly the prediction effect of predict engine that could not get rid of influence for noises, which are implied in raw wind power series [6]. To promote performance of forecasting results, Scholars have pay attention to the combined prediction structures, in which the combination for a series of individual approaches is implemented. For time series prediction, models based on decomposition techniques take advantage frequently to solve the strong non-stationary and volatility of wind power time series [7], [8]. Besides, the hybrid or ensemble approaches mainly consist of two parts: pre-treatment and post-treatment respectively. The process of pre-treatment utilize decomposition method to decompose and reconstruct time-series dataset for further usage, such as empirical wavelet transform (EWT), variational mode decomposition (VMD), empirical mode decomposition (EMD), singular spectrum analysis (SSA) as well as their improved forms. Moreover, the process of post-treatment usually refer to prediction modeling of data via pre-treatment using artificial intelligence(AI) models such as support vector regression (SVR), extreme learning machine (ELM), and deep learning models. For example, researchers proposed a novel robust hybrid method, namely IVMD-SE-MCC-LSTM, for short-term wind power forecasting, whose raw datasets are decomposed and reconstructed by improved variational mode decomposition (IVMD) proposed in their paper, thus to group modeling employed Long Short-term Memory (LSTM) [9]. Analogously, Safari et al. [10] developed a new multi-step framework for short-term wind power prediction applying multi-scale singular spectrum analysis (MSSSA) and least squares support vector machine (LSSVM), where the results demonstrate better performance in of the proposed model in the aspect of computational burden. Scholars have adopted long-short term memory network to predict oil production, where network architecture and window size are optimized by genetic algorithm, fed oil production time serials data of SL and JD oilfields as input, before which ensemble empirical mode decomposition (EEMD) with dynamic time warping (DTW) based decide the model subsequences through mean value comparison [11]. A time domain signal decomposing algorithm proposed by Huang et al. [12], reconstructs several intrinsic mode functions (IMFs), namely locally narrow band components, of which modal aliasing is inherent qualities yet [13]. Dragomiretskiy and Zosso [14] developed a novel decomposition approach, referred to as VMD, to turn decomposition for signal into variational solution problem and, transform signal into ensemble of band-limited modes. In spite of which the VMD has such advantages and been applied in many fields, such a decomposition method still not extricate itself from some shortcomings that the punishment factor and number of modes are not non-automatically determined [15]. SSA, firstly presented by Bonizzi et al. [16], has been widespread utilized in various research fields, as a well-known data analysis instrument for decomposition. However, there are comparatively low quality and scarce data that it is not well received frequently yet in some application scenarios [17]. Besides, embedding dimension should be selected carefully and experimentally in construction process of SSA, where the improper dimensions could bring to the effect of adverse decomposition results [18]. In contrast, EWT, a novel approach for prepossessing data with combined characteristics of adaptability of EMD and formalism of wavelet transform (WT), possesses an admirable advantage that is to remedy technology for WT’s demand of wavelet basis and parameters [19].

The prevailing wind power prediction approaches can be briefly divided into four categories, namely physical modeling, statistical modeling, intelligent modeling, and hybrid structure modeling. As to physical models, numerical weather prediction (NWP), which could generate forecasting results of long-term and large-scale and based on ensemble of sigmoidal power function, was developed by Pelikan et al. [20] to calculate week-ahead or day-ahead forecasting [21]. Nevertheless, technologies with physical-based requires typically fruitful characteristic value, such as atmospheric density, wind farm area topography and surface roughness of parts [9], whose applications are limited universally in short-term wind power forecasting [22], [23]. Statistical modeling methods, roughly including ARMA (autoregressive moving average) [24], [25], ARIMA (autoregressive integrated moving average) [26], [27], and f-ARIMA (fractional ARIMA) [28], [29], could much fit for prediction of short-term wind power, as considering the mapping relationship within the input and target variables [30], by which nonlinear and irregular features of wind power series could not be tackle with because of the prior assumption linear forms well [31]. In recent years, machine learning and deep learning approaches [32], developed widely in the field of prediction, could achieve high-predictions for short-term wind power series that possess non-stationary and nonlinear characteristics ordinarily. As a shallow learning model, extreme learning machine (ELM), a popular member of the machine learning family, is usually reported in literature for the research of prediction, owing to its preeminent capacity of adaptively learning nonlinear features. For instance, Majumder et al. [33] utilized ELM to obtain reliable and accurate solar power forecasting for 5-minute interval and 30-minute interval serials data respectively, whose final experiment results achieved performance of 4.23% and 5.26% respectively according to index.

Nevertheless, individual model or its improved type, namely non-combination mechanism, has not enough ability to capture the specific feature of each component, since there are distinguishing volatility and complex relevance in every components. In view of the above, approach of grouping
modeling is struggling to retain components characteristics and adapt to their complexity that is remarkably different for individuals, where the determination of complexity for each component is formulated by instrument quantified the tightness of wind power time series data. For the purpose of recession on forecasting difficulty of each component, which still remains disordered and non-stationary information, various subsequences should be categorized into several groups with similar feature using method that usually utilizes entropy value [34], [35]. As a tool to describe complexity of time series, developed from approximate entropy (AE), sample entropy (SE) has been widely applied in a myriad of academic fields [36]. For instance, a diagnosis method was present for diagnosis and early battery faults prediction, in which SE was established to estimate battery health [37]. Chen et al. [38] proposed newly hybrid runoff forecasting approach, where the hybrid model adopting sample entropy to measure the complexity of the components is proved to possess the benefits of reducing the time consumption as well as ensuring the prediction accuracy.

Thanks to the less intervention and extremely learning rate, extreme learning machine neural network [39], is more efficient and simpler than other neural networks. However, when the dataset is disturbed by outliers and noise in it, the robustness of ELM is decayed dramatically [40]. Hence a neoteric reduced kernel extreme learning machine (RKELM) is employed to build up the recognition model in initial offline phase, which is designed to enhance the abilities of excellent generalization and rapid learning speed. Additionally, it has been proved in experiments that the good generalization of such a machine learning method is mainly caused by the quality of the sequence rather than the length of the sequence [41]. For this point, RKELM is applied to forecast the relatively tight and complex subsequences for promotion of generalization ability of such sequences, as far as other subsets, at the same time, simplified gated recurrent unit network (SGRU), developed from recurrent unit network (GRU), is taken advantage of prediction in parallel.

Due to considering the characteristics of wind power time series data, a novel prediction structure, which involves EWT, sample entropy calculation, reduced kernel extreme learning machine, slime mould algorithm optimization algorithm and simplified gated recurrent unit network, is built in this article. For one thing, decomposition that is achieved by EWT with other kinds of wavelet transform method, EWT could analyze adaptive decomposition of signal due to utilizing Fourier spectrum. To achieve the aim above, the EWT employs several basic function, which mainly including empirical wavelet function and empirical scale function. In each frequency band, the empirical wavelet function and empirical scale function. To achieve the aim above, the EWT employs several basic function, which mainly including empirical wavelet function and empirical scale function. In each frequency band, the empirical wavelet function usually indicates the band-pass filters, while the empirical scale function defines the low-pass filters. The steps of EWT method can be summarized simply are shown as follows:

1) Owing to the insensitivity for parameters of EWT, complex time series data is decomposed and reconstructed by EWT method in order to weaken the non-stability of raw data.
2) For the better stability and accuracy of the combined wind power prediction model concurrently, the parameters of RKELM and the matrix shape of decomposed components are both optimized to be decided by SMA optimization.
3) A novel deep learning network layer that is improved by simplifying GRU, namely SGRU, possessing the characteristics of broadly similar computational performance and even less computing time, is designed in this study, after which forecasting indicators of cases has confirmed it actually in the tail of the article.
4) An hybrid prediction method based on EWT-RKELM-SGRU with SMA based consists tow branches, SGRU modeling and RKELM modeling, into which the subsequent is brought ultimately via judging from the largeness after target calculation by sample entropy. That is one with lower value belongs to the SGRU modeling group, while the others belong to the RKELM modeling group.

In addition, the theories foundation of EWT, SE, SMA, RKELM and SGRU are provided thoroughly in Section II. Section III describes the method and steps of prediction of wind power. In Section IV the research dataset is described briefly, after which experimental procedure and comparative assessments are demonstrated, respectively. Lastly, the conclusions are discussed in Section V.

II. BACKGROUND KNOWLEDGE
A. EMPIRICAL WAVELET TRANSFORM
The EWT was proposed by Gilles firstly in 2013 as a novel composition method for processing series data. Compared with other kinds of wavelet transform method, EWT could analyze adaptive decomposition of signal due to utilizing Fourier spectrum. To achieve the aim above, the EWT employs several basic function, which mainly including empirical wavelet function and empirical scale function. In each frequency band, the empirical wavelet function usually indicates the band-pass filters, while the empirical scale function defines the low-pass filters. The steps of EWT method can be summarized simply are shown as follows:

Step 1: The spectrum diagram of the signal is obtained by Fourier transform of it.
Step 2: Fourier spectrum is divided into $N$ consecutive segments firstly. Furthermore, segments could be sorted order by descend. Actually, we should retain $M$ maximums if $M$ is smaller than $N$, otherwise $N - 1$ maximum are kept.
Step 3: What was retained in step 2 builds the empirical scaling function and the empirical wavelets function according to (1) and (2), as shown at the bottom of the next page, respectively.
Step 4: The approximation parameters and the detailed parameters are calculated according to (3) and (4), respectively.

\[ \omega_f^j(0, t) = \langle f, \phi_1 \rangle \]

\[ = \int f(\tau)\phi_1(\tau - t) d\tau \]

\[ = \left[ \hat{f}(\omega)\hat{\phi}_1(\omega) \right]^\vee \]  

\[ \omega_f^j(n, t) = \langle f, \psi_1 \rangle \]

\[ = \int f(\tau)\psi_1(\tau - t) d\tau \]

\[ = \left[ \hat{f}(\omega)\hat{\psi}_1(\omega) \right]^\vee \]  

(3)  

(4)  

Step 5: The signal, we need finally, is reconstructed according to (5). where \( \phi_1 \) indicates empirical scale function, while \( W_f^j(0, \omega) \) and \( W_f^j(n, \omega) \) are transformed from \( W_f^j(0, t) \) and \( W_f^j(n, t) \) using the method of Fourier transform, respectively.

\[ f(t) = \omega_f^e(0, t) \]  

(5)  

B. SAMPLE ENTROPY

As one category of measuring the complexity of a dataset belonging time series, Sample Entropy has been utilized in various fields, including signal analysis, mechanical fault diagnosis and meterology research, which could reflect signal characteristics in view of frequency components of the time series. Compared with other entropies, such as AE, SE possesses not only a better performance for time series than AE in aspects of data length, but also consistent measurements, although which is a kind of development from AE. For the original bounded time series \( \{ x_i \mid i = 1, 2, \ldots, N \} \), the specific implementation steps of SE is exhibited briefly below:

Step1: Reconstruct the given time series \( \{ x_i \mid i = 1, 2, \ldots, N \} \) into a novel \( m \)-dimensional matrix \( X_i = [x_i, x_{i+1}, \ldots, x_{i+m-1}] \), where \( i = 1, 2, \ldots, N - m + 1 \) and \( m \) refers to the window length.

Step2: Calculate Chebyshev distance for each component, which is defined as \( D_m(X_i, X_j) = \max_{0 \leq m-1} |x_i+k - x_j+k| \).

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

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

Step3: Calculate the ratio \( B_m^r(r) \) corresponding to the total number of \( D_m(X_i, X_j) < r \) for the \( i \)-th vector, afterwards computing the mean of \( B_i^r(r) \), defined as \( B^m(r) \).

Step4: Deduce \( B^{m+1}(r) \) by iterating from Step 1 to Step 3, after selecting a new window length as \( B^m(r) \).

Step5: For the given bounded time series, the result of SE could be formulated as follows finally:

\[ SE(N, m, r) = -\ln \left( \frac{B^{m+1}(r)}{B^m(r)} \right) \]  

where the calculated SE value is defined as \( SE(N, m, r) \), \( N \) refers the length of series, \( m \) represents the window length and \( r \) is the similarity tolerance which is usually set in the range of \([0.1SD, 0.25SD]\) (SD indicates the standard deviation of the time series) [42]. In this study, the value of \( r \) is chosen as 0.1SD experimental.

C. SLIME MOULD ALGORITHM

Slime mould algorithm (SMA) is a novel nature behaviours-based stochastic optimization algorithm, whose imitation mechanisms are developed based on the oscillation mode of slime mould [43]. Among the optimization processes, the adaptive weights are introduced to mimic the positive and negative feedback process of slime mold propagation wave based on biological oscillator. Here, the central features of SMA mentioned above are exhibited as below:

1) APPROACH FOOD

The following formula is developed to construct the approaching mechanism of slime mould:

\[ X_{t+1} = \begin{cases} 
X_b(t) + v_b. (W. X_A(t) - X_B(t)) & r < p \\
v_c. X_t & r \geq p 
\end{cases} \]  

(7)  

where \( v_b \) is a parameter with a range of \([-a, a]\), \( v_c \) decreases linearly from one to zero, \( t \) represents the \( t \)-th iteration, \( X \) represents the location of slime mould, \( X_b \) represents the individual location with the highest odor concentration currently found, \( X_A \) and \( X_B \) are the random individuals selected from the population, \( W \) indicates the weight vector of slime
moulds, $S(i)$ represents the fitness value of $X$, $DF$ represents the best fitness obtained in all iterations. The detailed descriptions of the above arguments can be found in [43].

2) WRAP FOOD
The mathematical formula for wrapping food that representing the updating process of slime mould is as follows:

$$
X^* = \begin{cases} 
\text{rand}(UB - LB) + LB & \text{rand} < z \\
X_b(t) + v_b(WX_a(t) - X_b(t)) & r < p \\
v_cX(t) & r \geq p
\end{cases}
$$

where $LB$ and $UB$ denote the lower and upper boundaries of the search range, rand and $r$ denote the random value in [0, 1].

3) OSCILLATION
As mentioned in II-C1, $v_b$ oscillates randomly between $[-a, a]$, which will gradually decay to zero as the iterations increase. The value of $v_c$ oscillates between [-1, 1] and tends to zero eventually.

D. REDUCED KERNEL EXTREME LEARNING MACHINE
As a frequently-used learning method, ELM is put forward by Huang et al. in 2006 [44], which mainly was designed for single layer feed forward neural networks (SLFNs). Although ELM holds the ability of outstanding generalization and rapid learning [45], it is well known that the ELM has the drawback of tackling with morbid data not well [46]. For this purpose, Kernel Extreme Learning Machine (KELM) was developed, which, however, remains the fact that performance is barely satisfactory in real-time computation.

In order to remedy a defect of traditional KELM(kernel extreme learning machine) in aspect of extremely large time and space complexities in real-time computing, an advanced method, reduced kernel based extreme learning machine, is raised by Deng et al. [47]. In RKELM, $n$ samples has been selected from the raw dataset of $N$ samples, where $n \ll N$. Thus, both the complexities of time and space are reduced through the method mentioned above. The calculation formula is given as:

The algorithm of RKELM could be described briefly as following:

a. Select subset of $n$ samples from original dataset possessing $N$ samples, where $n \ll N$.
b. Rebuild the reduced kernel matrix with the shape of $N \times n$.
c. Calculate the values of final prediction which tends to better generalization performance via formula as follow:

$$
\mathbf{x} = \mathbf{K} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{K} \begin{bmatrix} 1 \\ \vdots \\ \frac{1}{C} \end{bmatrix} + \mathbf{K}(X, \tilde{X})^T \mathbf{K}(X, \tilde{X})^{-1} \mathbf{K}(X, \tilde{X})^T \mathbf{T} \tag{9}
$$

where $K$ refers to the matrix of rectangular kernel.

E. SIMPLIFIED GATED RECURRENT UNIT NETWORK
Similar to LSTM, a widely used typical cyclic neural networks and effectively solving problem of the disappearing gradient when time lags are continuing, GRU proposed as an development structure of LSTM [48], consists of update gate and reset gate, which leads to lower computing consumption than LSTM and could ignore the loss of performance for modeling, and its inner structure is shown in Figure 1. Taking into account the potentially exceedingly long computing times in practical modelling, a simplified version developed from GRU is proposed in this study, namely SGRU, among which only two activation functions and four gates are retained, and the internal data flow characteristics are basically Inherited from GRU in form. Figure 2 illustrates the internal structural composition of SGRU clearly, and also its mathematical expression is as follows:

$$
z_t = \sigma (W_z \cdot [t_{t-1}, x_t])
\tilde{t}_t = \tanh (W_t \cdot [(1 - z_t) \odot t_{t-1}, x_t])
t_t = (1 - z_t) \odot t_{t-1} + \tilde{t}_t \odot z_t \tag{10}
$$

where $x_t$, as output of the previous layer information, is passed into the structure, at the same time, $t_t$ that contains
the implied information of previous layer is also passed. After a gated calculation by sigmoid function, both $x_i$ and $t_i$ spliced together are sent into tanh activation function in order to scale the value into the range of $[-1, 1]$. Finally, output with implied information, namely $t_i$, is sent off to the next one, thus memorizing the current information. And sigmoid function and the tanh function are defined below respectively:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (11)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (12)$$
As can be seen from the formula above, the output of this structure can be obtained via relatively fewer calculations, which considerably reduces the cost of operation, in addition to guaranteeing the transfer of implied information that is involved in current layer.

### III. SHORT-TERM WIND POWER FORECASTING MODEL BASED ON EWT, SMA, AND RKELM MULTIPLE PARAMETERS OPTIMIZATION

#### A. MULTIPLE PARAMETERS OPTIMIZATION BASED ON SLIME MOULD ALGORITHM

There exist two dominating factors affecting the forecasting performance of the developed combined system, i.e., the rationality of the refactor form for input matrixes and the parameters of the forecasting model. In the previous investigations applying the decomposition methods and parameter optimization, the input matrixes for the predictor are fixed by the subjective experience of scholars, where the different components equipped with same reconstruction arguments should be reconsidered in accordance to the diversities of the input structures. For this purpose, phase space reconstruction, which transforms the unidimensional series into the high-dimensional feature space considering the time delay and embedded dimension simultaneously, is employed to construct the input matrixes appropriately. However, the optimization for phase space reconstruction and the predictors are generally implemented separately, which isolates the correlation between the arguments of the above modules as well as increasing the computational cost significantly. Focusing on improving the optimization efficiency and forecasting accuracy in this study, a multiple parameters optimization strategy is proposed to determine the optimal shape of input matrix and the appropriate arguments of RKELM.

#### B. MODEL ALLOCATION FOR VARIOUS COMPONENTS

The application of decomposition methods has been investigated in the field of series prediction widely. It is worth noting that a fixed forecasting model is employed for each decomposed component in the previous literature, despite some scholars pay attention to the adaptive parameter optimization for the prediction model on each component. However, the computational resource of the whole combined model will significantly increase with the number of components increase. Therefore, a model allocation strategy considering the balance of forecasting accuracy and computational complexity is developed in this study. Specifically, we apply SE to measure the complexity scale of each component, which is aimed at the partition of the tendency item and the random ones. It is worth mentioning that the series will be difficult to predict with the corresponding increase of SE value. Hence, the newly constructed deep network, namely SGRU, is adopted to predict the tendency component with minimum entropy value, while the random items with larger entropy values are predicted by RKELM with above multiple parameters optimization. Applying the model allocation for the components based on SE, the training time of the entire combined model can be significantly reduced without significantly forecasting capability decrease.

#### C. ENTIRE STRUCTURE OF THE PROPOSED MODEL

In summary, the detailed procedures of the proposed model adopting the aforementioned modules, i.e., EWT-based feature selection, SMA-based multiple parameter optimization as well as SE-based model allocation, can be summed up as follows:

- **Step 1**: Employ EWT to extract the tendency item and the random ones with appropriate parameters.
- **Step 2**: Determine the prediction model for each model in accordance to the corresponding SE value of each component.
- **Step 3**: Train the SGRU model for the determined tendency component, and train the optimal RKELM model for the random components using the SMA-based multiple parameter optimization.
- **Step 4**: Sum all the predicted results up to obtain the ultimate prediction values of training set.

Additionally, the entire flowchart of the composite model consisting of above supplemental strategies is exhibited in Figure 3.

### IV. EXPERIMENTAL DESIGN

#### A. DATA DESCRIPTION

To quantitatively testify the performance of the proposed composite forecasting model, two practical wind power generation datasets are collected from the wind farm located in Sotavento Galicia. Meanwhile, the consecutive values...
monitored from Jan. 18, 2020 to Jan. 29, 2020 and Jan. 30, 2020 to Feb. 10, 2020 are considered in this study, which is measured with time period of ten minutes as well as consisting of 1704 samples. The diagrams of above mentioned two wind power sequential series are depicted in Figures 4 and 5, respectively, which demonstrates significant intermittent, randomness and volatility. Besides, the datasets are partitioned in to three parts, namely training set, verification set and testing set, to achieve the model construction, optimal parameter selection and performance evaluation, respectively, where the corresponding partition ratio is 3:1:1.
B. EXPERIMENTAL DESCRIPTION

In this section, the particular description of the contrastive models and the employed performance evaluation indicators will be introduced in detail.

1) CONTRASTIVE MODELS

In this study, a composite forecasting model integrating decomposition approach, prediction model allocation strategy and SMA-based multiple parameters optimization is proposed to achieve accuracy prediction results. To comprehensively analyzed the contribution of each developed module, a series of single forecasting models and combined models are carried out for contrastive discussion. Among the single models, SVR, RKELM, BPNN, GRU and SGRU are applied as the benchmark models to reveal the difference of each single models as well as the improvement of the newly developed SGRU model. On the basis of RKELM, CEEMDAN and EWT are combined with RKELM successively, thus illustrating the performance improvements achieved by CEEMDAN and EWT. Moreover, EWT is further fused with SGRU to compare with EWT-RKELM, with which we intend to demonstrate the predictive performance robustness of RKELM and SGRU for various subseries.

2) PERFORMANCE EVALUATION

To quantitatively analyse the improvements achieved by different modification strategies as well as revealing the differences among the proposed method and the contrastive ones, the indicators including root-mean-square error (RMSE), mean absolute error (MAE) [49], grey relational analysis (GRA) [50] and Diebold-Mariano (DM) test [51] are employed in this study. Among the indexes mentioned above, RMSE, MAE, MSE and GRA are selected to reveal the forecasting accuracy of each model, where MAE can prevent errors from canceling each other out, reflecting the actual situation of the error. Additionally, the DM test can demonstrate the significance superiority of the proposed forecasting approach compared with the contrastive models. The detailed definition of RMSE, MAE, MSE are depicted as follows:

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \quad (13)
\]

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \quad (14)
\]

where \(y_i\) and \(\hat{y}_i\) denote the actual value and the predicted value, respectively. In addition, the specific formulation for GRA is sequential listed as: (1) normalized the reference and comparison sequences; (2) calculate the correlation coefficient among the above two normalized sequences at each time; and (3) formula the mean value of all the correlation...
coefficients to obtain the grey correlation degree. Furthermore, the specific calculation for the DM test is presented as below:

$$DM = \frac{1}{N} \sum_{i=1}^{N} \sqrt{\frac{(L(error_1^i) - L(error_2^i))^2}{s^2/N}}$$  \hspace{1cm} (15)$$

where $s^2$ denotes the estimation for the variance of $d_h = L(error_1^i) - L(error_2^i)$, where $L$ indicates the loss function adopted to evaluate the prediction accuracy of two models. For the given confidence level $\alpha$, the performance difference between the proposed approach and the comparison ones can be claimed as less applying null hypothesis $H_0$, while $H_1$ is the inverse hypothesis of $H_0$. Meanwhile, $H_0$ will be rejected in the condition of that the DM values are outside the interval of $[-Z_{\frac{\alpha}{2}}, Z_{\frac{\alpha}{2}}]$.

C. EXPERIMENTAL RESULTS AND ANALYSIS

The components obtained by performing EWT on the collected two datasets are depicted in Figures 6 and 7, where the corresponding variation of all the SE values are expressed beside the curves. It can be observed in conjunction with Figures 6 and 7 that the components with stable tendency conditions have the minimum SE values, indicating that such the components can be effectively predicted by the developed deep learning network. Subsequently, the blended model based on RKELM and the SMA-based multiple parameters optimization is performed for the remaining components.
possessing larger SE values, where the obtained optimal arguments for each component are expressed in Table 1.

**TABLE 1.** Obtained parameters for components with high SE values.

| Cases   | Components | d | $\tau$ | g | C  |
|---------|------------|---|--------|---|----|
| Case 1  | $r_1$      | 7 | 1      | 63.35 | 1000 |
|         | $r_2$      | 7 | 1      | 0.78  | 1000 |
| Case 2  | $r_1$      | 9 | 10     | 1000  | 1000 |
|         | $r_2$      | 7 | 1      | 1.02  | 1000 |

To verify the superiority and availability of the proposed model, representative single forecasting models and decomposition-based combined models are chosen for comparison. The forecasting results of different models among two cases are shown in TABLE 2 and TABLE 3 respectively. As to single forecasting model, a newly developed deep learning approach regarded as SGRU can achieve better forecasting results in all cases contrasted with SVR, RKELM, BPNN and GRU. Taking case 1 as an instance, the metrics of RMSE and MAE obtained by SGRU are 66.4231 (KW) and 54.0535 (KW) respectively, which has decreased 52.45% and 45.45% compared with the average value of remaining single models. The metric GRA of SGRU has increased 7.95% and 7.13% contrasted with BPNN and GRU respectively. It can be concluded that SGRU possesses better performance on excavating potential information of raw data. Thus, SGRU is proposed to predict wind power in this article.

As to decomposition-based combined models, CEEMDAN-RKELM, EWT-RKELM and EWT-SGRU can achieve better performance compared with RKELM and SGRU in all cases, which attributes that CEEMDAN and EWT can reduce the non-stationarity of raw wind power.
For case 2, the decreasing rates of RMSE and MAE between EWT-RKELM and RKELM are 50.77% and 50.61% respectively meanwhile the metric GRA of EWT-SGRU has improved 8.74% contrasted with SGRU. Additionally, EWT-RKELM can achieve better forecasting performance compared with CEEMDAN-RKELM, where the metrics RMSE and MAE are decreased by 43.6% and 33.14% respectively meanwhile the metric GRA is increased by 9.03%
in case 2. Accordingly, EWT, in this article, is selected as preprocessing approach, which possesses better capability in decomposition.

Additionally, the forecasting results of the proposed model are superior to EWT-RKELM and EWT-SGRU, where the descent rates of the metrics RMSE and MAE between the proposed model and EWT-SGRU are 7.89% and 2.78% respectively in case 2. The comparative experiment demonstrates that the forecasting accuracy can be improved when RKELM and SGRU are employed to analyze high frequency series and low frequency series severally. Therefore, the proposed model can achieve satisfactory result in the field of wind power forecasting.

In this section, the prediction result visualizations of contrastive models are demonstrated and discussed separately. The comparison between the proposed model and other models on metrics MSE(KW), RMSE(KW), MAE(KW) and GRA are displayed in Figures 8 - 13. It can be seen from Figures 8, 9 and Figures 11 and 12 that the proposed model possesses better forecasting results and shorter error blocks compared with remaining models. As to Figures 10 and 13, the maximum value of metric GRA are made of the proposed model among two cases, which further testifies that the proposed model can obtain higher forecasting accuracy. Subsequently, the fitting curves of models belonging to single one and combined one with decomposition-based are displayed in Figures 14, 15 and 16, 17, respectively. As can be observed from Figures 14 and 16, the fitting curves of the single models diverges from the actual values, while the fitting curves of the decomposition-based combined models are relatively close the actual value in Figures 15 and 17. This reason is that decomposition-based approach can weaken
the non-stationarity of raw wind power series. Moreover, the fitting curves of errors distribution for different models are shown in Figures 18 and 19 intuitively. Taking case 1 as an example, the proposed model possesses the lowest deviation compared with remaining models. Ultimately, Figure 20 has been demonstrated obviously that the scatter plots illustrate the correlation distribution of each model. At the same time, it also could be observed directly from Figure 20 that the model proposed is uniformly distributed over the regression line. Similar conclusions can also be observed from Figure 21. Based on above discussion, the superiority of higher forecasting accuracy obtained by the proposed model can be verified effectively. This reason is that decomposition-based approach can weaken the non-stationarity of raw wind power series.

V. CONCLUSION

In this study, a novel structure for wind power forecasting, based on the EWT decomposition, the SMA algorithm, the Sample Entropy theory, the RKELM and the SGRU network, is present for wind power forecasting. The EWT method, as a pre-dealing, is utilized to decompose raw wind power forecasting series data into three sub-sequences, which are generally regarded as primary trend, periodic trend and residual trend. The entropy value of those sub-sequences, which possess different complexities, could be calculated by taking advantage of Sample Entropy theory. As one part of prediction, sub-sequence, whose value of entropy is the lowest than others, is substituted into SGRU, a new deep learning structure improved from GRU, for prediction modeling. Meanwhile, RKELM network, combined with SMA optimization algorithm based, is exploited to determine the optimal model value of $d$, $\tau$, $g$ and $C$ for the remaining sub-sequences, after which the two of the original sub-sequences, with optimal parameters, are fed into RKELM network in order to build two prediction model as well. Then, the prediction results are finally obtained via merge all of the sub-sequences prediction data, which are produced from SGRU and RKELM, respectively. In summary, the experimental results obtained by the proposed hybrid prediction model performed on two short-time wind power serial datasets in various periods illustrate that the the proposed model achieves significantly improvements compared with the relevant contrast models.

REFERENCES

[1] L. Xi, J. Chen, Y. Huang, Y. Xu, L. Liu, Y. Zhou, and Y. Li, “Smart generation control based on multi-agent reinforcement learning with the idea of the time tunnel,” Energy, vol. 153, pp. 977–987, Jun. 2018.
[2] D. Xu, Q. Wu, B. Zhou, C. Li, L. Bai, and S. Huang, “Distributed multi-energy operation of coupled electricity, heating, and natural gas networks,” IEEE Trans. Sustain. Energy, vol. 11, no. 4, pp. 2457–2469, Oct. 2020.
[3] H. Liu and C. Chen, “Multi-objective data-ensemble wind speed forecasting model with stacked sparse autoencoder and adaptive decomposition-based error correction,” Appl. Energy, vol. 254, Nov. 2019, Nov. 113686.
[4] B. C. Ummels, M. Gibescu, E. Pelgrum, W. L. Kling, and A. J. Brand, “Impacts of wind power on thermal generation unit commitment and dispatch,” IEEE Trans. Energy Convers., vol. 22, no. 1, pp. 44–51, Mar. 2007.
[5] X. Mi and S. Zhao, “Wind speed prediction based on singular spectrum analysis and neural network structural learning,” Energy Convers. Manage., vol. 216, Jul. 2020, Art. no. 112956.
[6] Z. Sun and M. Zhao, “Short-term wind power forecasting based on VMD decomposition, ConvLSTM networks and error analysis,” IEEE Access, vol. 8, pp. 134422–134434, 2020.
[7] H. Liu, H.-Q. Tian, D.-F. Pan, and Y.-F. Li, “Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks,” Appl. Energy, vol. 107, pp. 191–208, Jul. 2013.
[8] A. Aghajani, R. Kazemzadeh, and A. Ebrahimi, “A novel hybrid approach for predicting wind farm power production based on wavelet transform, hybrid neural networks and imperialist competitive algorithm,” Energy Convers. Manage., vol. 121, pp. 232–240, Aug. 2016.
[9] J. Duan, P. Wang, W. Ma, X. Tian, S. Fang, Y. Cheng, and H. Liu, “Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and coренtropy long short-term memory neural network,” Energy, vol. 214, Jan. 2021, Art. no. 118980.
[10] N. Safari, C. Y. Chung, and G. C. D. Price, “Novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis,” IEEE Trans. Power Syst., vol. 33, no. 1, pp. 590–601, Jan. 2018.
[11] W. Liu, W. D. Liu, and J. Gu, “Forecasting oil production using ensemble empirical model decomposition based long short-term memory neural network,” J. Petroleum Sci. Eng., vol. 189, Jun. 2020, Art. no. 107013.
[12] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis,” Proc. R. Soc. London, A, Math., Phys. Eng. Sci., vol. 454, no. 1971, pp. 903–905, Mar. 1998.
[13] B. Ma and T. Zhang, “Single-channel blind source separation for vibration signals based on TVF-EMD and improved SCA,” IET Signal Process., vol. 14, no. 4, pp. 259–268, Jun. 2020.
[14] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” IEEE Trans. Signal Process., vol. 62, no. 3, pp. 531–544, Feb. 2014.
[15] I. Li, X. Yao, H. Wang, and I. Zhang, “Periodic impulses extraction based on improved adaptive VMD and sparse code shrinkage denoising and its application in rotating machinery fault diagnosis,” Mech. Syst. Signal Process., vol. 126, pp. 568–589, Jul. 2019.
[16] P. Bonizzi, J. M. H. Kareel, O. Meste, and R. L. M. Peeters, “Singular spectrum decomposition: A new method for time series decomposition,” Adv. Adapt. Data Anal., vol. 6, no. 4, Oct. 2014, Art. no. 145011.
[17] M. I. Gursoy, S. V. Usun, and A. S. Yilmaz, “An efficient DWT and EWT feature extraction methods for classification of real data PQ disturbances,” Uluslararası Muhendislik Arastirma ve Gelisimci Dergisi, vol. 10, no. 1, pp. 158–171, Jan. 2018.
[18] E. Pelikan, K. Eben, J. Resler, P. Jurus, P. Krc, M. Brabec, T. Brabec, and P. Musilek, “Wind power forecasting by an empirical model using NWP outputs,” in Proc. 9th Int. Conf. Environ. Elect. Eng., May 2010, pp. 45–48.
[19] C. Feng, M. Cui, B.-M. Hodge, and J. Zhang, “A data-driven multi-model methodology with deep feature selection for short-term wind forecasting,” Appl. Energy, vol. 190, pp. 1245–1257, Mar. 2017.
[20] X. Shi, X. Lei, Q. Huang, S. Huang, K. Ren, and Y. Hu, “Hourly day-ahead wind power prediction using the hybrid model of variational model decomposition and long short-term memory,” Energies, vol. 11, no. 11, pp. 3227, Nov. 2018.
[21] J. Zhang, J. Yan, D. Infeld, Y. Liu, and E.-S. Lien, “Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model,” Appl. Energy, vol. 241, pp. 229–244, May 2019.
[22] J. Shi, X. Qu, and S. Zeng, “Short-term wind power generation forecasting: Direct versus indirect arima-based approaches,” Int. J. Green Energy, vol. 8, no. 1, pp. 100–112, Feb. 2011.
[23] Y. Wang, D. Wang, and Y. Tang, “Clustered hybrid wind power prediction model based on ARMA, PSO-SVM, and clustering methods,” IEEE Access, vol. 8, pp. 17071–17079, 2020.
[26] S. Singh, K. S. Parmar, J. Kumar, and S. J. S. Makkhan, “Development of new hybrid model of discrete wavelet decomposition and autoregressive integrated moving average (ARIMA) models in application to one month forecast the casualties cases of COVID-19,” Chaos, Solitons Fractals, vol. 135, Jun. 2020, Art. no. 109866.

[27] H. Yu, L. J. Ming, R. Sumei, and Z. Shuping, “A hybrid model for financial time series forecasting—Integration of EWT, ARIMA with the improved ABC optimized ELM,” IEEE Access, vol. 8, pp. 84501–84518, 2020.

[28] R. G. Kavasseri and K. Seetharaman, “Day-ahead wind speed forecasting using f-ARIMA models,” Renew. Energy, vol. 34, no. 5, pp. 1388–1393, May 2009.

[29] A. Biernacki, “Improving quality of adaptive video by traffic prediction with (P)ARIMA models,” J. Commun. Netw., vol. 19, no. 5, pp. 521–530, Oct. 2017.

[30] X. Mi, H. Liu, and Y. Li, “Wind speed prediction model using singular spectrum analysis, empirical mode decomposition and convolutional support vector machine,” Energy Convers. Manage., vol. 180, pp. 196–205, Jan. 2019.

[31] J. Song, J. Wang, and H. Lu, “A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting,” Appl. Energy, vol. 215, pp. 643–658, Apr. 2018.

[32] H. Wang, Y. Liu, B. Zhou, C. Li, G. Cao, N. Voropai, and E. Barakhatenko, “Taxonomy research of artificial intelligence for deterministic solar power forecasting,” Energy Convers. Manage., vol. 214, Jun. 2020, Art. no. 112909.

[33] I. Majumder, M. K. Behera, and N. Nayak, “Solar power forecasting using a hybrid EMD-ELM method,” in Proc. Int. Conf. Circuit, Power Comput. Technol. (ICCCPT), Apr. 2017, pp. 1–6.

[34] D. Zhang, X. Peng, K. Pan, and Y. Liu, “A novel wind speed forecasting based on hybrid prediction model and online sequential outlier robust extreme learning machine,” Energy Convers. Manage., vol. 180, pp. 338–357, Jan. 2019.

[35] C. Wang, H. Zhang, W. Fan, and P. Ma, “A new chaotic time series hybrid prediction method of wind power based on EEMD-SE and full-parameters continued fraction,” Energy, vol. 138, pp. 977–990, Nov. 2017.

[36] J. S. Richman, D. E. Lake, and J. R. Moorman, “Sample entropy,” in Methods in Enzymology, vol. 384. Amsterdam, The Netherlands: Elsevier, 2004, pp. 172–184.

[37] Y. Shang, G. Lu, Y. Kang, Z. Zhou, B. Duan, and C. Zhang, “A multi-fault diagnosis method based on modified sample entropy for lithium-ion battery strings,” J. Power Sources, vol. 446, Jan. 2020, Art. no. 227275.

[38] S. Chen, S. Dong, Z. Cao, and J. Guo, “A compound approach for monthly runoff forecasting based on multiscale analysis and deep network with sequential structure,” Water, vol. 12, no. 8, p. 2274, 2020.

[39] 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.

[40] R. Bisoi, P. Dash, and A. Parida, “Hybrid variational mode decomposition and evolutionary robust kernel extreme learning machine for stock price and movement prediction on daily basis,” Appl. Soft Comput., vol. 74, pp. 652–678, Jan. 2019.

[41] W.-Y. Deng, Q.-H. Zheng, and Z.-M. Wang, “Cross-person activity recognition using reduced kernel extreme learning machine,” Neural Netw., vol. 53, pp. 1–7, May 2014, doi: 10.1016/j.neunet.2014.01.008.

[42] W. Sun and Y. Wang, “Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network,” Energy Convers. Manage., vol. 157, pp. 1–12, Feb. 2018.

[43] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, “Spline mould algorithm: A new method for stochastic optimization,” Future Gener. Comput. Syst., vol. 111, pp. 300–323, Oct. 2020.

[44] 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, 2006.

[45] Z. F. Liu, L. L. Li, M. L. Tseng, and M. K. Lim, “Prediction short-term photovoltaic power using improved chicken swarm optimizer-Extreme learning machine model,” J. Cleaner Prod., vol. 248, Mar. 2020, Art. no. 119272.

[46] H. Yildirim and M. R. Özkale, “The performance of ELM based ridge regression via the regularization parameters,” Expert Syst. Appl., vol. 134, pp. 225–233, Nov. 2019.

[47] W. Deng, Q. Zheng, and K. Zhang, “Reduced kernel extreme learning machine,” in Proc. 8th Int. Conf. Comput. Recognit. Syst. (CORES). Cham, Switzerland: Springer, 2013, pp. 63–69.

[48] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” 2014, arXiv:1412.3555. [Online]. Available: http://arxiv.org/abs/1412.3555

[49] Y. Wang, Q. Hu, D. Meng, and P. Zhu, “Deterministic and probabilistic wind power forecasting using a variational Bayesian-based adaptive robust multi-kernel regression model,” Appl. Energy, vol. 208, pp. 1097–1112, Dec. 2017.

[50] J.-Z. Wang, Y. Wang, and P. Jiang, “The study and application of a novel hybrid forecasting model—A case study of wind speed forecasting in China,” Appl. Energy, vol. 143, pp. 472–488, Apr. 2015.

[51] F. Diebold and R. Mariano, “Comparing predictive accuracy,” J. Bus. Econ. Statist., vol. 20, no. 1, pp. 134–144, Jan. 2002.

HUA YAN was born in Hubei, China, in 1983. He received the B.S. degree from the Continuing Education College, China Three Gorges University, in 2010. He is currently pursuing the master’s degree with the College of Electrical Engineering and New Energy, China Three Gorges University. His research interests include the Internet of Things and deep learning algorithms.

ZHENGPING WU was born in Hunan, China, in 1966. He received the Ph.D. degree from the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, in 2007. In 2007, he went to China Three Gorges University to conduct teaching and research work. His research interests include design and optimization of control systems, artificial intelligent algorithms for control systems, image recognition of substation, and the Internet of Things.