Research article

A novel Bayesian ensembling model for wind power forecasting

Jingwei Tang a,b, Jianming Hu a,*, Jiani Heng c, Zhi Liu b

a College of Economics and Statistics, Guangzhou University, Guangzhou, China
b Department of Mathematics, Faculty of Science and Technology, University of Macau, Macau, China
c Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

ARTICLE INFO

Keywords:
Bayesian ensembling
Echo state network
Generalized mixture function

ABSTRACT

Precise and robust wind power prediction can effectively alleviate the problem caused by the randomness and volatility of wind power. Ensemble learning can successfully improve forecasting precision and robustness, and quantify the uncertainty of the prediction. This paper presents a new ensemble probabilistic forecasting framework, based on modified randomized maximum a posteriori (MAP) sampling technique, echo state network (ESN) and generalized mixture (GM) function to bring superior forecasting results. The proposed model first trains a set of independent ESN models for probabilistic forecasting using the modified randomized MAP sampling technique, and then dynamically weighs and ensembles the base model forecasting through the GM function. The proposed model and other benchmark models have been implemented on four wind power datasets from different places to illustrate the advantage of the proposed method. The compared result indicates that the suggested model outperforms some state-of-the-art models and can successfully achieve dynamic ensemble probabilistic prediction.

1. Introduction

As a kind of clean and renewable energy, wind energy has been widely used and developed all over the world. According to the Global Wind Energy Council (GWEC), 94 GW of both onshore and offshore wind turbines have been installed in 2021, bringing total capacity up to 837 GW [1]. Wind energy will continue to grow and have an essential role in the global energy transition to a low-carbon economy. With the increasing penetration of wind energy, its influence on the utility system planning and operation for generation and transmission has gradually increased [2]. But due to the inherent intermittency, uncertainty and volatility of wind energy, the reliability and security of the power system are being threatened. In view of these, regulations for generators or grid ancillary services are carried out on the power grid. However, all these measures will impose an additional economic burden on the operation of the grid. Therefore, high accuracy wind power prediction is urgently needed to handle these problems, for it can promote wind power into the grid system, thereby reducing forecasting risk, enhancing the security of the grid and improving economic and social benefits.

Up to now, many pieces of research have been devoted to the study of prediction methods to enhance the accurateness and efficiency of wind power forecasting. These forecasting methods can be classified into physical methods, statistical methods, artificial intelligence methods, and combination prediction methods.

Physical methods mainly adopt topographic features and meteorological factors as input variables to predict the trend of future wind power, including the fifth-generation mesoscale model (MM5) [3], numerical weather prediction (NWP) model [4, 5, 6] and computational fluid dynamics (CFD) model [7], etc. For the reason that there is no need for massive historical data, these methods have an advantage in long-term prediction. However, these methods are more suitable for professionals since they have requirements for meteorological knowledge. With respect to statistical methods, historical data is utilized to model its relationship with predicted data and random errors, which can be expressed in the form of a function and the short-term wind power can be effectively predicted. Statistical models include regression analysis model [7], kernel regression (KR) [8], vector autoregression (VAR) [9], and exponential smoothing [10], etc. However, these methods not only require a large amount of historical data, but also its prediction error increases over time, making it not suitable for long-term prediction. With the development of information technology, artificial intelligence methods with outstanding learning performance have been widely used.

* Corresponding author.
E-mail address: hujm17@gzhu.edu.cn (J. Hu).
for wind power prediction, such as neural network (NN) [11, 12], gated recurrent neural network (RNN) [13, 14] and extreme learning machine model [15, 16], etc. This kind of prediction methods built on artificial intelligence architecture can explore the complex implicit relationship between input and output without determining the specific function expression, only need to train and learn the model established by the historical data. But such methods may suffer from the problems of overlearning or local optimum.

The above-mentioned methods all fall into the single prediction method. From the perspective of forecasting performance, each model has its strengths and weaknesses since its output is influenced by input variables and the internal structure of the model. A single model may possess good behavior in specific forecasting circumstances, but it will also result in a large error at certain observed data. There is no optimal model that can guarantee to obtain the best result for all cases, and then different models need to be measured to obtain the optimal one before application [17]. Therefore, researchers have turned their attention to the study of combination prediction models.

There are two main definitions of the combination. One is presented in the form of taking advantage of different methodologies purposefully, the other is in the form of generating the weighting factor based on the prediction evaluation of each combination member [18].

With respect to the former category, they are generally composed of two kinds of different processes, one is for the primary forecasting process and the other is for the auxiliary process such as data pre-processing, parameter selection and optimization, and data post-processing [18]. For instance, Du et al. [19] proposed an improved empirical mode decomposition (EMD) with adaptive noise technology to decompose the raw data and extract the pivotal feature. Li et al. [20] implemented adaptive learning factor and differential evolution strategy to improve the dragonfly algorithm to perform the process of optimal parameter selection. Wu et al. [21] proposed a complete ensemble EMD to divide the original data series into a group of inherent mode functions and adopted a multi-objective grey wolf algorithm to optimize the parameter of the extreme machine. Luo et al. [22] took the wavelet transform (WT) decomposition method as pre-processing method and applied a multi-objective cuckoo search (MOCS) algorithm to the hyperparameter of Elman neural network. Improved hybrid time series decomposition strategy (HTD), proposed by Lv et al. [23], can simultaneously abstract the linear mode and frequency-domain information from time series and used the sequence-to-sequence model to predict the wind speed at different temporal resolutions. Although the prediction accuracy of these models, based on only one predictor, has improved, these approaches struggle to handle all the different data features and scenarios.

For the latter category, which is also named as weighting-based combination model, it takes the relative effectiveness of each combination member into consideration and thereby arranging them an appropriate weight [18], wherein the crucial point lies in the technique of combination [24]. As a consequence, many optimization algorithms have been used for optimizing the weights including the multi-objective multi-verse optimizer [25], the multi-objective dragonfly algorithm [26], the flower pollination algorithm [27], the multi-objective grey wolf optimizer [28, 29], bat algorithm [30] and so on. In addition, there are some other methods to calculate the combination weights of every single model. For instance, in ref [31], the weight of each predictor was assigned in terms of their inverse of the mean squared error (MSE) on validation dataset. In ref [32], the combination weights were determined by the multivariate statistical method partial least squares. Although the performance of these combination models has improved, they used a static way rather than a dynamic way to determine combination weights. Namely, these studies defined weights during the training or validation process and used them throughout the testing process [33], neglecting the dynamic change of input data and model performance. This can eventually result in the inefficient of the combination forecasting using such kind of weight, since the positive information behind each model cannot be timely captured. Therefore, it is necessary to adopt dynamic weighting method rather than static weighting method.

In the process of dynamic ensembling, the results of all base models are combined and the combination weights vary with the model capacity. Some researchers have applied the dynamic weighting method to combination model. For example, in the research of Jiang [34] and Li [35], the variable weighted combination theory was implemented and the regression coefficients were adapted as combination weights. In ref [36], the in-sample time-varying combination weights were optimized by quadratic programming and then the out-of-sample time-varying adaptive weights were predicted by the high-order Markov chain model to update the in-sample weights. In ref [37], an in-sample training-validation pair-based neural network weighting method was proposed, which can acquire the weight information of predictors and forecast the acquired combination weights. Duan et al. [38] proposed a nonlinear weighted combination method to integrate the multiple subsequence prediction models by the deep belief networks and the particle swarm optimization algorithm. Moreover, in the research of Yin [39] and Chen [40], the optimal ensemble weights of the base learners are dynamically selected by a deep reinforcement learning model. However, in most of these researches, an additional process is required to construct the combination weights, which complicates the model structure and computational process.

In addition, given the problem that quantifying the uncertainty of wind power is essential, ensembling provides a pragmatic and scalable solution for it. Through ensembling, results of multiple single estimators are integrated, which are trained using different initializations and sometimes on noisy versions of the training data [41]. Unfortunately, the weakness of this solution is that its inference is not Bayesian since the concept of prior is not preserved. Some scholars have done related research to handle this problem. In ref [42], proper scoring rules were used for training the NNs, and the proposal was easier to implement than Markov chain Monte Carlo, requiring fewer NNs modifications. In ref [43], each ensemble member was individually regularized to a distinct random prior function. Although it successfully constructed the uncertainty measurement mechanism, it adds noise to both the regularization and target terms, and cannot reproduce the true posterior. Meanwhile, the idea that adding a regularization term to the loss function can produce the MAP estimation of parameters has attracted the attention of scholars, since injecting noise into the regularization term or target sets of parameters and sample repeatedly, the distribution of MAP solution that approximates the true posterior distribution can be generated. This method can be considered as a useful method for sampling from a higher-dimensional posterior [44].

In consideration of the above discussions, to fill the gaps in the current forecasting models and ameliorate the prediction effect, this paper...
proposes a novel ensemble forecasting framework, which is composed of ESN, modified randomized MAP and GM function. First, the ESN is used as an individual model for wind power prediction. Second, the modified randomized MAP is utilized to repeat sampling to generate several single models. Finally, all the single predictions are dynamically integrated to get the final prediction results by the GM function. The major contributions are concluded as follows:

1) A new and advanced ensembling model based on ESN and modified randomized MAP and GM function is developed in this...
2. Methodology

2.1. Echo state networks

The ESN proposed by Jaeger, based on reservoir computing method, is an extremely efficient learning algorithm [45]. In general, the classical RNNs need to train all parameters and utilizes a gradient descent method to learn the model. This approach has the disadvantages of converging slowly, falling into local optimum easily, bifurcating, and gradient vanishing or exploding, thus it is difficult to be applied to practical problems [45]. As a newly RNNs, the ESNs get over the above difficulties by using the randomly generated sparse networks (namely reservoir) as the medium of information processing, that maps the input signal from low dimension into high dimensional state space, and then derive the output through training partial connection weights in the state space [46]. An illustration of the ESN is shown in Figure 1. It is worth noting that the weights being trained are only the output weights, while the reservoir weights are kept unvaried [47].

We first consider an ESN with N input features, L reservoir units and M outputs. The state of input vector at any time t can be represented as

\[ \mathbf{v}_t = (v_t(1), v_t(2), \ldots, v_t(N)) \],

of reservoir units is \( \mathbf{r}_t = (r_t(1), r_t(2), \ldots, r_t(L)) \), and of outputs is \( \mathbf{p}_t = (p_t(1), p_t(2), \ldots, p_t(M)) \). The propagation process of the ESN, in more detail, can be expressed as Eq. (1):

\[
\mathbf{r}_{t+1} = (1 - \lambda) \mathbf{g}(\mathbf{U}_r + \mathbf{U}_i \mathbf{v}_{t+1} + \mathbf{U}_p \mathbf{p}_t) + \lambda \mathbf{r}_t
\]

in which \( \lambda \) is the leak rate, \( \mathbf{g} (\cdot) \) is the nonlinear function, \( \mathbf{U}_r \in \mathbb{R}^{L \times L} \) is the internal reservoir state weight, \( \mathbf{U}_i \in \mathbb{R}^{L \times N} \) is the input weight, \( \mathbf{U}_p \in \mathbb{R}^{L \times M} \) is the target output weight and the predicted output units \( \mathbf{p}_t \) are calculated through Eq. (2):

\[
\mathbf{p}_{t+1} = \mathbf{U}_p \mathbf{r}_{t+1} + \mathbf{v}_{t+1}
\]

in which \( \mathbf{U}_p \in \mathbb{R}^{M \times (N+L)} \) is the output weight and \( \mathbf{r}_{t+1}; \mathbf{v}_{t+1} \) is the extension matrix of reservoir units and input.

It is worth noting that there will be a shortage of memory of the ESN since only the output weight \( \mathbf{U}_p \) will be learned. Nevertheless, the units \( \mathbf{r}_t \) must have the ability of memory to capture the temporal characteristic in the time series data. In Maass et al. [48], this problem can be addressed by adding two conditions. First, each unit in \( \mathbf{r}_t \) should be separable, and the diversities of each input in \( \mathbf{v}_t \) should be captured in the relevant unit in \( \mathbf{r}_t \). Second, the units \( \mathbf{r}_t \) should be propagated to the output vector \( \mathbf{p}_t \),

The rest of this paper is organized as follows: Section 2 describes the methodologies that constitute the proposed combination prediction method. Section 3 presents the results and discussions of the benchmarking experiments, and Section 4 concludes this paper.

---

### Table 1. Statistical indicators of Datasets A-D.

| Dataset | Time interval | Number | Mean (kW) | Max. (kW) | Min. (kW) | Med. (kW) | Std. (kW) |
|---------|---------------|--------|-----------|-----------|-----------|-----------|-----------|
| All samples | 15-min | 11920 | 24466.65 | 195140.5 | -968.68 | 3433.84 | 41820.78 |
| Training set | 2000 | 9536 | 23557.66 | 195140.5 | -968.68 | 2924.31 | 42440.39 |
| Testing set | 600 | 2384 | 28683.88 | 184228.2 | -9 | 6353.14 | 39030.19 |
| All samples | 1-h | 11920 | 3052.04 | 16163.41 | 0 | 1419.22 | 3484.61 |
| Training set | 6353.14 | 9536 | 2832.45 | 16163.41 | 0 | 2877.24 | 3785.35 |
| Testing set | 3940 | 2384 | 3931.13 | 14841.15 | 0 | 1661.66 | 3574.04 |
| Dataset A | 10-min | 19700 | 97.58 | 1080.00 | 10.00 | 60.00 | 102.49 |
| Training set | 19700 | 15760 | 98.09 | 1080.00 | 10.00 | 60.00 | 105.26 |
| Testing set | 3940 | 2384 | 95.56 | 850.00 | 20.00 | 60.00 | 90.54 |
| Dataset C | 5-min | 105115 | 6767.02 | 23709.69 | 0 | 2492.27 | 6440.36 |
| All samples | 105115 | 84092 | 6842.28 | 23709.69 | 0 | 4910.00 | 6564.41 |
| Testing set | 6466.02 | 21023 | 6466.02 | 23709.69 | -277.67 | 4926.27 | 6440.36 |

---

### Table 2. The setting of hyper-parameters.

| Learning rate | Decay rate | Hidden size | Epoch | Batch size |
|---------------|------------|-------------|-------|------------|
| GRU 0.09      | 0.985      | 20          | 2000  | 600        |
| LSTM 0.09     | 0.985      | 20          | 2000  | 600        |
| MLP 0.09      | 0.985      | 50          | 2000  | 600        |
| ESN 0.09      | 0.985      | 50          | 2000  | 600        |
Figure 3. The fitting curves of six single models and their combination models for Datasets A to D.
through the readout function based on the required precision. Therefore, the ESN has the capability of processing time series data.

### 2.2. Anchored ensembles of neural networks

Understanding the information behind the uncertainty of prediction of the Neural Networks (NNs) is vitally important. The Bayesian framework is one of the methods to handle this issue, since it can model the uncertainty and provide predictive distribution, but its implementation in NNs is greatly limited by the large scale of parameters and data. Ensembling offers another choice to learn about the uncertainty, and provide predictive distribution, but its implementation is also disappointing since all ensemble members are encouraged to propose a modiﬁed ensembling process on the basis of the Bayesian framework, named as randomized anchored MAP sampling, abbreviated anchored ensemble. The major contributions of the new scheme are: (1) consider a general model with normally prior distribution and normally likelihood distribution are both multivariate normal, the posterior distribution is also normal, \( N(\mu_{\text{post}}, \Sigma_{\text{post}}) \propto N(\mu_{\text{prior}}, \Sigma_{\text{prior}}) \cdot N(\mu_{\text{like}}, \Sigma_{\text{like}}) \). Without loss of generality, \( \mu_{\text{post}} \) can be obtained as Eq. (4):

\[
\mu_{\text{post}} = \left( \Sigma_{\text{like}}^{-1} + \Sigma_{\text{prior}}^{-1} \right)^{-1} \left( \Sigma_{\text{like}}^{-1} \mu_{\text{like}} + \Sigma_{\text{prior}}^{-1} \mu_{\text{prior}} \right)
\]

Figure 4. The partial combination weights of the combination models for Datasets A to D.

In general, \( \mu_{\text{prior}} \) and \( \gamma \) are chosen to transform into noisy random variables, so that \( \text{Var}(\gamma_{\text{post}}) = \Sigma_{\text{post}} \) can be achieved. For the reason that it may be impossible to extract element that exactly includes \( y \) from nonlinear regression model. Moreover, for a NN model, before adding noise to \( y \), it is required to assign different noise distributions to different parameters, but it is not possible because these parameters have a single target set. Consequently, it is infeasible to transform target \( y \) into noise variable in the case of NNs, and then \( \gamma_{\text{prior}} \) becomes the only choice to be injected into noise.

Substitute noisy random variable \( \gamma_{0} \) for \( \gamma_{\text{prior}} \), take samples from the prior distribution, and the MAP estimate of \( \mu_{\text{prior}}(\theta_{0}) \) can be represented as Eq. (5):

\[
\mu_{\text{MAP}}^{\text{like}}(\theta_{0}) = \left( \Sigma_{\text{like}}^{-1} + \Sigma_{\text{prior}}^{-1} \right)^{-1} \left( \Sigma_{\text{like}}^{-1} \mu_{\text{like}} + \Sigma_{\text{prior}}^{-1} \theta_{0} \right)
\]

wherein the variance of it is set to \( \text{Var}(\mu_{\text{MAP}}(\theta_{0})) = \Sigma_{\text{post}} \), and \( \theta_{0} \sim N(\mu_{0}, \Sigma_{0}) \), \( \mu_{0} = \mu_{\text{prior}}, \Sigma_{0} = \Sigma_{\text{prior}} + \Sigma_{\text{prior}} \Sigma_{\text{like}} \Sigma_{\text{like}}^{-1} \Sigma_{\text{prior}} \).

Due to the space limit, we will not present the derivation of the distribution of \( \theta_{0} \) here. Conceive a NN with hidden width \( H \), we find that \( \text{diag}(\Sigma_{0}) \geq \text{diag}(\Sigma_{\text{prior}}) \), the matrix \( \Sigma_{\text{prior}} \Sigma_{\text{like}}^{-1} \) tends to 0 as the width increases. Using \( \Sigma_{0} = \Sigma_{\text{prior}} \) also presents good empirical performance. Thus, it is feasible for us to denote \( \Sigma_{0} \approx \Sigma_{\text{prior}} \) under the case that the posterior distribution is governed by the prior. This allows us to relax the hypothesis to normally prior distribution and normally likelihood distribution.

Therefore, a NN can be trained by minimizing the newly regularized loss function as Eq. (6):

\[
\text{Loss}_{\text{regularized}} = \frac{1}{N} |y - \hat{y}|^2 + \frac{1}{N} \| \Gamma^{1/2} \theta \|^2
\]
The specific process of the anchored ensembles of neural networks is presented in Algorithm I.

Algorithm I. Anchored Ensembles of Neural Networks

Input: a sequence of training dataset \((X_i, y_i)\) and testing dataset \(X\)

Preset: ensemble size \(M\), noisy random variable \(\theta_0 \sim N(0, \sigma_{\theta_0}^2)\),
estimated data error \(\epsilon \sim N(0, \sigma_\epsilon^2)\),
regularization matrix \(\lambda \sim \lambda_0 + \lambda_1\)

\(\#\) training ensemble process
for \((i \in 1:M)\)

- initialize \(N\) from \(N(\mu_{\omega_0}, \Sigma_{\omega_0})\)
- train \((X_i, y_i, \Gamma, \theta_0)\) with loss function Eq. (6)

end for

\(\#\) forecasting process with ensemble \((i \in 1:M)\)

\[ \hat{y}_i = N(\mu_{\omega_i}, \Sigma_{\omega_i}) \]

end for

Output: predicted distribution mean \(\hat{y}\) and variance \(\sigma_\epsilon^2\)

2.3. Generalized mixture functions

The GM function [33] is a combination approach, which is the generalization of the ordered weighted averaging (OWA) and mixture function, and can significantly improve the efficiency of the combination process since it dynamically determines weights according to the performance of each combination member and the relationship among the performance of all members.

With respect to the OWA, it relates weights to all the components \(x_i\) of the input vector and is defined as Eq. (7):

\[
\text{OWA}_w(x_1, ..., x_n) = \sum_{i=1}^{n} w_i x_i
\]

where \(w = (w_1, ..., w_n)\) is the predefined vector of weights, with \(\sum_{i=1}^{n} w_i = 1\), and permutation \((x_1, ..., x_n)\) is the descending sequence of \((x_1, ..., x_n)\).

Similar to the OWA, the GM function proposed by Pereira et al. [49, 50] also has the ability to consider the corresponding weight according to the characteristics of the input vector. The difference is that these weights are not fixed but can be determined dynamically.

Consider a finite family of functions \(T = \{g_i : [0, 1]^n \rightarrow [0, 1], 1 \leq i \leq n\}\), with \(\sum_{i=1}^{n} g_i(x_1, ..., x_n) = 1\), is referred to as the family of weight functions. Hence, the generalized mixture function associated with \(T\) is the function with the form of \(GM_T(x_1, ..., x_n) = \sum_{i=1}^{n} g_i(x_1, ..., x_n)\).

Table 3. Forecasting performance of six single models and their combination models for Dataset A to D.

| Dataset | Model | Evaluation criteria | RMSE | MAE | U1 |
|---------|-------|---------------------|------|-----|----|
| A       | KR    | 6422.44             | 3351.73 | 6.71 |
|         | GRU   | 6585.67             | 3722.61 | 6.99 |
|         | LSTM  | 6505.41             | 3627.15 | 6.88 |
|         | MLP   | 6415.74             | 3487.55 | 6.66 |
|         | SVR   | 6424.67             | 3979.08 | 6.68 |
|         | ESN   | 6275.28             | 3269.61 | 6.56 |
|         | Comb_s| 6268.52             | 3421.12 | 6.57 |
|         | Comb_m| 6267.25             | 3407.12 | 6.57 |
|         | Comb_g| 6266.35             | 3390.16 | 6.57 |
| B       | KR    | 1389.18             | 947.93  | 13.04 |
|         | GRU   | 1415.45             | 966.56  | 12.93 |
|         | LSTM  | 1414.08             | 951.44  | 13.17 |
|         | MLP   | 1380.6              | 925.74  | 12.83 |
|         | SVR   | 1408.72             | 957.28  | 13.09 |
|         | ESN   | 1378.75             | 949.59  | 12.87 |
|         | Comb_s| 1371.85             | 932.82  | 12.75 |
|         | Comb_m| 1371.78             | 932.74  | 12.75 |
|         | Comb_g| 1371.69             | 933.31  | 12.73 |
| C       | KR    | 89.53               | 57.55   | 67.96 |
|         | GRU   | 65.17               | 35.02   | 49.47 |
|         | LSTM  | 64.37               | 34.52   | 48.87 |
|         | MLP   | 66.03               | 34.01   | 50.13 |
|         | SVR   | 90.74               | 45.63   | 68.88 |
|         | ESN   | 63.91               | 34.82   | 48.72 |
|         | Comb_s| 63.97               | 34.79   | 48.56 |
|         | Comb_m| 62.8                | 33.77   | 47.67 |
|         | Comb_g| 62.41               | 33.04   | 47.38 |
| D       | KR    | 798.33              | 484.67  | 9.12  |
|         | GRU   | 750.26              | 459.71  | 8.57  |
|         | LSTM  | 799.35              | 506.07  | 9.13  |
|         | MLP   | 759.37              | 478.43  | 8.67  |
|         | SVR   | 781.06              | 532.85  | 8.92  |
|         | ESN   | 744.47              | 459.56  | 8.5   |
|         | Comb_s| 739.5               | 458.33  | 8.45  |
|         | Comb_m| 739.29              | 457.45  | 8.44  |
|         | Comb_g| 737.88              | 455.75  | 8.43  |

\[
\text{Loss}_{\text{anchor}_i} = \frac{1}{N} \| y - \hat{y}_i \|^2 + \frac{1}{N} \left\| \Gamma X_{\theta_j} \phi_{\theta_j} \right\|^2
\]

wherein \(\text{diag}(\Gamma_k) = \phi_{\theta_j} \phi_{\theta_j}^\top \sim N(\mu_{\omega_0}, \Sigma_{\omega_0})\).

Table 4. Forecasting performance of the combination models composed of six models for Dataset A.

| Model | MLP   | RMSE  | MAE   | U1 × 10^{-2} |
|-------|-------|-------|-------|--------------|
| #1    | 6494.28| 3387.46| 6.77  | 1.64 × 10^{-2} |
| #2    | 6322.12| 3276.08| 6.64  | 1.64 × 10^{-2} |
| #3    | 6349.11| 3607.57| 6.65  | 1.64 × 10^{-2} |
| #4    | 6284.32| 3480.05| 6.58  | 1.64 × 10^{-2} |
| #5    | 6288.97| 3214.52| 6.57  | 1.64 × 10^{-2} |
| #6    | 6263.63| 3141.35| 6.59  | 1.64 × 10^{-2} |
| Comb #1-6_s| 6171.27| 3185.80| 6.64  | 1.64 × 10^{-2} |
| Comb #1-6_m| 6171.67| 3180.97| 6.64  | 1.64 × 10^{-2} |
Table 5. Forecasting performance of the combination models composed of six models for Dataset B.

| Model | MLP | MAE | U1 \( \times 10^{-2} \) | GRU | MAE | U1 \( \times 10^{-2} \) | LSTM | MAE | U1 \( \times 10^{-2} \) | ESN | MAE | U1 \( \times 10^{-2} \) |
|-------|-----|-----|----------------|-----|-----|----------------|-------|-----|----------------|-----|-----|----------------|
| Model #1 | 65.08 | 30.86 | 49.40 | 66.03 | 34.01 | 50.13 | 65.66 | 30.94 | 49.84 | 64.37 | 34.52 | 48.87 |
| Model #2 | 65.17 | 30.52 | 49.47 | 66.25 | 33.68 | 50.29 | 66.61 | 32.27 | 50.56 | 62.99 | 32.72 | 47.82 |
| Model #3 | 62.76 | 33.54 | 47.64 | 64.53 | 32.71 | 48.98 | 68.14 | 32.96 | 51.73 | 62.58 | 31.20 | 47.50 |
| Model #4 | 63.13 | 29.76 | 47.92 | 63.18 | 30.17 | 47.90 | 66.82 | 35.74 | 50.72 | 63.37 | 33.33 | 48.10 |
| Model #5 | 64.80 | 32.57 | 49.19 | 72.35 | 37.55 | 54.92 | 70.14 | 33.93 | 53.24 | 63.01 | 32.43 | 47.83 |

Table 6. Forecasting performance of the combination models composed of six models for Dataset C.

| Model | MLP | MAE | U1 \( \times 10^{-2} \) | GRU | MAE | U1 \( \times 10^{-2} \) | LSTM | MAE | U1 \( \times 10^{-2} \) | ESN | MAE | U1 \( \times 10^{-2} \) |
|-------|-----|-----|----------------|-----|-----|----------------|-------|-----|----------------|-----|-----|----------------|
| Model #1 | 747.19 | 458.49 | 8.53 | 738.69 | 449.75 | 8.44 | 799.35 | 506.07 | 9.13 | 746.46 | 461.69 | 8.52 |
| Model #2 | 747.61 | 462.48 | 8.54 | 741.19 | 449.64 | 8.46 | 754.27 | 469.52 | 8.61 | 736.00 | 451.65 | 8.41 |
| Model #3 | 753.33 | 465.50 | 8.60 | 794.95 | 496.29 | 9.08 | 794.92 | 497.68 | 9.08 | 737.69 | 451.39 | 8.42 |
| Model #4 | 784.77 | 484.45 | 8.96 | 750.26 | 458.17 | 8.57 | 757.02 | 473.77 | 8.65 | 745.40 | 473.61 | 8.51 |
| Model #5 | 760.68 | 473.43 | 8.69 | 738.78 | 447.18 | 8.44 | 770.16 | 477.16 | 8.80 | 745.73 | 464.76 | 8.52 |
| Model #6 | 746.56 | 459.76 | 8.53 | 808.54 | 506.66 | 9.23 | 804.72 | 510.51 | 9.19 | 735.55 | 453.32 | 8.40 |

Table 7. Forecasting performance of the combination models composed of six models for Dataset D.

| Model | MLP | MAE | U1 \( \times 10^{-2} \) | GRU | MAE | U1 \( \times 10^{-2} \) | LSTM | MAE | U1 \( \times 10^{-2} \) | ESN | MAE | U1 \( \times 10^{-2} \) |
|-------|-----|-----|----------------|-----|-----|----------------|-------|-----|----------------|-----|-----|----------------|
| Model #1 | 10_gm | 3320.20 | 6.59 | 3320.20 | 3381.36 | 6.78 | 6456.89 | 3433.19 | 6.69 | 6427.78 | 3288.37 | 6.68 |
| Model #2 | 6484.15 | 3397.83 | 6.75 | 6532.31 | 3436.84 | 6.70 | 6485.15 | 3487.30 | 6.68 | 6129.47 | 3127.17 | 6.39 |
| Model #3 | 6407.69 | 3385.56 | 6.66 | 6261.84 | 3209.58 | 6.59 | 6413.80 | 3853.51 | 6.75 | 6125.86 | 3152.82 | 6.40 |
| Model #4 | 6390.65 | 3408.57 | 6.67 | 6220.99 | 3203.46 | 6.46 | 6354.49 | 3408.81 | 6.60 | 6144.27 | 3189.05 | 6.42 |

Table 8. Forecasting performance of the combination models composed of ten models for Dataset A.

| Model | MLP | MAE | U1 \( \times 10^{-2} \) | GRU | MAE | U1 \( \times 10^{-2} \) | LSTM | MAE | U1 \( \times 10^{-2} \) | ESN | MAE | U1 \( \times 10^{-2} \) |
|-------|-----|-----|----------------|-----|-----|----------------|-------|-----|----------------|-----|-----|----------------|
| Model #7 | 6173.93 | 3100.00 | 6.42 | 6294.65 | 3267.01 | 6.54 | 6382.11 | 3387.07 | 6.62 | 6122.73 | 3092.34 | 6.39 |
and normalize them.

variables, follow the principle of the partial autocorrelation function

\( \Theta \)

as Eq. (8),

\( \mathbf{H}_n(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i \Theta(x_1, \ldots, x_n) \)

or otherwise

(8)

For instance, function \( \Theta \) can be the form of median, maximum, minimum and arithmetic mean.

2.4. Proposed ensemble forecasting framework

This subsection presents the ensemble forecasting composed of the ESN, the anchored ensembles and the GM function is proposed, named as EAEG. The flow diagram of the proposed method is shown in Figure 2, and the specific explanation of the entire forecasting framework is as follows.

Firstly, choose original wind power and speed data as explanatory variables, follow the principle of the partial autocorrelation function (PACF) [51], select an appropriate lag order for explanatory variables and normalize them.

Secondly, in order to reduce the calculative burden and accelerate the execution time, the ESN is regarded as the basic model, by utilizing the technique of the anchored ensembles to sampling generate several ESN models, and then trained network by using different initializations. Beyond that, to avoid building an extra model, and capture the information of time-varying data the GM function is applied to dynamically assign the combination weights according to the prediction evaluation of models to combine the prediction results of each model. By comparing with other single models, the proposed combination model can handle the complex features of wind power generation and make the fluctuation trend of the final prediction result more consistent with that of the actual power.

Thirdly, provide an empirical comparison of the proposed model with other benchmark models on several real-world datasets, discuss and analyze the experimental results to evaluate the validity and generalization of the proposed method.

3. Experiment and analysis

For the purpose of testing the performance of the presented approach, two experiments based on real-world data are carried out in this section, which is denoted as Experiment I and Experiment II, respectively. In Experiment I, the combination approach of six single models is conducted using three different combination methods, and the demonstration of the effectiveness of the combination models is achieved by...
Figure 5. The fitting curves of the combination models composed of six ESNs for Datasets A to D.
Figure 6. The fitting curves of the combination models composed of ten ESNs for Datasets A to D.
Figure 7. The prediction intervals of the combination model composed of ten ESNs for Datasets A to D, using the GM function as combination method.
comparing them with the single benchmark models, including kernel regression (KR) [8], GRU [13], LSTM [14], multi-layer perception (MLP) [12], support vector regression (SVR) [52], which are widely used in wind power prediction. In Experiment II, for the sake of verifying the superiority of the improved ensemble forecasting framework, a comparison is made between the proposed method and other benchmark combination methods, which are composed of sampling of the benchmark models. In order to score the deterministic forecasting performance of the developed method, three statistical metrics, including root mean square error (RMSE) as Eq. (9), mean absolute error (MAE) as Eq. (10), and Theil U statistic (U1) as Eq. (11), are employed to measure forecasting validity. All experiment codes are executed on Python 3.9, with the computer configuration is 4.00 GHz CPU, 16.0 GB RAM, Ubuntu.

- Root mean square error (RMSE)

\[
RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2} \tag{9}
\]

- Mean absolute error (MAE)

\[
MAE = \frac{1}{T} \sum_{t=1}^{T} |y_t - \hat{y}_t| \tag{10}
\]

- Theil’s U statistic (U1)

\[
U1 = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2} \left/ \sqrt{\frac{1}{T} \sum_{t=1}^{T} y_t^2 + \frac{1}{T} \sum_{t=1}^{T} \hat{y}_t^2} \right. \tag{11}
\]

3.1. Dataset description and setting

In this subsection, the effectiveness of the proposed ensembling method is proved by the different time interval data from different places. The experiment datasets are collected from the Guazhou Wind Farm (40°31’12” North latitude and 95°46’56” East longitude, China), the Sotavento Experimental Wind Farm (43°21’15” North latitude and 7°52’52” West longitude, Spain), the low energy building (the public appliances energy prediction dataset in UCI Machine Learning Repository) and the Ararat Wind Farm (37°14’24” South latitude and 142°58’48” East longitude, Australia). More specifically, data from the Guazhou Wind Farm is denoted as Dataset A, with a 15-minute period; data collected from the Sotavento Experimental Wind Farm is denoted as Dataset B, with 1-hour period; data from the low energy building is denoted as Dataset C, with a 10-minute period; data collected from the Ararat Wind Farm is denoted as Dataset D, with 5-minute period. There is no exact theory to divide the proportion between the training dataset and the testing dataset. However, there’s no denying that it is extremely significant to reasonably determine the proportion between the two. Too small training samples will generate an underfitting model, and too large training samples will give rise to overfitting [53]. On the basis of the size and feature of the datasets, the ratio between the training samples and the testing sample is 4:1 in the experiments. Detailed pieces of information from the datasets are exhibited in Table 1.

With respect to the data preprocessing, firstly check the PACF of data. As for the dataset used in the experiments, according to the PACF, the autocorrelation of data features is eliminated when the lag order is 2. Secondly, normalize the data to the range of 0–1 to overcome the effect of index dimension. After performing the above process, the normalized wind power and wind speed features from lag-2 to lag-1 as input features for prediction models.
Figure 9. The fitting curves of the combination models of Four different benchmark models for datasets A to D.
3.2. Experiment I

In order to demonstrate the effectiveness of constructing a combination model, four datasets collected from different places with different time intervals are utilized in this experiment. An agglomeration of based models such as KR, GRU, LSTM, MLP, SVR and ESN are employed to perform one-step-ahead forecasting and the forecasting results are compared with those combination models combined by these models. Furthermore, by comparing the prediction results of combination models constructed using different combination methods (i.e., softmax function, mean function and GM function, and the corresponding combination models are denoted as Comb_s, Comb_m and Comb_gm, respectively), the contribution of the combination method used in the proposed approach can be verified. The setting of hyper-parameters of these NNs is shown in Table 2.

Figure 3 presents the fitting curves of six single models and their combination models on the partial testing set of Datasets A-D, as well as their partial enlarged detail. It can be clearly seen that the forecasting performance of different single models is different, especially the fluctuation in Dataset B and Dataset C is dramatic. By combining these models together, the fluctuation is effectively mitigated and the prediction accuracy is improved.

Figure 4 presents partial combination weights of the combination models for Datasets A-D. It can be observed that the combination weights obtained using the mean function are the same for each model, the combination weights obtained using the softmax function are different for each model, but are the same for each prediction time under the same model. With respect to the combination weights obtained using the GM function are not only different for each model, but also vary with the prediction time. These indicate that the GM function can effectively grasp the alterations of input data over time and balance the significance of each model in the combination. Therefore, it can be concluded that the combination model constructed by the GM function is more advantageous overall than that constructed by other combination methods.

Table 3 present the detailed evaluation indicators of the single models and combination models for Datasets A-D. It can be observed that the forecasting performances of different indicators vary in different models. In all the Datasets, the ESN model has the best overall performance compared with other single models. For example, the appropriate values of the ESN on the RMSE, MAE and U1 in Dataset D are 744.47, 455.45 and 0.085, respectively. In addition, the forecasting evaluation indicators of these models in terms of the RMSE in ascending order are the ESN, GRU, MLP, SVR, KR and LSTM. The forecasting performance of these models in terms of the MAE are ESN, GRU, MLP, KR, LSTM and SVR. The outstanding result of the ESN can be retrospected to the fact that it can learn the nonlinear relationship and capture dynamic information over time.

Concerning the forecasting performance of the combination models, it can be clearly observed that compared with single models, the overall performance of them has been significantly improved. Take Dataset D as an example, by comparison with the indicator values of the KR model, the RMSE, MAE and U1 values of the Comb_gm are reduced by 8.19%, 6.35% and 8.19%, respectively. Moreover, the Comb_gm has the best overall performance compared with other single models in Dataset D, the corresponding values on the RMSE, MAE and U1 are 737.88, 455.75 and 0.0843. Simultaneously, it is found that the Comb_gm exhibited the best performances for different Datasets from the combined methods. This is mainly because combination model can effectively consider the characteristics of different models and combine the forecasting results of multiple combined members into one optimal result. Although Datasets A-C are collected from different places than Dataset D, with different time intervals and different data characteristics, the above conclusion can also be drawn from Datasets A-C.

**Remark.** From the above comparisons, for different datasets, the combination models have better forecasting performance than most of their combination members, indicating that the construct combination model has good generalization capability and can better execute the forecasting task. Especially, the GM function has advantage over other combination functions used in this paper.

3.3. Experiment II

The intention of this experiment is to verify the advantage of using anchored ensembles and the GM function to construct a combination model. By introducing four benchmark models (i.e., the MLP, GRU, LSTM and ESN), the anchored ensemble is used for sampling of them to generate several single models, hence the final combination model can be constructed, respectively. The one-step-ahead prediction evaluation metrics of the proposed combination model and comparative models for Datasets A-D are presented in Tables 4, 5, 6, 7, 8, 9, 10, and 11. It should be noted that Model #i, represents the single model generated by sampling and Comb _#ij#f#i represents that the combination model is composed of Model #i to Model #j (i, j = 1, 2, ...), with the combination method is f (#f #s, m, gm).

Tables 4, 5, 6, and 7 present the forecasting performance of the combination models composed of six combination members for Datasets A-D. It is obvious that there is a certain difference in the forecasting result between the single models generated by sampling of the benchmark model, which indicates that there exists instability when using a single model for wind power forecasting. With respect to the combination models consisted of these single models, their forecasting performance have been significantly improved. Take the forecasting performance of the MLP models in Dataset A as an example, the RMSE, MAE and U1 values of the Comb #1–6_gm are 5.04%, 6.63% and 4.67% smaller than the corresponding index values of Model #1.

Furthermore, combined with Tables 3, 4, 5, 6, and 7, it can be observed that the combination models outperform the relevant base models. Take the MLP model as an example, by comparison with the index values of the MLP presented in Dataset A of Table 3, the RMSE, MAE and U1 values of the corresponding Comb #1–6_gm in Table 4 are reduced by 3.87%, 9.31% and 3.14%, respectively. Such improvement of the performance show from another angle that the construction of a combination model is of great significance to improve the forecasting ability.

As for the combined size of the combination model, compare Tables 8, 9, 10, and 11 with Tables 4, 5, 6, and 7, it is obvious that with the increase of combination members, the forecasting performance of each combination model is improved with different degrees. Take the forecasting performance of the ESN models in Dataset C as an example, the RMSE, MAE and U1 values of the Comb #1–10_gm in Table 10 are 1.63%, 16.39% and 1.62.% smaller than the corresponding index values of the Comb #1–6_gm in Table 6.

Figures 5 and 6 show the fitting curves of the combination models composed of six and ten ESNs for Datasets A-D, respectively. It can be observed that the forecasting behavior of each combination member is diverse, and the overall fitting curves of the combination models are closer to the real value than that of each combination member, which indicates that the combination of models can fully exploit the effective information of each member and lead to the enhancement of generalization ability and the improvement of prediction accuracy.

Figure 7 illustrate the prediction intervals of the ESNs combination model for Datasets A–D, respectively. The provided high-quality prediction intervals indicate that the proposed approach could capture the uncertain characteristics of wind power properly and has a good capability of wind power forecasting. What is noteworthy is that the combination model generated by anchored ensembling belongs to the category of Bayesian, wherein the mean and variance of model can be derived from Bayesian inference rather than heuristic method.
Especially, compared with the combination methods, regardless of the number of ensembles or different base models in Tables 4, 5, 6, 7, 8, 9, 10, and 11, it shows that the proposed combination method of the GM can provide better performance than some other methods since it has the most significant values of all evaluation metrics. For example, with respect to the ESN-based model in Dataset A, the RMSE, MAE and U1 values of the Comb #1–6_gm in Table 4 are 6127.59, 3099.00 and 0.0639, respectively. Additionally, Figure 8 presents the partial combination weights of the combination models composed of ten ESNs for Datasets A-D. It can be clearly observed that the weights of the softmax and mean functions are static, while the weights of the GM are dynamic. The changing weights indicate that the GM function can effectively mine the positive information of each single model and capture the variation characteristics of wind power.

By comparing the benchmark models with each other, it is observed that the ESN-based models outperform other models both for single models and combination models, which demonstrates that for the dataset applied in this experiment, the ESN possesses an absolute advantage over other models in forecasting. In contrast, the LSTM-based models do not perform well overall and are not competitive, which is a consequence of the lack of long short-term memory in the time series data used in the experiment. For instance, for Dataset A, the index values of the best overall performing single models of the ESN on the RMSE, MAE and U1 are 6129.47, 3122.17 and 0.0639, separately, and the relevant values of the LSTM are 6329.11, 3392.32 and 0.0662, respectively. Figure 9 presents the fitting curves of the combination models of four different benchmark models for Datasets A–D. It can be seen that compared with the LSTMs combination model, the fluctuation trend of the ESNs combination model is more in line with that of the real wind power value as a whole.

Remark. In accordance with the above analysis, it is observed that the proposed forecasting method can generate more satisfactory results than other single and combination models, the contribution of using dynamic weight approach and the anchored ensembles have been approved. Notably, the combined size is expected to be as small as possible while producing a good approximation. Bigger combined size will provide better accuracy, but at a higher cost of computation. Thus, the combined size should be carefully considered according to the actual requirement.

4. Conclusions

In order to generate accurate and stable prediction of wind power, this paper proposed a new dynamic combined probabilistic forecasting method, which is constructed based on the ESN, the anchored ensembles and the GM function. The proposed method successfully compensates for the defects of the existing literature in wind power forecasting and validates its effectiveness through the simulation application on four data-sets. The main conclusions of this paper are:

1) A single model cannot fully grasp the characteristics of wind power generation, while the proposed model can.

2) The results of point forecasting are intuitive and easy to understand, and the result of probabilistic forecasting can provide more uncertain information about wind power forecasting, which can be realized through anchored ensembling.

3) Compared with the static weighted combination forecasting, dynamic weighted combination forecasting can better capture the positive information of each model and improve combination model forecasting ability and flexibility.

4) The proposal outperforms the benchmark models, and this can be attributed to the fact that it generates several ESNs using the anchored ensembles to execute wind power forecasting and then dynamically combines the forecasting result using the GM function.

In the future research, we will focus on the research which related to online learning and multi-step-ahead prediction based on the proposed Bayesian ensembling method, which is possible to extend the developed model in the fields of other renewable energies.

Declarations

Author contribution statement

Jingwei Tang: Conceived and designed the experiments; Performed the experiments; analyzed and interpreted the data; wrote the paper.

Jianming Hu: conceived and designed the experiments; performed the experiments; contributed reagents, materials, analysis tools or data; wrote the paper.

Jiani Heng and Zhi Liu: analyzed and interpreted the data; wrote the paper.

Funding statement

Ph.D. Jianming Hu was supported by Natural Science Foundation of Guangdong Province [2020A151501527], Joint Foundation of Guangzhou city and Guangzhou University [202210102228]. Jianming Heng was supported by National Natural Science Foundation of China [72103186].

Data availability statement

The authors do not have permission to share data.

Declaration of interest's statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

[1] Global Wind Energy Council, GWEC: Global Wind Report 2022. https://gwec.net/global-wind-report-2022/.

[2] J.C. Smith, M.R. Milligan, E.A. Demeo, B. Parsons, Utility wind integration and operating impact state of the art, IEEE Trans. Power Syst. 22 (2007) 900–908.

[3] S. Salcedo-Sanz, M. Pérez-Bellido A, E.G. Ortiz-Garcia, et al., Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction, Renew. Energy 34 (6) (2009) 1451–1457.

[4] M.A. Prisner, C. Otero-Casal, P.C. Fernández, G. Mínguez-Macho, Wind power forecasting for a real onshore wind farm on complex terrain using WRF high resolution simulations, Renew. Energy 135 (2019) 674–686.

[5] N. González-Alonso de Linaje, C. Matur, D. Borrvaran, Quantifying the wind energy potential differences using different WRF initial conditions on Mediterranean coast of Chile, Energy 188 (2019), 116027.

[6] D.E.K. Dzebre, J. Amofo, M.S. Adaramola, An assessment of high-resolution wind speeds downscaled with the Weather Research and Forecasting Model for coastal areas in Ghana, Heliyon 7 (8) (2021), e07768.

[7] A. Dupré, P. Drobiński, B. Alonzo, J. Badros, C. Briend, R. Plougonven, Sub-hourly forecasting of wind speed and wind energy, Renew. Energy 145 (2020) 2373–2379.

[8] J. Naik, P. Sarapathy, P.K. Dash, Short-term wind speed and wind power prediction using hybrid empirical mode decomposition and kernel ridge regression, Appl. Soft Comput. 70 (2018) 1167–1188.

[9] J. Dowell, P. Pinson, Very-short-term probabilistic wind power forecasts by sparse vector autoregression, IEEE Trans. Smart Grid 7 (2) (2015) 763–770.

[10] P. De Falco, L.P. Di Noia, R. Rizzo, Exponential smoothing model for photovoltaic power forecasting[C]//2021 9th international conference on modern power systems (MPS), IEEE (2021) 1–5.

[11] F. Rodríguez, A.M. Florez-Tapia, L. Fontán, A. Galarza, Very short-term wind power density forecasting through artificial neural networks for microgrid control, Renew. Energy 145 (2020) 1517–1527.

[12] E. Ogliari, M. Guiltizoni, A. Giglio, et al., Wind power 24-h ahead forecast by an artificial neural network and an hybrid model: comparison of the predictive performance, Renew. Energy 178 (2021) 1466–1474.
[13] A. Kiwarvi, Z. Lin, X. Liu, Wind power forecasting: A data-driven method along with gated recurrent neural network, Renew. Energy 163 (2021) 1895–1909.

[14] M. Zhou, B. Wang, S. Guo, et al., Multi-objective prediction intervals for wind power forecast based on deep neural networks, Inf. Sci. (2020).

[15] L.L. Li, Z.F. Liu, M.L. Tseng, et al., Using enhanced crow search algorithm optimization-extreme learning machine model to forecast short-term wind power, Expert Syst. Appl. 184 (2021), 115579.

[16] Q. Li, X. Zhang, T. Ma, et al., A multi-step ahead photovoltaic power prediction model based on similar day, enhanced colliding bodies optimization, variational mode decomposition, and deep extreme learning machine, Energy 224 (2021), 120994.

[17] Z.H. Guo, J. Wu, H.Y. Lu, J.Z. Wang, A case study on a hybrid wind speed forecasting method using BP neural network, Knowl. Base Syst. 24 (2011) 1048–1056.

[18] A. Tascikaraoglu, M. Uzunoglu, A review of combined approaches for prediction of short-term wind speed and power, Renew. Sustain. Energy Rev. 34 (2014) 243–254.

[19] P. Du, J. Wang, W. Yang, T. Niu, A novel hybrid model for short-term wind power forecasting, Appl. Soft Comput. 80 (2019) 93–106.

[20] L.L. Li, X. Zhao, M.L. Tseng, R.R. Tan, Short-term wind power forecasting based on support vector machine with improved dragonfly algorithm, J. Clean. Prod. 242 (2020), 118447.

[21] C. Wu, J. Wang, X. Chen, P. Du, W. Yang, A novel hybrid system based on multi-objective optimization for wind speed forecasting, Renew. Energy 146 (2020) 149–165.

[22] L. Luo, H. Li, J. Wang, et al., Design of a combined wind speed forecasting system based on decomposition-ensemble and multi-objective optimization approach, Appl. Math. Model. 89 (2021) 49–72.

[23] X.X. Lv, L. Wang, Deep learning combined wind speed forecasting with hybrid time series decomposition and multi-objective parameter optimization, Appl. Energy 311 (2021), 117674.

[24] W. Zheng, R. Pestana, S.-L. Feng, H. Shen, L. Rosa, Short-term wind power combination forecasting method based on dynamic coefficient updating, Power Syst. Technol. 41 (2) (2017) 500–507.

[25] Y. Wang, J. Wang, Z. Li, et al., Design of a combined system based on two-stage data preprocessing and multi-objective optimization for wind speed prediction, Energy 231 (2021), 121125.

[26] S. Wang, J. Wang, H. Lu, et al., A novel combined model for wind speed prediction—Combination of linear model, shallow neural networks, and deep learning approaches, Energy 234 (2021), 121275.

[27] W. Zhang, Z. Qu, K. Zhang, W. Mao, Y. Ma, X. Fan, A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting, Energy Convers. Manag. 136 (2017) 439–451.

[28] A. Altant, S. Karasu, E. Zio, A new hybrid model for wind speed forecasting combining long short-term memory neural network, decomposition methods and grey wolf optimizer, Appl. Soft Comput. 100 (2021), 106996.

[29] J. Wang, Y. An, Z. Li, et al., A novel combined forecasting model based on neural networks, deep learning approaches, and multi-objective optimization for short-term wind speed forecasting, Energy 251 (2022), 122960.

[30] J. Wang, J. Heng, L. Xiao, C. Wang, Research and application of a combined model based on multi-objective optimization for multi-step ahead wind speed forecasting, Energy 125 (2017) 591–613.

[31] J. Heinemann, O. Kramer, Machine learning ensembles for wind power prediction, Renew. Energy 89 (2016) 671–679.

[32] P. Jiang, C. Li, Research and application of an innovative combined model based on a modified optimization algorithm for wind speed forecasting, Measurement 124 (2018) 395–412.

[33] V.S. Costa, A.D.S. Farias, B. Bedregal, R.H.N. Santiago, AMDp, Canuto, Combining multiple algorithms in classifier ensembles using generalized mixture functions, Neurocomputing 313 (2018) 402–414.

[34] P. Jiang, Z. Liu, Variable weights combined model based on multi-objective optimization for short-term wind speed forecasting, Appl. Soft Comput. 82 (2019), 105587.

[35] H. Li, J. Wang, H. Lu, Z. Guo, Research and application of a combined model based on variable weight for short term wind speed forecasting, Renew. Energy 116 (2018) 669–684.

[36] W. Zhao, J. Wang, H. Lu, Combining forecast of electricity consumption in China with time-varying weights updated by a high-order Markov chain model, Omega 45 (2014) 80–91.

[37] J. Wang, N. Zhang, H. Lu, A novel system based on neural networks with linear combination framework for wind speed forecasting, Energy Convers. Manag. 181 (2019) 425–442.

[38] J. Duan, P. Wang, W. Ma, et al., A novel hybrid model based on nonlinear weighted combination for short-term wind power forecasting, Int. J. Electr. Power Energy Syst. 134 (2022), 107452.

[39] S. Yin, H. Liu, Wind power prediction based on outlier correction, ensemble reinforcement learning, and residual correction, Energy 250 (2022), 123857.

[40] C. Chen, H. Liu, Dynamic ensemble wind speed prediction model based on hybrid deep reinforcement learning, Adv. Eng. Inf. 46 (2021), 101290.

[41] T. Pearce, M. Zaki, A. Britnup, N. Anastassacos, A. Neely, Uncertainty in Neural Networks: Bayesian Ensembling, 2018.

[42] B. Laksminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, Adv. Neural Inf. Process. Syst. (2017) 6405–6416.

[43] I. Osband, J. Aalst, A. Casielles, Randomized prior functions for deep reinforcement learning, in: 32nd Conference on Neural Information Processing Systems (NIPS 2018), 2018.

[44] J.M. Bardsey, A. Solonen, H. Haario, M. Laine, Randomize-then-optimize: a method for sampling from posterior distributions in nonlinear inverse problems, SIAM J. Sci. Comput. 36 (4) (2014) 1895–1910.

[45] H. Jaeger, The “echo State” Approach to Analysing and Training Recurrent Neural Networks—With an Erratum Note, 2001.

[46] C. Du, F. Cai, M.A. Zidan, W. Ma, S.H. Lee, W.D. Lu, Reservoir computing using dynamic memristors for temporal information processing, Nat. Commun. 8 (1) (2017) 2204.

[47] M. Lukovnicic, H. Jaeger, Reservoir computing approaches to recurrent neural network training, Comput. Sci. Rev. 3 (3) (2009) 127–149.

[48] W. Maass, T. Natschlager, H. Markram, Real-time computing without stable states: a new framework for neural computation based on perturbations, Neural Comput. 14 (11) (2002) 2531–2560.

[49] M.R.A. Pereira, G. Pan, On non-monotonic aggregation: mixture operators, in: Proceedings of the 4th Meeting of the EURO Working Group on Fuzzy Sets (EUROFUSE 1999) and 2nd International Conference on Soft and Intelligent Computing (SIC 1999), 1999, pp. 513–517.

[50] R.A.M. Pereira, The orness of mixture operators: the exponential case, in: Proceedings of the 8th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2000), 2000.

[51] H. Witten Ian, Eibe Frank, Data Mining Practical Machine Learning Tools and Techniques, second ed., Morgan Kaufmann Press, 2011.