

Article

Multi-Step Ahead Wind Power Generation Prediction Based on Hybrid Machine Learning Techniques

Wei Dong ¹, Qiang Yang ^{1,2,*}  and Xinli Fang ^{1,3}

¹ College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China; dong_w@zju.edu.cn (W.D.); fxlily2008@163.com (X.F.)

² Jiangsu Key Construction Laboratory of IoT Application Technology, Taihu University of Wuxi, Wuxi 214064, China

³ Power China Hua Dong Engineering Corporation Limited, Hangzhou 311122, China

* Correspondence: qyang@zju.edu.cn; Tel.: +86-151-6713-8974; Fax: +86-571-8795-1625

Received: 11 July 2018; Accepted: 25 July 2018; Published: 30 July 2018



Abstract: Accurate generation prediction at multiple time-steps is of paramount importance for reliable and economical operation of wind farms. This study proposed a novel algorithmic solution using various forms of machine learning techniques in a hybrid manner, including phase space reconstruction (PSR), input variable selection (IVS), K-means clustering and adaptive neuro-fuzzy inference system (ANFIS). The PSR technique transforms the historical time series into a set of phase-space variables combining with the numerical weather prediction (NWP) data to prepare candidate inputs. A minimal redundancy maximal relevance (mRMR) criterion based filtering approach is used to automatically select the optimal input variables for the multi-step ahead prediction. Then, the input instances are divided into a set of subsets using the K-means clustering to train the ANFIS. The ANFIS parameters are further optimized to improve the prediction performance by the use of particle swarm optimization (PSO) algorithm. The proposed solution is extensively evaluated through case studies of two realistic wind farms and the numerical results clearly confirm its effectiveness and improved prediction accuracy compared to benchmark solutions.

Keywords: multi-step ahead prediction; phase space reconstruction; input variable selection; K-means clustering; neuro-fuzzy inference system; wind power prediction

1. Introduction

Currently, the urgent pursuit of low-carbon economy and advances of wind power technologies are strongly driving the rapid sustainable transition in the energy sector as well as the wind power development across the world [1,2]. Due to the intermittent and stochastic nature, the power generation of wind farms needs to be accurately predicted at different time-scales (e.g., daily, hourly or even less) and timely reported to the dispatch center. Accurate short-term wind power forecasting can improve wind power utilization, increase system reliability, reduce operational cost and allow flexible dispatch strategies [3]. In particular, the ultra-short-term wind power prediction for a couple of hours ahead with the small prediction cycle (e.g., 15 min) can provide strong support for frequency modulation and spinning reserve optimization. However, such ultra-short-term prediction is often a non-trivial task that demands advanced algorithmic solutions and tools with sufficient accuracy and acceptable computational complexity.

In the literature, much research effort has been made to address the prediction issues from different aspects, e.g., electricity pricing [4,5] and power generation [6], in energy systems. The available prediction tools and solutions of ultra-short-term prediction can be categorized into three classes: physical-based methods, statistical-based methods, and machine learning-based methods.

The physical-based methods focus on the spatial and temporal factors in a full fluid-dynamics atmosphere model [7], which can generally perform well in longer horizons. The statistical-based methods, e.g., autoregressive model (AR), moving-average (MA), auto-regressive integrated moving average (ARIMA) and Kalman filters, carry out statistical analysis based on the historical data to identify the internal regularity and the tendency of variations to deduce the prediction results. The machine learning-based methods include several supervised learning models, e.g., artificial neural networks (ANNs) [8], support vector machines (SVMs) [9], adaptive neuro-fuzzy inference systems (ANFISs) [10] and Gaussian processes (GPs) [11]. In addition, some combined or hybrid methods have been proposed aim to improve the prediction performance. For example, a hybrid intelligent method was proposed in [3] using multiple support vector regression (SVR) models with the parameters estimated based on an enhanced harmony search (EHS) algorithm. In [12], a hybrid forecasting model based on K-means clustering and an a priori algorithm was developed for short-term wind speed prediction and the prediction errors are corrected by associated rules.

In recent years, the feature selection and analysis of machine-learning based models for wind power prediction have received much attention. The prediction accuracy can be improved by exploring the information obtained from of the historical wind speed and generation time series data. In [13], the features are firstly extracted from the historical power generation data, and then the dataset is split into subsets based on the stationary patterns. In [14], a novel decomposition approach to fully consider the chaotic nature of wind power time series was proposed. The time series data were separated into different frequency characteristics using ensemble empirical mode decomposition (EMD) before carrying out the chaotic time series analysis and singular spectrum analysis (SSA). A forecasting model combining a support vector machine (SVM) optimized by a genetic algorithm and feature selection based on the phase space reconstruction was presented in [15] for short-term wind speed prediction. In addition, numerical weather prediction (NWP) data (including wind speed, direction, temperature, humidity, atmospheric pressure, etc.) were adopted as the input variables for supervised models. In [16], output data from different NWP models were used and the data with the minimum training error were selected to be used in both ANN and SVM models. Afterwards, the forecasting errors were corrected based on the model output statistics (MOS). The study in [17] proposed a wind power prediction model based on the composite covariance function considering the joint effects among features of NWP data. In [18], a data-driven feature extraction approach was developed to utilize unlabeled NWP data which can be used in the supervised forecasting models.

It should be noted that, as the dimensionality of input variable increases, irrelevant and redundant variables can deteriorate the prediction performance. Therefore, the selection of appropriate variables through dimensionality reduction approaches is needed [19], and two dimensionality reduction techniques, feature selection (variable selection) and feature extraction (feature transform), are often used [20]. The latter can produce a new feature space through mapping the original features into lower dimensional ones, e.g., singular value decomposition (SVD), principal component analysis (PCA) and locally linear embedding (LLE). However, such feature extraction methods may often lose physical properties of the original variables, and also difficult to be interpreted. In time series data analysis, the variable selection that filters out some meaningless attributes without any transformation can be more attractive than feature extraction. The variable selection methods can select the compact subset from the original dataset to improve the performance and interpretability of the prediction model.

In general, three types of feature selection methods, filter methods, wrapper methods and embedded methods, are considered [21]. The filter method ranks the input variables with a correlation or mutual information (MI) criteria and selects the variable with the highest ranking. It is effectively a pre-processing step before the development of the predictive model [21]. The wrapper method identifies and evaluates the subsets of input variables based on the accuracy contributed by the given output variable. Similarly, the embedded one builds the close-loop search into a classifier construction in the training process [22]. The wrapper and embedded techniques can generally achieve better

accuracy than the filter technique, but the filter method is less likely to lead to over-fitting and with less computational complexity [23].

It should be highlighted that the existing solutions either have not been able to fully consider the available information (e.g., historical data and NWP) or select the appropriate variables for improving the prediction accuracy. To the author's best knowledge, the technical challenge of ultra-short-term prediction of wind power generation remains and the hybrid approach based on data mining and machine learning techniques has not been thoroughly exploited.

This paper attempts to address the challenge of ultra-short-term power generation prediction in wind farms. The main technical contributions made in this work are summarized as follows: a novel hybrid algorithmic solution is presented which considers both historical generation data and NWP data, and selects the optimal combination of input features using a filter method for different prediction steps. The proposed algorithmic solution was extensively evaluated and validated through case studies of two realistic wind farms. The basic idea behind the proposed prediction solution is illustrated in Figure 1. The long-term nonlinear dynamic characteristics of wind power time series data are extracted and recovered by using phase space reconstruction in C_C method. Afterwards, the most appropriate input variables are selected from the reconstructed phase and NWP features with respect to different forecasting steps based on the minimal redundancy and maximal relevance criterion. Finally, an adaptive neuro-fuzzy inference system based algorithmic solution with heuristically optimized parameters is adopted by using the selected input variables to produce the prediction results.

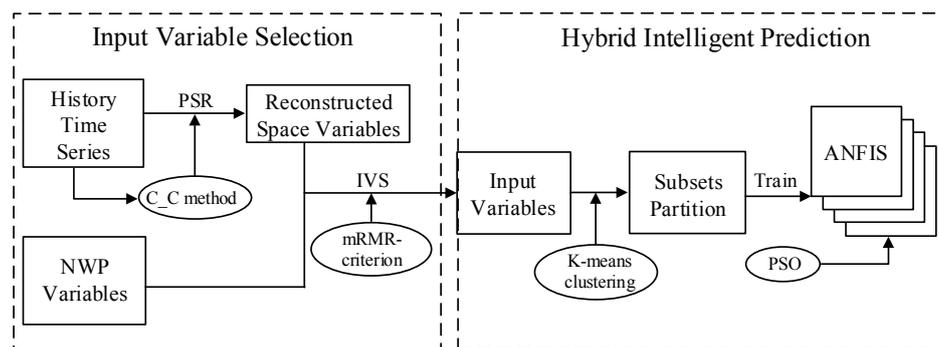


Figure 1. The framework of the proposed hybrid algorithmic solution.

The rest of this paper is organized as follows: Section 2 describes the input variable selection (IVS) solution based on (PSR) technique and mRMR criterion. Section 3 presents the framework of the proposed hybrid intelligence prediction model. Section 4 carries out the case studies and presents a set of key numerical results. Finally, the conclusions are given in Section 5.

2. Input Variable Selection (IVS)

Due to the chaotic property of the weather system, the evolution of dynamic characteristics has initial sensitivity. The correlation between historical time series and future wind power generation will decay rapidly with the increase of forecasting time step, and even deteriorate the prediction performance. Thus, the adoption of both historical generation data and the numerical weather prediction (NWP) data as the input variables is required. NWP aims to predict the variation of weather through solving the process of thermodynamics and hydrodynamics equations based on the meteorological data of the system. However, it can only provide short-term surface wind and other weather characteristics prognoses roughly, which are not entirely adequate for specific local conditions [24]. NWP data are often adopted to provide ancillary information, e.g., wind speed, wind direction, temperature, humidity and air pressure, for prediction. For different step-ahead prediction, these input variables can have different impacts on the forecasting targets. Therefore, a two-stage input variable selection is used in the proposed prediction solution. Firstly, the initial

variables can be selected from the historical series data through the phase space reconstruction technique (PSRT), and the initial variables further can be combined with NWP information as the candidate inputs. Secondly, the optimal input variables are filtered based on mRMR criterion.

2.1. The Initial Input Variable Selection of Historical Series Using PSRT

The Lyapunov exponent can be used to prove that the wind power generation time series has chaotic characteristics. Therefore, the nonlinear dynamic characteristics of wind power time series can be extracted and recovered by using phase space reconstruction theory. In [25], the time-delay technique was used to reconstruct a finite dimensional phase space of sampled system’s time evolution. In the time delay coordinate reconstruction, it is not only very important but also difficult to choose an appropriate time delay τ and a good embedding dimension m since real datasets are finite and noisy. There are currently two different viewpoints for the estimation of the aforementioned parameters. One holds that they are irrelevant and should be chosen independently. To choose time delay τ , one can use methods including autocorrelation function, multiple-autocorrelation, mutual information, and so on. G-P algorithm or False Nearest Neighbor can be used to find the embedding dimension m . However, it is suggested that the delay time and embedding dimension are dependent mutually. The delay time window ω should be estimated for the choice of m and τ . The delay time window ω can be estimated using C_C method [26]. The C_C method was used to determine the optimal input variables from the historical generation with reduced computational complexity and enhanced efficiency [27].

The phase space reconstruction is an efficient tool to analyze the dynamic pattern of a chaotic time series data. The delay-coordinate method was presented by Takens et al. to perform the phase space reconstruction. The time series $x = \{x_i, i = 1, 2, \dots, N\}$ can be reconstructed into a multi-dimensional phase space $X = \{X_i\}$ to represent the dynamic system, according to:

$$X_i = [x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}] \tag{1}$$

where $i = 1, 2, \dots, M, M = N - (m - 1)\tau, m$ is the embedding dimension, and τ is the delay time. In this study, the C_C method [26] was constructed via two correlation integrals, developed to reconstruct the given time series $x = \{x_i\}$ to simplify the candidate input forms.

As suggested in [26], the correlation integral for the embedded time series is defined as:

$$C(m, N, r, t) = \frac{2}{M(M-1)} \sum_{1 \leq i < j \leq M} \theta(r - d_{ij}), r > 0 \tag{2}$$

where $d_{ij} = \|x_i - x_j\|_\infty, \|\cdot\|_\infty$ represents the infinite norm, t denotes the index lag, and θ is the Heaviside function, $\theta(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$. The correlation integral is a cumulative distribution function, which denotes the probability with the distance less than search radius r between any two points in the phase space. To study the nonlinear dependence and eliminate spurious temporal correlations, the given time series $x = \{x_i\}$ must be divided into t disjoint sub-sequences. The disjoint time series can be expressed as (3):

$$\begin{aligned} & \left\{ x_1, x_{1+t}, \dots, x_{1+(l-1)t} \right\} \\ & \left\{ x_2, x_{2+t}, \dots, x_{2+(l-1)t} \right\} \\ & \dots \dots \\ & \left\{ x_t, x_{2t}, \dots, x_{lt} \right\} \end{aligned} \tag{3}$$

where l is the length of subseries, $l = INT(N/t)$, and $INT(\cdot)$ denotes reserving integer of the value.

Let us construct a statistic function $S(m, N, r, t)$ as the serial correlation of a nonlinear time series, which is a dimensionless measure of nonlinear dependence. For general t in the above disjoint time series expressed in Equation (3), $S(m, N, r, t)$ is defined as Equation (4):

$$S(m, N, r, t) = \frac{1}{t} \sum_{s=1}^t \left[C_s \left(m, \frac{N}{t}, r, t \right) - C_s^m \left(1, \frac{N}{t}, r, t \right) \right] \quad (4)$$

Finally, when $N \rightarrow \infty$, the following can be obtained:

$$S(m, r, t) = \frac{1}{t} \sum_{s=1}^t [C_s(m, r, t) - C_s^m(1, r, t)] \quad (5)$$

If the time series data follows an independent and identical distribution, $S(m, r, t)$ is equal to zero constantly for fixed value m, t and $N \rightarrow \infty$. However, as the real dataset is finite and the components of series are correlated, $S(m, r, t)$ is non-zero [26]. The maximum deviation of $S(m, r_j, t)$ for all radius r can be defined as (6):

$$\Delta S(m, t) = \max \{ S(m, r_j, t) \} - \min \{ S(m, r_j, t) \} \quad (6)$$

Here, N, m and r can be estimated based on the Brock–Dechert–Scheinkman (BDS) statistics as $N = 3000, m = 2, 3, 4, 5, r_i = i\sigma/2, i = 1, 2, 3, 4$, respectively, where $\sigma = std(x)$ denotes the standard deviation of the time series. Then, Equations (7)–(9) can be obtained as:

$$\bar{S}(t) = \frac{1}{16} \sum_{m=2}^5 \sum_{j=1}^4 S(m, r_j, t) \quad (7)$$

$$\Delta \bar{S}(t) = \frac{1}{4} \sum_{m=2}^5 \Delta S(m, t) \quad (8)$$

$$S_{cor}(t) = \Delta \bar{S}(t) + |\bar{S}(t)| \quad (9)$$

The optimal delay time τ is determined when the value of $\bar{S}(t)$ first reaches zero or when $\Delta \bar{S}(t)$ reaches the first minimum point. The optimal embedding window ω corresponds to the global minimum point of S_{cor} . Furthermore, the embedding dimension m can be obtained by the following formula: $m = \omega / \tau + 1$.

Once the reconstruction parameters, delay time τ and embedding dimension m , are determined by C_C method, the initial variables related to the historical sequence are obtained as $His_{Inputs}(t) = [x(t), x(t - \tau), \dots, x(t - (m - 1)\tau)]$, where $x(t)$ is the power generation values observed at current time t . The forecasting weather variables provided by NWP can be written as $NWP_{Inputs}(t') = [V_{wind}(t'), (\sin(D_{wind}(t')) + \cos(D_{wind}(t'))), T(t'), H(t'), P(t')]$, where $V_{wind}(t')$, $D_{wind}(t')$, $T(t')$, $H(t')$ and $P(t')$ in turn represent wind speed, wind direction, temperature, humidity and air pressure at the predicted time t' . Therefore, candidate input variables set can be combined into $V = [His_{Inputs}(t), NWP_{Inputs}(t')]$, and the input set dimension is $|V| = m + 5$.

2.2. The Optimal Selection of Candidate Input Variables Using mRMR-Criterion Ranking

The input variables of historical generation and NWP are selected using the minimal redundancy maximal relevance (mRMR) criterion based on mutual information (MI) [28]. As MI can measure both the linear and nonlinear dependency between variables, it has been applied for correlation measurement and variable selection [29]. The basic idea of variable selection algorithm based on MI is to select the best subset S from the original dataset X by maximizing the joint MI between S and target output Y , namely $I(S; Y)$. In the literature, many MI-based variable selection algorithms are available, e.g., mutual information feature selection (MIFS) [30], mutual information feature selection under uniform information distribution (MIFS-U) [31], the minimal redundancy maximal relevance (mRMR) [28], and normalized mutual information feature selection (NMIFS) [32]. In this work, the mutual-information-based mRMR criterion is adopted to find the compact and informative input space. The mRMR technique aims to find a subset of candidate variables with maximal dependency

(with respect to the target to be predicted) as well as minimal redundancy (between the variables in the subset). The concept of MI is based on entropy that is described as follows.

The entropy of a random variable indicates the required average amount of information to describe the random variable [33] which has been adopted in many studies [34,35]. The entropy of a discrete random variable $X = (x_1, x_2, \dots, x_N)$ is denoted by $H(X)$, where x_i refers to the possible values that X can take for discrete variable or the possible value range for continuous variable. $H(X)$ is defined as:

$$H(X) = -\sum_{i=1}^N p(x_i) \log(p(x_i)) \quad (10)$$

where $p(x_i)$ is the probability mass function.

For any two random variables, X and $Y = (y_1, y_2, \dots, y_M)$, the joint entropy is defined as:

$$H(X, Y) = -\sum_{j=1}^M \sum_{i=1}^N p(x_i, y_j) \log(p(x_i, y_j)) \quad (11)$$

where $p(x_i, y_j)$ is the joint probability mass function of X and Y . The conditional entropy of X given Y is defined as:

$$H(Y|X) = -\sum_{j=1}^M \sum_{i=1}^N p(x_i, y_j) \log(p(y_j|x_i)) \quad (12)$$

The conditional entropy is the amount of uncertainty left in Y when a variable X is introduced, so it is less than or equal to the entropy of both variables. The conditional entropy is equal to the entropy if, and only if, the two variables are independent. Mutual Information (MI) is the amount of information that both variables share, and is defined as:

$$I(X; Y) = H(Y) - H(Y|X) \quad (13)$$

MI can be expressed as the amount of information provided by variable X , which reduces the uncertainty of variable Y . MI is zero if the random variables are statistically independent. MI is symmetric, so:

$$I(X; Y) = I(Y; X) \quad (14)$$

The minimal-redundancy-maximal-relevance criterion (mRMR) aims to identify a compact subset of informative input variables by simultaneously considering the maximum relevance scheme and minimum redundancy. The simple combination of individually informative input variables does not necessarily achieve a good forecasting performance [28]. Therefore, both the informativeness of individual input variables and redundancy between them should be considered. Thus, the informativeness score for individual variable v_i based mRMR criterion is given by:

$$J(v_i, S) = I(v_i; t) - \frac{1}{|S|} \sum_{v_j \in S} I(v_i; v_j) \quad v_i \in (V - S) \quad (15)$$

where V is the total candidate variables, S is the selected input variables, and $|\cdot|$ is the number of variables. The mutual information $I(v_i; t)$ is used between the target t and the candidate input variable v_i to measure the strength of v_i relative to the forecasting process. The goal of second item is to optimally select those variables that reveal a minimum of resemblance or redundancy between them, thus making the selected set more representative or informative of the whole set.

In the implementation, a stepwise search, incremental forward selection (IFS) method, is used to select input variables according to Equation (15), in which greater $J(v_i, S)$ scores indicate more promising input variable v_i . In the first step, Max-Relevance score of all candidate input variable is calculated, where the variable with the maximum $I(v_i; t)$ score is determined as the first promising input variable:

$$v_1 = \arg \max_{v_i \in V} I(v_i; t) \quad (16)$$

The rest of the variables are selected step by step according to the criterion in Equation (15). In step m ($2 \leq m \leq |V|$), it is supposed that an input variable subset S_{m-1} , composed of $m - 1$ promising variables, s_1, s_2, \dots, s_{m-1} , that has been selected from previous step (e.g., step $m - 1$). The m -th promising variable can be selected from $V - S_{m-1}$ at step m by optimizing the following condition:

$$s_m = \arg \max_{v_i \in V - S_{m-1}} J(v_i, S_{m-1}) \quad (17)$$

As one input variable represents one step forward, the promising variables can be incrementally retrieved until step $|V|$ where a total of input variables V are selected. The variables are also ranked in selection process and the informativeness score (InSc) in m -th step is given by:

$$InSc_m = J(s_m, S_{m-1}) \quad m = 2, 3, \dots, |V| \quad (18)$$

where s_m is the most promising variable to be selected in m -th step according to Equation (17). Thus, the priority of candidate input variables can be ranked through the mRMR-based incremental forward selection (IFS) method. The cumulative amount of $InSc$, denoted as $CumInSc_m = \sum_{i=1}^m InSc_i$, indicates the information contributed from the newly added variable. The final optimal number of input variables can be determined according to the change trend of $CumInSc$.

3. Hybrid Intelligent Prediction Model

In this work, the proposed hybrid algorithmic solution combines the K-means clustering, Particle Swarm Optimization (PSO) and adaptive neuro-fuzzy interference system (ANFIS) in the prediction model.

3.1. Subsets Partition Using K-Means Algorithm

The obtained historical dataset is divided into subsets and the data in the same class are with the similar meteorological features. Consequently, the complexity of network training can be significantly reduced with improved generalization capability. The dataset partitioning is implemented using K-means algorithm as follows.

Here, data centralization and normalization are needed before clustering. Z-score standardization of the dataset is expressed as follows:

$$z_{i,j} = \frac{v_{i,j} - \mu_j}{\sigma_j} \quad i = 1, 2, \dots, N; \quad j = 1, 2, \dots, D \quad (19)$$

where μ_j is the mean of column i , σ_j is the standard deviation of column, N is the number of instances, and D is the dimensionality of input variables.

Given a training set $\{z_1, z_2, \dots, z_N\}$, the K-means clustering algorithmic can partition the dataset into k cohesive groups through an unsupervised learning process [36]. Here, $z_i \in \mathbb{R}^D$ and first choosing cluster centroids $\{a_1, a_2, \dots, a_k\} \in \mathbb{R}^D$ randomly. Then, the fundamental purpose of K-means algorithm is to minimize the following cost function:

$$C_N = \frac{1}{N} \sum_{i=1}^N \min_{1 \leq j \leq k} \|z_i - a_j\|^2 \quad (20)$$

where $\|\cdot\|$ is the function representing the usual Euclidean norm. After determining the cluster center, the training samples are grouped into the subsets of the nearest cluster centers. For k subsets, $U^l = \{u_1, u_2, \dots, u_{N_l}\}$, $l = 1, 2, \dots, k$, the number of samples in each subset is recorded as

$N_l, l = 1, 2, \dots, k$. Each subset U^l determines a set of independent ANFIS network parameters during training process. The vector nearest to the cluster center is adopted as the network input when performing the online prediction.

3.2. Adaptive Neuro-Fuzzy Inference System

The adaptive neuro-fuzzy inference system (ANFIS) is a data-driven modelling technique [37] to address the multivariable nonlinear system prediction through nonlinear neural network and adaptive fuzzy reasoning process. The fuzzy membership function and fuzzy rules of the system can be obtained through learning from historical data, rather than expert experience or intuition. Figure 2 illustrates the typical structure of two-input ANFIS model. The ANFIS is based on Takagi–Sugeno inference approach that creates a nonlinear mapping from input space to the output space through using the fuzzy if–then rules. The ANFIS is comprised of five layers as follows.

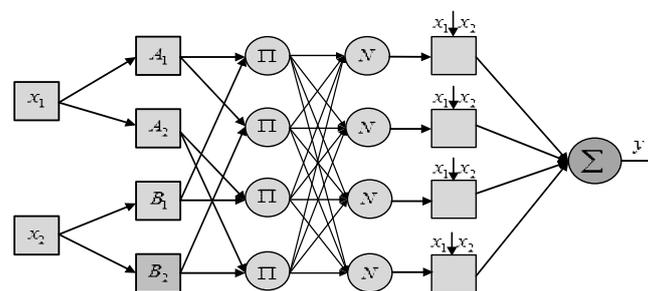


Figure 2. Typical two-input ANFIS structure.

Fuzzifier (Layer 1): Neurons in this layer perform the fuzzification operations, and the membership degree of input in different fuzzy sets (e.g., A_1, A_2, B_1 , and B_2) can be obtained. Fuzzification is represented by fuzzy membership function f , and the output membership degree μ_A, μ_B for x_1, x_2 can be expressed as:

$$\mu_{A_i}(x_1) = f(x_1); \mu_{B_j}(x_2) = f(x_2), i, j = 1, 2 \quad (21)$$

Rules inference (Layer 2): The rules neurons receive input from their respective fuzzified neurons and calculate the rules active intensity ω_n :

$$\omega_n = \mu_{A_i}(x_1)\mu_{B_j}(x_2), i, j = 1, 2, n = 1, 2, 3, 4 \quad (22)$$

Normalization (Layer 3): Each neuron in this layer receives all neuronal inputs from the previous layer and calculates the normalized active intensity for a given rule $\bar{\omega}_n$:

$$\bar{\omega}_n = \omega_n / \sum \omega_n \quad (23)$$

Defuzzifier (Layer 4): This layer computes the posteriori value with weight of given rule f_n :

$$\bar{\omega}_n f_n = \bar{\omega}_n (p_n x_1 + p_n x_2 + r_n) \quad (24)$$

Output (Layer 5): This layer sums all defuzzified neuron outputs to arrive at the final ANFIS output y :

$$y = \sum_n \bar{\omega}_n f_n = \frac{\sum_n \omega_n f_n}{\sum_n \omega_n} \quad (25)$$

In this work, ANFIS is adopted to fit individual sub-training sets with the parameters heuristically optimized by the use of Particle Swarm Optimization (PSO) algorithm. The structure and parameters

of ANFIS are firstly determined using the fuzzy C-Means clustering (FCM) algorithm [38]. Afterwards, PSO is adopted to optimize the parameters. Each particle in PSO can identify and maintain its locally optimal solution (Pbest), and also collectively search for the global optimal solution (Gbest) in the swarm [39]. The location and velocity function in PSO can be expressed as (26).

$$\begin{cases} p_i^d(t+1) = p_i^d(t) + v_i^d(t+1) \\ v_i^d(t+1) = w(t) \cdot v_i^d(t) + c_1 r_{i1} \cdot (pbest_i^d - p_i^d(t)) + c_2 r_{i2} \cdot (gbest^d - p_i^d(t)) \end{cases} \quad (26)$$

where r_{i1} and r_{i2} are two random variables in the range of $[0, 1]$, c_1 and c_2 are positive constants, and w is the inertia weight. $P_i = (p_i^1, p_i^2, \dots, p_i^n)$ and $V_i = (v_i^1, v_i^2, \dots, v_i^n)$ represents the position and velocity of the i -th particle, respectively $pbest_i = (pbest_i^1, pbest_i^2, \dots, pbest_i^n)$ and $gbest_i = (gbest_i^1, gbest_i^2, \dots, gbest_i^n)$ represent the best previous position of the i -th particle and the best previous position among all the particles in the population, respectively.

4. Performance Assessment and Numerical Result

To extensively verify the reliability of the proposed prediction solution, the performance assessment based on the data collected in two real wind farms with different locations and seasons are carried out: Anzishan wind farm (capacity of 45 MW, Henan, China, hub height of 70 m) in June 2017, and Xuqiao wind farm (capacity of 94 MW, Anhui, China, hub height of 90 m) in December 2017. The reference height of these two wind farms is 10 m. The power generation (directly measured) and NWP data with a 15-min interval (e.g., 96 observation values per day) are obtained from these two wind farms. The three-month data (about 8640 observation values) prior to the test month, March to May 2017 and September to November 2017, are used in the training process for Anzishan farm and Xuqiao farm, respectively. The proposed hybrid solution is implemented in the MATLAB (version 8.3, MathWorks, Natick, MA, USA) programming environment.

4.1. Input Variable Selection Process

The time series data of power generation in the previous month are used to determine the phase-space reconstruction parameters through the C_C method. With different spatial and temporal characteristics, the parameters of each wind farm will be determined, respectively. For Anzishan farm, it can be observed in Figure 3 that ΔS have the first local minimum when t is equal to 26, thus the optimal delay time τ is set to 26. The global minimum point of S_{cor} corresponds to the optimal embedding window $\omega = 52$ as shown in the figure. Thus, the embedding dimension m can be calculated as 3. As for Xuqiao farm, the optimal delay time τ is also equal to 26 as shown in Figure 4. The embedding window ω is observed as 134, thus the optimal embedding dimension m is 6.

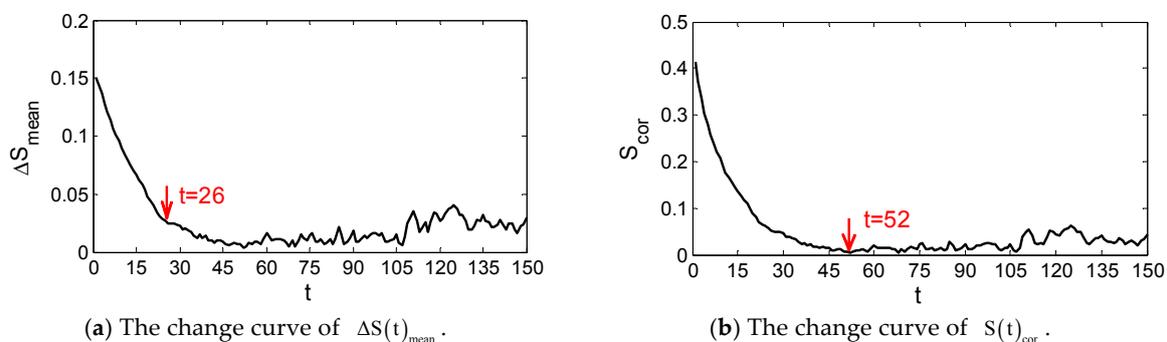


Figure 3. Results of $\Delta S(t)_{mean}$ and $S(t)_{cor}$ produced by the C_C method in Anzishan farm.

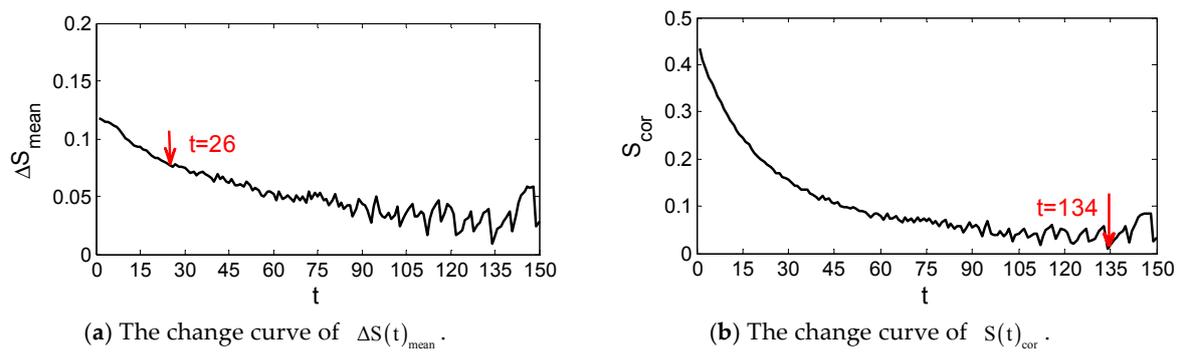


Figure 4. Results of $\Delta S(t)_{\text{mean}}$ and $S(t)_{\text{cor}}$ produced by the C_C method in Xuqiao farm.

The historical time series data can be selected as m -dimension input variables using the phase-space reconstruction. For the current time t , the initial input variables related to the historical data are $His_{Inputs}(t) = [x(t), x(t - \tau), \dots, x(t - (m - 1)\tau)]$, where $x(t)$ is the power generation values observed at current time t . The m -dimensional input variables are in turn denoted as the set: $\{Hin1, Hin2, \dots, Hinm\}$. In the same way, the forecasting weather variables provided by NWP data include wind speed, trigonometric wind direction, temperature, humidity, and atmospheric pressure. $NWP_{Inputs}(t') = [V_{wind}(t'), (\sin(D_{wind}(t')) + \cos(D_{wind}(t'))), T(t'), H(t'), P(t')]$ is in turn denoted as the set: $\{windV, windD, Temp, Hum, airP\}$. Based on the obtained reconstruction parameters by the C_C method, the candidate input variables set of two wind farms can be obtained as follows:

$$\text{Anzishan} : \{Hin1, Hin2, Hin3, WindV, WindD, Temp, Hum, AirP\}$$

$$\text{Xuqiao} : \{Hin1, Hin2, Hin3, Hin4, Hin5, Hin6, WindV, WindD, Temp, Hum, AirP\}$$

Afterwards, the candidate variables are sorted through mRMR criterion to rank the predictive strength of each input variable. By observing the variation trend of the cumulative amount of $InSc$, that is, when the $CumInSc$ curve no longer increases or grows very slowly, the optimal number of input variables is selected. As the input variable selection is carried out adaptively in the prediction model, the selected input variables may vary for different prediction steps.

Based on proposed mRMR-criterion filter solution, Figure 5 shows the changes of $CumInSc$ curves of Anzishan and Xuqiao wind farms in different predicted steps (e.g., 1 h, 2 h, 3 h and 4 h ahead), respectively. The number of input variables is selected accordingly as shown in Figure 5. In this study, when the $CumInSc$ curve reaches the maxima, it is believed that adding the following variables at the back of extreme point will not add more useful information. Therefore, the candidate variables before the cumulative information maximum are regarded as the optimal or near optimal input variables to the prediction model. The detailed ranking of candidate variables and the number of input choices in multi-step ahead prediction for each farm are shown in Tables 1 and 2, respectively, where the selected input variables are highlighted in shade. It can be seen that the variables ranking in the step of proximity is similar and asymptotic. For different step-ahead prediction, the proposed hybrid solution with ranking the predictive strength candidate variables can select a compact subset of informative input variables based on the max-relevance and min-redundancy, which can effectively reduce the input dimension and interference information.

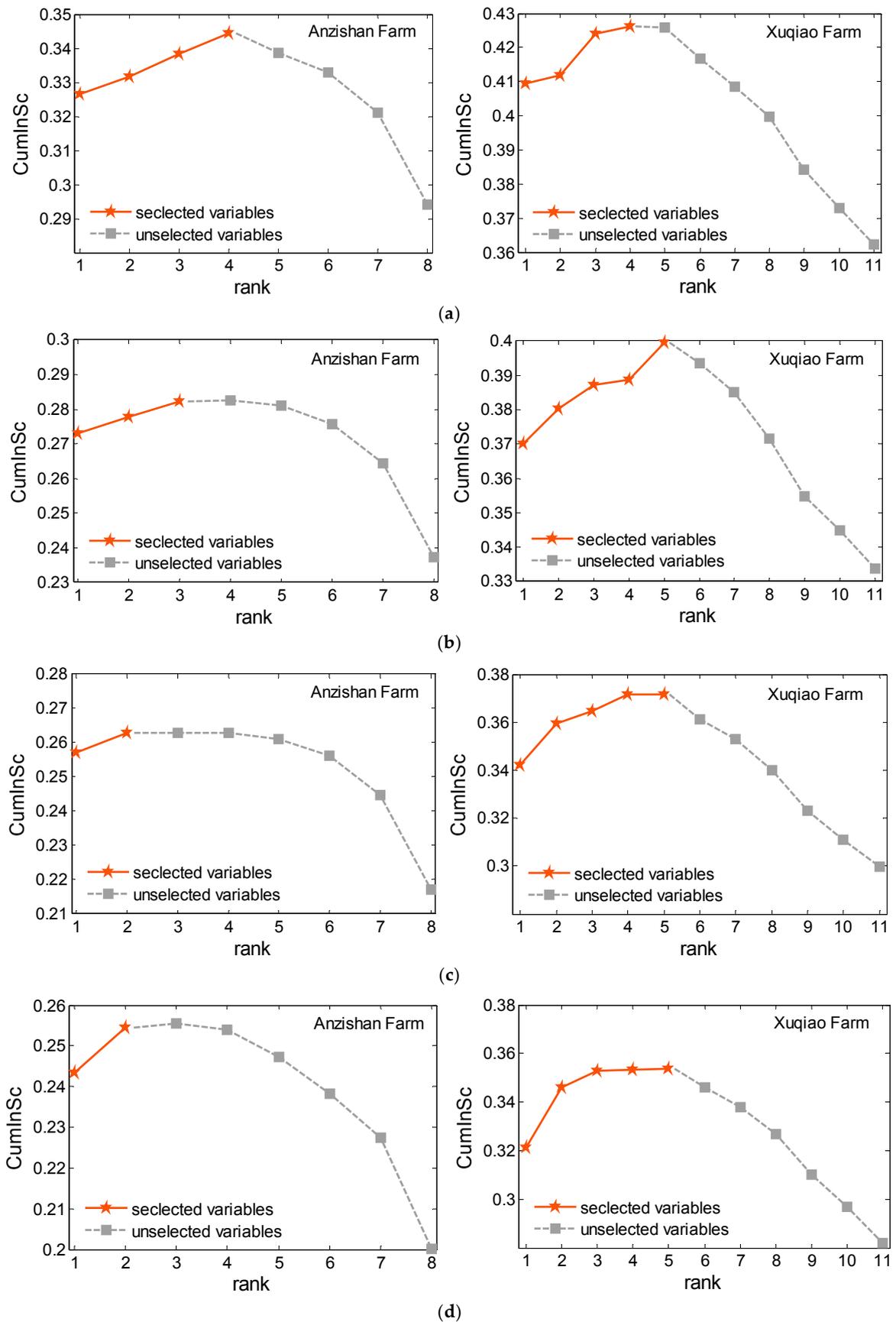


Figure 5. Cumulative informativeness score curve for IVS of Anzishan and Xuqiao farms: (a) one-hour ahead prediction; (b) two-hour ahead prediction; (c) three-hour ahead prediction; and (d) four-hour ahead prediction.

Table 1. Ranking and selection of candidate variables for Anzishan Farm.

Prediction Time Steps (15 min/step)	Input Variables Ranking	Number of Selected Variables
#1	<i>Hin1, WindD, Hin2, WindV; Hin3, Hum, AirP, Temp</i>	4
#2	<i>Hin1, Hin3, WindV, WindD, Hin2; Hum, AirP, Temp</i>	5
#3	<i>Hin1, WindD, WindV, Hin2; Hin3, Hum, AirP, Temp</i>	4
#4	<i>Hin1, WindD, WindV, Hin2; Hin3, Hum, AirP, Temp</i>	4
#5	<i>Hin1, WindD, WindV, Hin2; Hum, Hin3, AirP, Temp</i>	4
#6	<i>Hin1, WindV, Hin2, WindD; Hum, Hin3, AirP, Temp</i>	4
#7	<i>Hin1, WindV, Hin3; WindD, Hin2, Hum, AirP, Temp</i>	3
#8	<i>Hin1, WindV, Hin3; WindD, Hin2, Hum, AirP, Temp</i>	3
#9	<i>Hin1, WindV, Hin3; WindD, Hin2, Hum, AirP, Temp</i>	3
#10	<i>Hin1, WindV, Hin3; WindD, Hin2, Hum, AirP, Temp</i>	3
#11	<i>Hin1, WindV, Hin3; WindD, Hin2, Hum, AirP, Temp</i>	3
#12	<i>Hin1, WindV; Hin3, WindD, Hin2, Hum, AirP, Temp</i>	2
#13	<i>Hin1, WindV; Hin3, WindD, Hin2, Hum, AirP, Temp</i>	2
#14	<i>Hin1, WindV, Hin3; WindD, Hin2, AirP, Hum, Temp</i>	3
#15	<i>Hin1, WindV; Hin3, WindD, Hin2, AirP, Hum, Temp</i>	2
#16	<i>Hin1, WindV; WindD, Hin2, Hum, Hin3, AirP, Temp</i>	2

Table 2. Ranking and selection of candidate variables for Xuqiao Farm.

Prediction Time Steps (15 min/step)	Input Variables Ranking	Number of Selected Variables
#1	<i>Hin1, Hum, WindV, Hin2, Hin4; Hin3, Temp, Hin5, WindD, Hin6, AirP</i>	5
#2	<i>Hin1, Hum, WindV, Hin2, Hin4; Hin3, Temp, Hin5, WindD, Hin6, AirP</i>	5
#3	<i>Hin1, WindV, Hin2, Hin4; Hum, Hin3, Temp, Hin5, WindD, AirP, Hin6</i>	4
#4	<i>Hin1, WindV, Hin2, Hin4; Hum, Hin3, Temp, Hin5, WindD, Hin6, AirP</i>	4
#5	<i>Hin1, Hum, WindV, Hin2, Hin4; Hin3, Temp, Hin5, WindD, Hin6, AirP</i>	5
#6	<i>Hin1, WindV, Hin2, Hum, Hin4; Hin3, Temp, Hin5, WindD, Hin6, AirP</i>	5
#7	<i>Hin1, WindV, Hin2, Hin4, Hum; Hin6, Temp, Hin3, WindD, Hin5, AirP</i>	5
#8	<i>Hin1, WindV, Hin3, Hum, Hin2; Hin4, Temp, Hin6, WindD, Hin5, AirP</i>	5
#9	<i>Hin1, WindV, Hin3, Hum, Hin2; Hin4, Temp, Hin6, WindD, Hin5, AirP</i>	5
#10	<i>Hin1, WindV, Hin4, Hin2, Hum; Temp, Hin3, Hin6, WindD, Hin5, AirP</i>	4
#11	<i>Hin1, WindV, Hin3, Hum, Hin2; Hin4, Temp, Hin6, WindD, AirP, Hin5</i>	5
#12	<i>Hin1, WindV, Hin4, Hin2, Hum; Temp, Hin3, Hin6, WindD, Hin5, AirP</i>	5
#13	<i>Hin1, WindV, Hin4, Hin2, Hum; Temp, Hin3, Hin6, WindD, AirP, Hin5</i>	5
#14	<i>Hin1, WindV, Hin3, Hum, Hin2; Hin4, Temp, Hin6, WindD, AirP, Hin5</i>	5
#15	<i>Hin1, WindV, Hin3, Hin2, Hum; Temp, Hin4, Hin6, WindD, AirP, Hin5</i>	5
#16	<i>Hin1, WindV, Hin3, Hum, Hin2; Hin4, Temp, Hin6, WindD, AirP, Hin5</i>	5

4.2. Case Study and Numerical Result

For different prediction steps, the corresponding selected samples are used to train the aforementioned hybrid prediction solution in Section 3. Here, the main parameters and settings for training the optimal ANFIS are summarized in Table 3. After determining the ANFIS parameters by training samples, the multi-step ahead prediction results in the test month can be obtained. Figure 6 presents the prediction result of 150 h from the test month for two wind farms, respectively.

Table 3. Main setting parameters for training.

Main Techniques	Parameters and Settings
Hyperparameters of PSO	Inertia weight $w_0 = 1$
	Inertia weight damping ratio $w_{damp} = 0.99$
	Individual learning coefficient $c_1 = 1$
	Global learning coefficient $c_2 = 2$
	Numbers of particles $n_{pop} = 30$
	Maximum number of iterations $n_{ite} = 100$
Hyperparameters of ANFIS	Number of fuzzy set in each input variable $n_{fuzzy} = 5$
Hyperparameters of K-means	Number of clustering centers $k = 6$
Number of samples in the two farms	Number of training samples $n_{train}^{Az} = 8832, n_{train}^{Xq} = 8736$
	Number of test samples $n_{test}^{Az} = 2880, n_{test}^{Xq} = 2796$

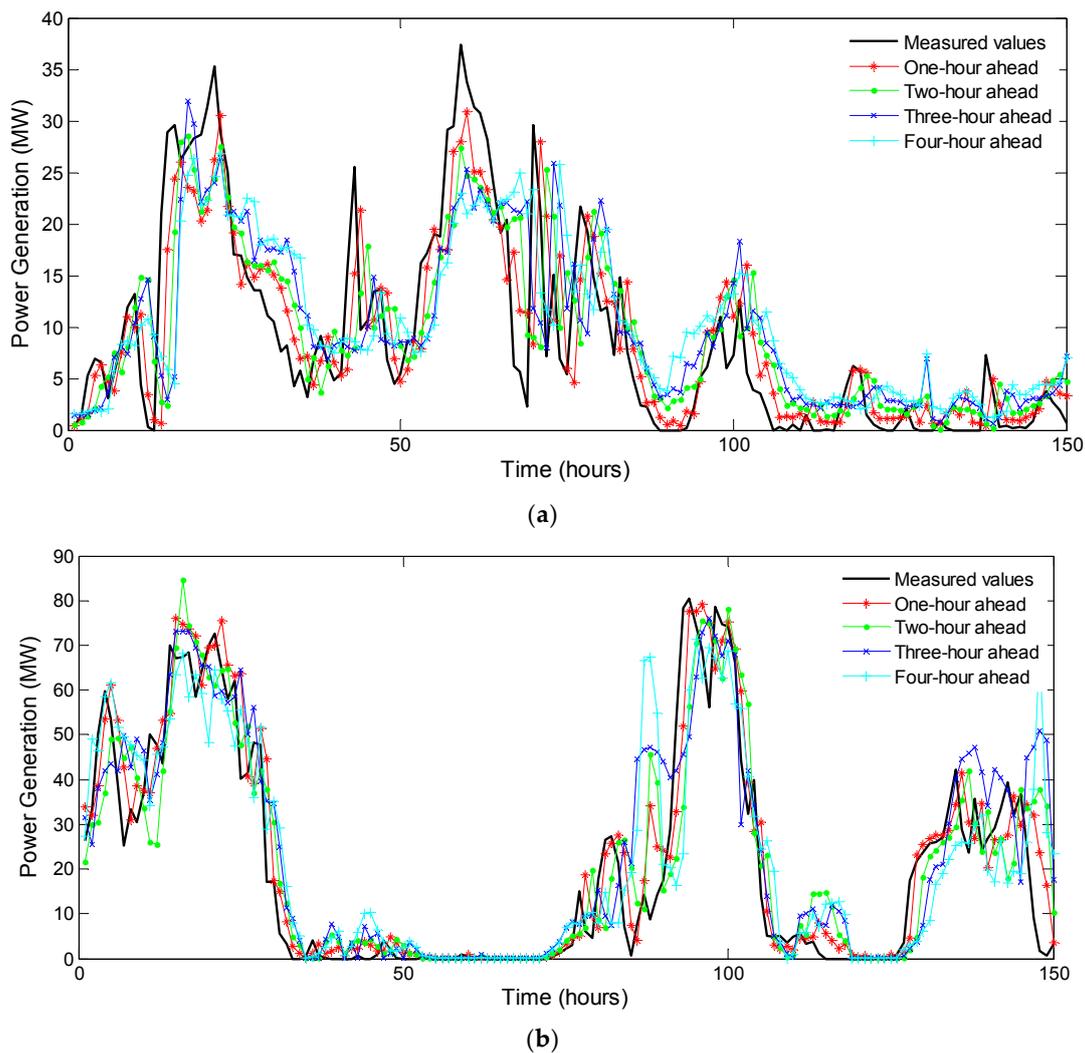


Figure 6. Partial results of multi-step ahead prediction in the test month: (a) power generation prediction of 150 h in Anzishan wind farm (June 2017); and (b) power generation prediction of 150 h in Xuqiao wind farm (December 2017).

To evaluate the effectiveness and accuracy of the prediction solution, the performance metric in terms of normalized root mean squared error (nRMSE) and normalized mean absolute error (nMAE) are adopted [40], as given in Equations (27) and (28), respectively. In general, smaller values of these measures indicate that the corresponding solution has better prediction performance.

$$\text{nRMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{P_{mi} - P_{pi}}{C_i} \right)^2} \quad (27)$$

$$\text{nMAE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{P_{mi} - P_{pi}}{C_i} \right| \quad (28)$$

where P_{pi} is the predicted power of time point i , P_{mi} is the measured mean power of time point i , N is the number of prediction samples, and C_i is the operating capacity of time point i .

To evaluate the proposed mRMR-criterion input variable selection (IVS) solution based prediction approach, mRMR-IVS model, a detailed comparative study was conducted for multi-step ahead wind power generation prediction. Two IVS based prediction approaches, the phase space reconstruction

based (PSR-IVS) model and principal component analysis based (PCA-IVS) model, were selected as the comparative benchmark. For PSR-IVS based prediction model, the input variables included the NWP variables and the phase space reconstruction variables of the historical time series. The reconstruction was determined based on C_C method as well. In this model, the input variables were the ones which are candidates in proposed mRMR-IVS solution, without further exquisite selection. For PCA-IVS based model, the input variables were transformed from the combination of NWP variables and a time series of 2 h, which use the principal component analysis (PCA) technique [41] to map the dataset from the original space to the principal component space. In this model, the original attribute variables are automatically reduced to appropriate input variables and the independent principal components can well maintain the key characteristics of the original variables. After selecting or extracting the input variables, all the IVS-based approaches used the hybrid model introduced in Section 3 to carry out the prediction.

Two error criteria, nRMSE and nMAE, were used to assess the performance of all considered prediction models. Tables 4 and 5 show the comparison of multi-step ahead prediction performance of different models for two wind farms, respectively. As shown in Tables 4 and 5, the proposed model demonstrates the smallest error over all 16 steps of prediction compared with both benchmark models. In addition, compared with the principal component analysis-based model (PCA-IVS), the phase space reconstruction based (PSR-IVS) model performs better in short step prediction, but has larger error in long prediction period. Due to the sophisticated and targeted input variable selection, the proposed model has a better performance than the benchmark models in the overall multi-step ahead prediction.

To present the comparison more intuitively, Figures 7 and 8 show the broken line in two wind farms of three cases based on the values of nRMSE and nMAE of different prediction steps. The curve trend shows that the error level rises with the increase of prediction step, meeting the objective expectation. It is shown that there are fluctuations in the curve of two benchmark models, especially in the intermediate prediction period, in that the selection of input variables cannot adapt to each prediction step. In the proposed model, the error trend increased smoothly, indicating that the mRMR-IVS based model can automatically select the optimal or nearly optimal input variables for different prediction steps to reduce the error. It means that the proposed hybrid solution can select suitable input variables effectively in different geographical environments and seasons, showing better adaptability and robustness.

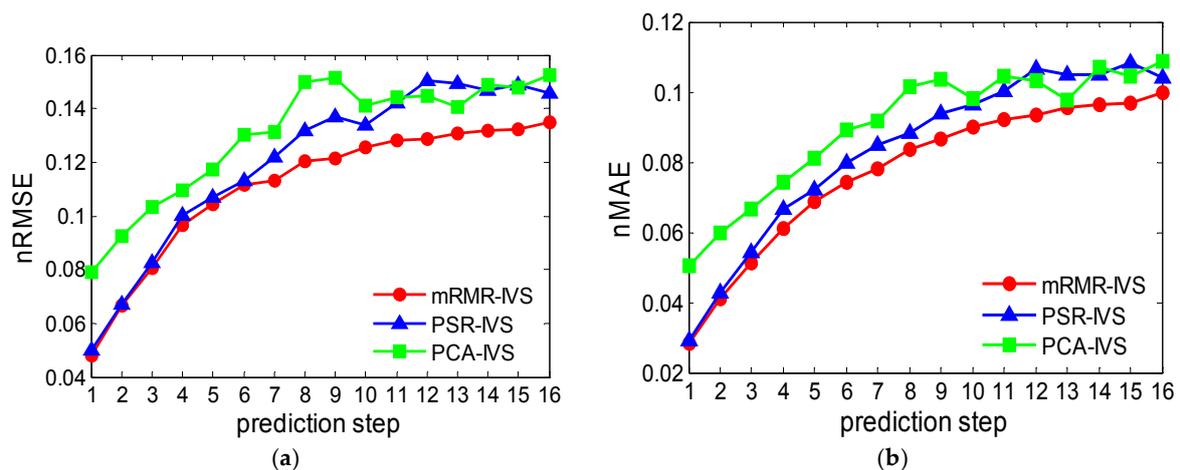


Figure 7. Error trend in different prediction steps of Anzishan farm: (a) nRMSE of test month; and (b) nMAE of test month.

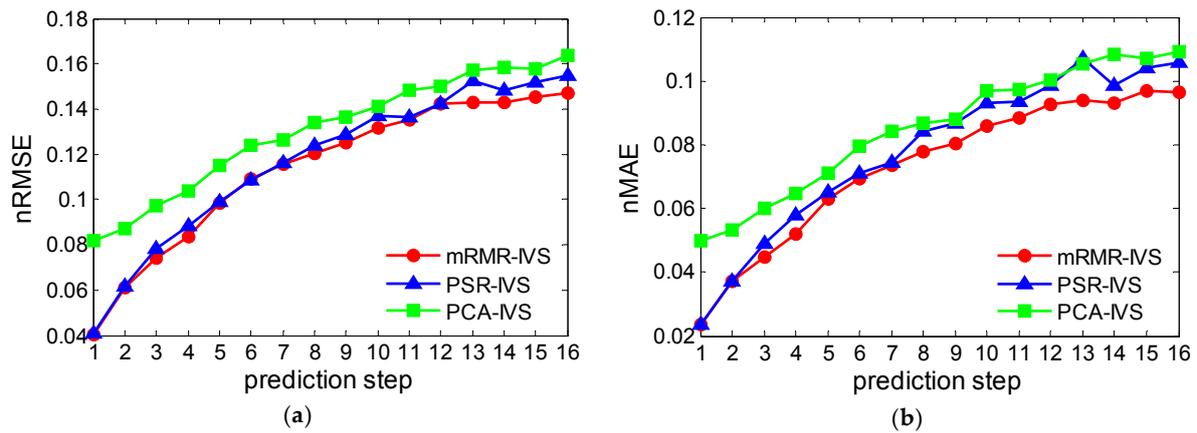


Figure 8. Error trend in different prediction steps of Xuqiao farm: (a) nRMSE of test month; and (b) nMAE of test month.

Table 4. Comparison of the multi-step ahead prediction performance of Anzishan farm.

Prediction Time Steps	mRMR-IVS		PSR-IVS		PCA-IVS	
	nRMSE (%)	nMAE (%)	nRMSE (%)	nMAE (%)	nRMSE (%)	nMAE (%)
#1	4.78	2.84	5.01	2.90	7.91	5.04
#2	6.67	4.10	6.74	4.27	9.25	5.98
#3	8.04	5.13	8.28	5.43	10.36	6.66
#4	9.64	6.12	10.01	6.64	10.98	7.43
#5	10.44	6.87	10.69	7.21	11.74	8.12
#6	11.18	7.43	11.31	7.96	13.03	8.93
#7	11.32	7.82	12.21	8.48	13.12	9.19
#8	12.02	8.38	13.19	8.81	14.97	10.15
#9	12.17	8.68	13.71	9.37	15.17	10.35
#10	12.54	8.99	13.37	9.64	14.11	9.80
#11	12.80	9.21	14.22	10.04	14.41	10.46
#12	12.88	9.36	15.02	10.66	14.50	10.30
#13	13.06	9.55	14.92	10.48	14.05	9.77
#14	13.17	9.65	14.69	10.49	14.89	10.69
#15	13.24	9.70	14.87	10.82	14.78	10.43
#16	13.51	9.97	14.60	10.39	15.27	10.88

Table 5. Comparison of the multi-step ahead prediction performance of Xuqiao farm.

Prediction Time Steps	mRMR-IVS		PSR-IVS		PCA-IVS	
	nRMSE (%)	nMAE (%)	nRMSE (%)	nMAE (%)	nRMSE (%)	nMAE (%)
#1	4.04	2.34	4.06	2.35	8.19	4.98
#2	6.12	3.71	6.17	3.70	8.69	5.33
#3	7.39	4.47	7.81	4.91	9.72	6.02
#4	8.37	5.22	8.85	5.80	10.40	6.45
#5	9.86	6.28	9.89	6.50	11.53	7.10
#6	10.89	6.93	10.87	7.12	12.37	7.95
#7	11.55	7.34	11.62	7.45	12.65	8.41
#8	12.05	7.78	12.38	8.41	13.40	8.69
#9	12.50	8.02	12.86	8.68	13.66	8.81
#10	13.19	8.57	13.71	9.29	14.11	9.69
#11	13.55	8.85	13.62	9.34	14.82	9.72
#12	14.23	9.28	14.24	9.87	15.02	10.02
#13	14.27	9.39	15.23	10.69	15.69	10.53
#14	14.29	9.33	14.81	9.87	15.82	10.84
#15	14.54	9.68	15.21	10.41	15.79	10.72
#16	14.71	9.65	15.50	10.56	16.35	10.94

To further validate the integral time-scales of ultra-short-term wind power prediction, which is generally 0–4 h in the future with a resolution of 15 min, 100 integrated time series with consecutive

4-h prediction were randomly selected during the test-month forecasting process to calculate the performance metric in statistics. The nRMSE and nMAE indicators of each integrated time series were calculated. Then, the probability of different error levels could be obtained according to the statistics of appearing frequency in these 100 results. The average errors and the probability distribution of typical error levels of Anzishan and Xuqiao wind farms are shown in Tables 6 and 7, respectively. For both cases, the proposed mRMR-IVS based solution demonstrates the minimum mean errors. The results given in Tables 6 and 7 verify that the proposed model is more competitive for most probability distributions at different error levels. This clearly confirms that the proposed solution can provide improved prediction performance than the comparison benchmarks.

Table 6. Mean errors and probability distributions of Anzishan farm.

Models	Mean nRMSE	Error Level for $P\{nRMSE < 0.2\}$	Error Level for $P\{nRMSE < 0.15\}$	Error Level for $P\{nRMSE < 0.05\}$
mRMR-IVS	9.95%	93.80%	82.40%	16.8%
PSR-IVS	10.75%	89.80%	82.40%	10.90%
PCA-IVS	11.79%	88.20%	76.10%	9.80%
Models	Mean nMAE	Error Level for $P\{nMAE < 0.2\}$	Error Level for $P\{nMAE < 0.15\}$	Error Level for $P\{nMAE < 0.05\}$
mRMR-IVS	7.84%	98.20%	92.80%	32.10%
PSR-IVS	8.88%	94.00%	87.40%	21.20%
PCA-IVS	9.65%	93.30%	84.90%	18.10%

Table 7. Mean errors and probability distributions of Xuqiao farm.

Models	Mean nRMSE	Error Level for $P\{nRMSE < 0.2\}$	Error Level for $P\{nRMSE < 0.15\}$	Error Level for $P\{nRMSE < 0.05\}$
mRMR-IVS	9.71%	91.60%	80.80%	27.00%
PSR-IVS	10.41%	90.00%	75.30%	26.60%
PCA-IVS	11.47%	84.00%	69.60%	25.30%
Models	Mean nMAE	Error Level for $P\{nMAE < 0.2\}$	Error Level for $P\{nMAE < 0.15\}$	Error Level for $P\{nMAE < 0.05\}$
mRMR-IVS	7.62%	96.80%	89.10%	38.40%
PSR-IVS	8.70%	94.00%	84.00%	32.30%
PCA-IVS	9.73%	89.80%	77.40%	31.50%

5. Conclusions and Future Work

This paper develops a novel algorithmic solution for ultra-short-term wind power generation prediction using hybrid machine learning techniques. The proposed solution is implemented through two steps: firstly, the input variable selection (IVS) is carried out using phase space reconstruction (PSR) technique and minimal redundancy maximal relevance (mRMR) criterion. Secondly, the input instances are divided into a set of subsets using the K-means clustering to train the ANFIS with parameters optimized using PSO. The proposed solution was extensively evaluated and validated through case studies based on real wind farms. The numerical results demonstrate the superiority of the proposed model compared with the benchmark models.

In the future, two research directions are considered worth further research effort. The proposed prediction solution can be further exploited using other supervised learning algorithms and advanced error correction techniques as well as extensively validated based on more field measurements. The hybrid machine learning techniques can be further extended and incorporated into control and management strategies of renewable energy systems (e.g., [42–45]) to improve the system operational performance.

Author Contributions: W.D. and Q.Y. conceived and designed the experiments; W.D. performed the experiments and analyzed the data; W.D., X.F. and Q.Y. contributed reagents/materials/analysis tools; and W.D. and Q.Y. wrote the paper.

Funding: This research was funded by the Natural Science Foundation of China (51777183), the Natural Science Foundation of Zhejiang Province (LZ15E070001) and the Natural Science Foundation of Jiangsu Province (BK20161142).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations and Nomenclature

PSR	Phase Space Reconstruction
IVS	Input Variable Selection
ANFIS	Adaptive Neuro-fuzzy Inference System
NWP	Numerical Weather Prediction
mRMR	Minimal Redundancy Maximal Relevance
PSO	Particle Swarm Optimization Algorithm
AR	Autoregressive Model
MA	Moving-average Model
ARIMA	Auto-regressive Integrated Moving Average
ANN	Artificial Neural Network
SVM	Support Vector Machine
GP	Gaussian Process
SVR	Support Vector Regression
EHS	Enhanced Harmony Search Algorithm
SSA	Singular Spectrum Analysis
MOS	Model Output Statistics
SVD	Singular Value Decomposition
PCA	Principal Component Analysis
LLE	Locally Linear Embedding
MI	Mutual Information
MIFS	Mutual Information Feature Selection
MIFS-U	Mutual Information Feature Selection under Uniform Information Distribution
NMIFS	Normalized Mutual Information Feature Selection
IFS	Incremental Forward Selection
FCM	Fuzzy C-Means Clustering Algorithm
nRMSE	Normalized Root Mean Squared Error
nMAE	Normalized Mean Absolute Error
mRMR-IVS	mRMR-criterion Input Variable Selection
PSR-IVS	Phase Space Reconstruction Input Variable Selection
PCA-IVS	Principal Component Analysis Input Variable Selection

References

1. Ponta, L.; Raberto, M.; Teglio, A.; Cincotti, S. An agent-based stock-flow consistent model of the sustainable transition in the energy sector. *Ecol. Econ.* **2018**, *145*, 274–300. [[CrossRef](#)]
2. Filippo, A.D.; Lombardi, M.; Milano, M. User-aware electricity price optimization for the competitive market. *Energies* **2017**, *10*, 1378. [[CrossRef](#)]
3. Huang, C.M.; Kuo, C.J.; Huang, Y.C. Short-term wind power forecasting and uncertainty analysis using a hybrid intelligent method. *IET Renew. Power Gen.* **2017**, *11*, 678–687. [[CrossRef](#)]
4. Weron, R. Electricity price forecasting: a review of the state-of-the-art with a look into the future. *Int. J.* **2014**, *30*, 1030–1081. [[CrossRef](#)]
5. Cincotti, S.; Gallo, G.; Ponta, L.; Raberto, M. Modeling and forecasting of electricity spot-prices: computational intelligence vs classical econometrics. *Ai Commun.* **2014**, *27*, 301–314.

6. Giebel, G.; Brownsword, R.; Kariniotakis, G. *The State-Of-The-Art in Short-Term Prediction of Wind Power: A Literature Overview*; Technical Report for ANEMOS.plus: Roskilde, Denmark, 2011.
7. Costa, A.; Crespo, A.; Navarro, J.; Lizcano, G.; Madsen, H.; Feitosa, E. A review on the young history of the wind power short-term prediction. *Renew. Sustain. Energy Rev.* **2008**, *12*, 1725–1744. [[CrossRef](#)]
8. Li, G.; Shi, J. On comparing three artificial neural networks for wind speed forecasting. *Appl. Energy* **2010**, *87*, 2313–2320. [[CrossRef](#)]
9. Zhou, J.; Shi, J.; Li, G. Fine tuning support vector machines for short-term wind speed forecasting. *Energy Convers. Manag.* **2011**, *52*, 1990–1998. [[CrossRef](#)]
10. Taslimi Renani, E.; Elias, M.F.M.; Rahim, N.A. Using data-driven approach for wind power prediction: A comparative study. *Energy Convers. Manag.* **2016**, *118*, 193–203. [[CrossRef](#)]
11. Lee, D.; Baldick, R. Short-term wind power ensemble prediction based on gaussian processes and neural networks. *IEEE Trans. Smart Grid* **2014**, *5*, 501–510. [[CrossRef](#)]
12. Guo, Z.; Chi, D.; Wu, J.; Zhang, W. A new wind speed forecasting strategy based on the chaotic time series modelling technique and the apriori algorithm. *Energy Convers. Manag.* **2014**, *84*, 140–151. [[CrossRef](#)]
13. Xue, Y.; Yu, C.; Li, K.; Wen, F.; Ding, Y.; Wu, Q. Adaptive ultra-short-term wind power prediction based on risk assessment. *Csee J. Power Energy Syst.* **2016**, *2*, 59–64. [[CrossRef](#)]
14. Safari, N.; Chung, C.Y.; Price, G.C.D. A novel multi-step short-term wind power prediction framework based on chaotic time series analysis and singular spectrum analysis. *IEEE Trans. Power Syst.* **2017**, *99*, 590–601. [[CrossRef](#)]
15. Santamaria-Bonfil, G.; Reyes-Ballesteros, A.; Gershenson, C. Wind speed forecasting for wind farms: A method based on support vector regression. *Renew. Energy* **2016**, *85*, 790–809. [[CrossRef](#)]
16. Buhan, S.; Cadirci, I. Multistage wind-electric power forecast by using a combination of advanced statistical methods. *IEEE Trans. Ind. Inform.* **2015**, *11*, 1231–1242. [[CrossRef](#)]
17. Fang, S.; Chiang, H.D. A high-accuracy wind power forecasting model. *IEEE Trans. Power Syst.* **2017**, *32*, 1589–1590. [[CrossRef](#)]
18. Fang, S.; Chiang, H.D. Improving supervised wind power forecasting models using extended numerical weather variables and unlabelled data. *IET Renew. Power Gen.* **2016**, *10*, 1616–1624. [[CrossRef](#)]
19. Fu, T. A review on time series data mining. *Appl. Artif. Intell.* **2011**, *24*, 164–181. [[CrossRef](#)]
20. Zhao, X.; Deng, W.; Shi, Y. Feature selection with attributes clustering by maximal information coefficient. *Procedia Comput. Sci.* **2013**, *17*, 70–79. [[CrossRef](#)]
21. Zhao, H.; Magoulès, F. Feature selection for predicting building energy consumption based on statistical learning method. *J. Algorithms Comput. Technol.* **2012**, *6*, 59–78. [[CrossRef](#)]
22. Kapetanakis, D.-S.; Mangina, E.; Finn, D.P. Input variable selection for thermal load predictive models of commercial buildings. *Energy Build.* **2017**, *137*, 13–26. [[CrossRef](#)]
23. Sharma, A.; Paliwal, K.K.; Imoto, S.; Miyano, S. A feature selection method using improved regularized linear discriminant analysis. *Mach. Vis. Appl.* **2013**, *25*, 775–786. [[CrossRef](#)]
24. Zjavka, L.; Misak, S. Direct wind power forecasting using a polynomial decomposition of the general differential equation. *IEEE Trans. Sustain. Energy* **2018**, *99*. [[CrossRef](#)]
25. Takens, F. Detecting strange attractors in turbulence. Lecture Notes in Mathematics-Springer-verlag-(Lect Notes Math). In *Lecture Notes Math*; Rand, D.A., Young, L.-S., Eds.; Springer: Berlin, Germany, 1981; Volume 898, pp. 366–381.
26. Kim, H.S.; Eykholt, R.; Salas, J.D. Nonlinear dynamics, delay times, and embedding windows. *Phys. D Nonlinear Phenom.* **1999**, *127*, 48–60. [[CrossRef](#)]
27. Niu, T.; Wang, J.; Lu, H.; Du, P. Uncertainty modeling for chaotic time series based on optimal multi-input multi-output architecture: Application to offshore wind speed. *Energy Convers. Manag.* **2018**, *156*, 597–617. [[CrossRef](#)]
28. Maji, P.; Garai, P. On fuzzy-rough attribute selection: criteria of max-dependency, max-relevance, min-redundancy, and max-significance. *Appl. Soft Comput.* **2013**, *13*, 3968–3980. [[CrossRef](#)]
29. Wang, X.; Han, M.; Wang, J. Applying input variables selection technique on input weighted support vector machine modeling for bof endpoint prediction. *Appl. Artif. Intell.* **2010**, *23*, 1012–1018. [[CrossRef](#)]
30. Battiti, R. Using mutual information for selecting features in supervised neural net learning. *IEEE Trans. Neural Netw.* **1994**, *5*, 537–550. [[CrossRef](#)] [[PubMed](#)]

31. Kwak, N.; Choi, C.-H. Input feature selection by mutual information based on Parzen window. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 1667–1671. [[CrossRef](#)]
32. Estevez, P.A.; Tesmer, M.; Perez, C.A.; Zurada, J.M. Normalized mutual information feature selection. *IEEE Trans. Neural Netw.* **2009**, *20*, 189–201. [[CrossRef](#)] [[PubMed](#)]
33. Cover, T.; Thomas, J. *Elements of Information Theory*; Wiley-Blackwell: Hoboken, NJ, USA, 2006.
34. Ponta, L.; Carbone, A. Information measure for financial time series: quantifying short-term market heterogeneity. *Phys. A Stat. Mech. Appl.* **2018**, *510*, 123–144. [[CrossRef](#)]
35. Bandt, C.; Pompe, B. Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.* **2002**, *88*. [[CrossRef](#)] [[PubMed](#)]
36. Celebi, M.E.; Kingravi, H.A.; Vela, P.A. A comparative study of efficient initialization methods for the k-means clustering algorithm. *Expert Syst. Appl.* **2013**, *40*, 200–210. [[CrossRef](#)]
37. Jang, J.-S.R. Anfis: Adaptive-network-based fuzzy inference system. *IEEE Trans. Syst. Man. Cybern.* **1993**, *23*, 665–685. [[CrossRef](#)]
38. Ghosh, S.; Kumar, S. Comparative analysis of k-means and fuzzy c-means algorithms. *Int. J. Adv. Comput. Sci. Appl.* **2013**, *4*. [[CrossRef](#)]
39. Kennedy, J.; Eberhart, R.C. Particle swarm optimization. In Proceedings of the IEEE International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; pp. 1942–1948.
40. Madsen, H.; Pinson, P.; Kariniotakis, G.; Nielsen, H.A.; Nielsen, T.S. Standardizing the performance evaluation of short term wind power prediction models. *Wind Eng.* **2005**, *29*, 475–489. [[CrossRef](#)]
41. Wu, Q.; Peng, C. Wind power generation forecasting using least squares support vector machine combined with ensemble empirical mode decomposition, principal component analysis and a bat algorithm. *Energies* **2016**, *9*, 261. [[CrossRef](#)]
42. Baghaee, H.R.; Mirsalim, M.; Gharehpetian, G.B.; Talebi, H.A. Three-phase ac/dc power-flow for balanced/unbalanced microgrids including wind/solar, droop-controlled and electronically-coupled distributed energy resources using radial basis function neural networks. *IET Power Electron.* **2017**, *10*, 313–328. [[CrossRef](#)]
43. Baghaee, H.R.; Mirsalim, M.; Gharehpetian, G.B. Power Calculation Using RBF Neural Networks to Improve Power Sharing of Hierarchical Control Scheme in Multi-DER Microgrids. *IEEE J. Emerg. Sel. Top. Power Electron.* **2016**, *4*, 1217–1225. [[CrossRef](#)]
44. Baghaee, H.R.; Mirsalim, M.; Gharehpetan, G.B.; Talebi, H.A. Nonlinear load sharing and voltage compensation of microgrids based on harmonic power-flow calculations using radial basis function neural networks. *IEEE Syst. J.* **2016**, *PP*, 1–11. [[CrossRef](#)]
45. Morshed, M.J.; Hmida, J.B.; Fekih, A. A probabilistic multi-objective approach for power flow optimization in hybrid wind-pv-pev systems. *Appl. Energy* **2018**, *211*, 1136–1149. [[CrossRef](#)]

