

Article

A Hybrid Wind Speed Forecasting System Based on a ‘Decomposition and Ensemble’ Strategy and Fuzzy Time Series

Hufang Yang ¹, Zaiping Jiang ^{1,*} and Haiyan Lu ²

¹ School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China; dufesos_1530@126.com

² Faculty of Engineering and Information Technology, University of Technology, Sydney 20000, Australia; haiyan.lu@uts.edu.au

* Correspondence: jiangzaiping@dufe.edu.cn; Tel.: +86-159-4081-7715

Academic Editor: Francisco Martínez-Álvarez

Received: 25 August 2017; Accepted: 13 September 2017; Published: 16 September 2017

Abstract: Accurate and stable wind speed forecasting is of critical importance in the wind power industry and has measurable influence on power-system management and the stability of market economics. However, most traditional wind speed forecasting models require a large amount of historical data and face restrictions due to assumptions, such as normality postulates. Additionally, any data volatility leads to increased forecasting instability. Therefore, in this paper, a hybrid forecasting system, which combines the ‘decomposition and ensemble’ strategy and fuzzy time series forecasting algorithm, is proposed that comprises two modules—data pre-processing and forecasting. Moreover, the statistical model, artificial neural network, and Support Vector Regression model are employed to compare with the proposed hybrid system, which is proven to be very effective in forecasting wind speed data affected by noise and instability. The results of these comparisons demonstrate that the hybrid forecasting system can improve the forecasting accuracy and stability significantly, and supervised discretization methods outperform the unsupervised methods for fuzzy time series in most cases.

Keywords: wind speed forecasting; hybrid forecasting system; data pre-processing; fuzzy time series; comprehensive evaluation

1. Introduction

Energy is a vital input for social and economic development [1]. The energy crisis has been proven to be one of the major factors that limit the development of the economy, and this has been increasingly emphasized by the increasing energy demands for rapid economic development [2]. With the continuous increase in energy demand, the consumption of non-renewable energy sources, such as coal and oil, has become alarmingly serious, resulting in an ever-growing energy crisis. This is due to the fact that fossil fuels, such as coal and oil, are slowly drying up, and non-renewable energy will become history in the near future [3]. In view of this present situation, people have gradually turned their attention to the development and utilization of new energy sources and have tried to change the trend in energy consumption to relieve, to some extent, the double pressure caused by the dry up of conventional energy and worsening of the global ecological environment [4].

Wind energy, one of the most important renewable energy resources, is drawing increasing attention by virtue of its prominent characteristics, such as wide distribution and prodigious reserves [5]. The development of wind energy, as an efficient and clean energy resource, is well known and establishes a good base for the strategic transformation of economic development from

relying on traditional fossil fuels to utilization of renewable energy sources [6]. Wind energy utilization has been around for than a century, and wind power generation has also been substantially explored by humans in the past. Wind power generation technology has been developed through a long process has become increasingly mature [7]. Moreover, there is a huge amount of wind energy in the world [8]. By the end of 2016, the worldwide wind capacity reached 486,661 MW, of which, 54,846 MW of energy were added in 2016. This represents a growth rate of 11.8% (17.2% in 2015). All wind turbines installed around the globe by the end of 2016 can generate around 5% of the world's total electricity demand [9].

As we all recognize, China has a large population, and its economy has been predicted to maintain good momentum of development. Thus, the above problems become more prominent due to the amazing energy consumption and the growth speed of traditional fossil fuel exploitation. In the near future, the supply of fossil fuel will not keep up with the demand which may hold back economic development. At the same time, the pressure of environmental degradation is also a problem that people have to face. Therefore, it is urgent to rationally adjust the energy structure for the sustainable development of the economy. In view of these reasons, the research about new energy, especially the wind power industry becomes more necessary. The wind power industry in China, through the government's great attention, is playing a positive role in optimizing the energy structure, promoting changes in energy production methods, and promoting transformation in the energy consumption of modern industrial systems [10].

Moreover, in wind data, it is necessary to consider and discuss the frequency of data sampling. According to State Grid Dispatching arrangement and plan in China, 144 wind speed datapoints should be obtained per day (24 h). In other words, the sampling interval is supposed to be 10 min. Ten minute wind speed forecasting has contributed to scientific and rational arrangements for the shut-down and start-up of the generators in the net so that the system can maintain a rotational reserve capacity within a reasonable and safe range [11]. Moreover, the minimum time interval recorded by the anemometer is 10 min at present. Thus, the sampling interval is set to 10 min and sampling frequency is 144 times per day in most researches [12] to meet the requirement of power grid scheduling in China.

While the potential of wind power as an energy resource is fully ascertained, its controllability needs to be improved. This controllability of wind power can be improved if the wind speed and the power output of a wind farm can be forecasted as accurately as possible and changes in wind speed can be predicted well in advance [13]. This would also help mitigate a series of adverse effects that result from wind power grid integration. Wind speed is influenced by several factors, such as air pressure, temperature, and humidity, which lead to randomness and volatility in wind speed prediction [14]. Wind speed forecasting has been an important link in the planning and working of power grid system; this is a heavy and high repetitive work. Moreover, wind speed forecasting is the basis of wind power and an important prerequisite for wind-power generation capacity forecasting. Thus, wind speed forecasting is a significant task and establishing a high accuracy of the wind speed forecasting model becomes a pressing concern [15].

The rest of this paper is organized as follows: Section 2 reviews and discusses the extant studies on wind speed forecasting. The methods used in this study are introduced in Section 3. Section 4 describes the datasets and setup. Section 5 describes the experimental results obtained from the datasets, while Section 6 analyses and discusses the forecasting results. Section 7 discusses parameters of the hybrid forecasting system. Section 8 further carries out the experiment for hourly time-horizon wind speed forecasting, and Section 9 gives the conclusion. Figure 1 clearly explains this structure.

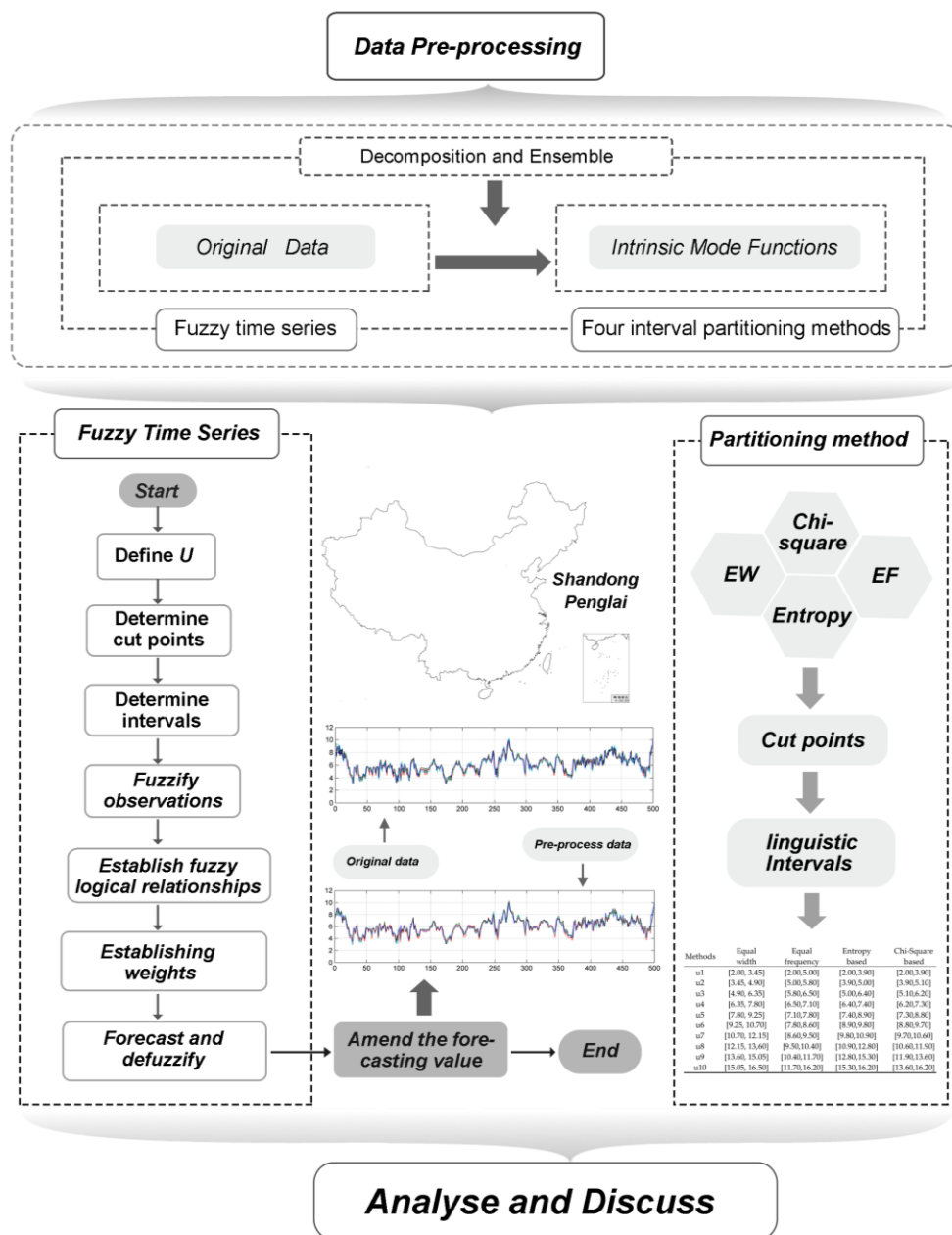


Figure 1. The flow chart of the proposed hybrid forecasting system.

2. Review and Discussion for Previous Works

Based on the discussion presented in Section 1 above, it can be appreciated that wind speed forecasting is a challenging yet crucial task. The accuracy and stability of such a forecasting is, perhaps, the single most significant issue, and as such, numerous extant researches have been targeted at addressing this concern.

Two prominent models used at present for wind speed forecasting include the single model [16–18] and hybrid model. The single model mainly comprises of a physical model, statistical model and an artificial neural network model. The physical model essentially utilizes a dynamic atmosphere model to simulate and forecast the wind speed. In the real-world scenario, hydrodynamic and thermodynamic equations that model changes in the weather pattern are used along with specified initial and boundary conditions to model the exact situation to be simulated by a megacomputer [19].

Time series is a set of values wherein all values of one index are arranged in chronological order. The main utility of the time series model is to forecast the future based on historical data. The traditional statistical models, such as Autoregressive (AR) [20,21], Autoregressive Moving Average (ARMA) [22], Autoregressive Integrated Moving Average (ARIMA) [23], and exponential smoothing (ES) [24], have widely used and reported in literature for their utility in wind speed forecasting, which was originally developed by Kendall and Ord [25].

Artificial neural network models have attracted extensive attention of scholars in various fields as they are capable of modeling linear as well as nonlinear functions arbitrarily. The use of artificial neural networks is a popular method for wind speed forecasting. Li et al. [26] compared three different neural networks for wind forecasting, including the adaptive linear element, back propagation, and radial basis function, and demonstrated that no single model is superior to others for all evaluation metrics. Hervás-Martínez et al. [27] proposed the hyperbolic tangent basis function neural network for wind forecasting, and the results demonstrate that their model improved the performance of the previous multilayer perceptron. Salcedo-Sanz et al. [28] forecasted the short-term wind speed by applying the Coral Reefs Optimization (CRO) algorithm and an Extreme Learning Machine (ELM). A Feature Selection Problem (FSP) was carried out to prove that the CRO-ELM approach had an excellent performance in wind speed forecasting. A further study showed that better results could be obtained by using ELM in conjunction with a CRO-Harmony Search (HS) optimization algorithm [29]. In addition to these above-stated models, other popular models employed in wind forecasting include support vector regression [30–33], Bayesian mode [34], and regression trees [35].

As mentioned above, no single model can obtain optimum results under all situations and perform better than others on all fronts. Therefore, some hybrid models have been proposed to remedy some of the weaknesses [36–39]. A hybridization of the fifth generation mesoscale model with neural networks was employed to address the short-term wind speed forecasting issue [40]. Similarly, the hybridization of global and mesoscale weather forecasting models with neural networks was also employed for short-term wind speed forecasting. The results prove that the hybrid weather forecast model's neural network approach can achieve great forecasting results for short-term wind speeds under specific situations [41]. Hervas-Martinez et al. proposed a hybrid model that combines the physical, statistical, and artificial neural networks, and achieves great forecasting accuracy [42]. Zhang et al. [43] developed a novel wavelet transform technique (WTT)-seasonal adjustment method (SAM)-radial basis function neural network (RBFNN) for short-term wind speed forecasting, which was proved to be an effective approach to improve the forecasting performance. Compared to the single model, the hybrid model was found to effectively improve the forecasting accuracy.

In addition to the choice of the forecasting model, de-noising of raw data also makes a significant contribution to the prediction accuracy. Wind signal de-noising methods, such as empirical mode decomposition [44,45], secondary decomposition [46], and fast ensemble empirical mode decomposition [47] algorithms, can effectively reduce noise in the wind speed time series signal and greatly improve the prediction accuracy.

Additionally, in the physical model, results of the numerical simulation greatly influence forecasting accuracy. The physical model is based on a large amount of historical data and requires specific and accurate physical information, such as pressure, temperature, and terrain, which may result in the systematic errors [48].

As for the time series methods, they, too, often require a large amount of historical data and face restrictions imposed by assumptions, such as normality postulates [49]. At the same time, models based on artificial intelligence often suffer from over-fitting or the difficulty of parameter setting. Moreover, over a long period, the existing forecasting models forecast wind speed by mostly using the original wind speed data recorded directly from wind farms, and as such, the high volatility of this data and outliers, which are not accounted for in the model, seriously influence the forecasting accuracy [50,51].

Hence, for the more accurate and stable forecasting results, a hybrid forecasting system, which combines a ‘decomposition and ensemble’ strategy and fuzzy time series model, is proposed in this paper. The proposed system includes two modules—data pre-processing and forecasting—to achieve better forecasting performance. In the data pre-processing module, ensemble empirical mode decomposition is employed to decompose the time series into finite number of intrinsic mode functions and reconstruct the raw wind data to overcome any non-stationary features. Next, in the forecasting module, a fuzzy time series, constructed by fuzzy sets, is developed to carry out wind speed forecasting. In fuzzy time series algorithm, a set of continuous numbers are assigned with linguistic value according to different interval partitioning methods which will also be discussed and compared in this paper. Furthermore, a set of comprehensive evaluating indicator system are established to compare different models’ performance. Accordingly, features of the developed hybrid forecasting system and our main contributions through this study are as follows:

1. A hybrid forecasting system is developed including two modules—data pre-processing and forecasting. Unlike previous time series models that dealt with continuous numbers, the fuzzy time series model is handled by fuzzy sets, which solve the weakness of traditional models requiring extensive historical data and assumptions. The effectiveness of this hybrid system is tested and is found to significantly enhance forecasting performance.
2. The pre-processing of raw data for wind speed forecasting makes significant contribution to forecasting accuracy. However, in most extant studies, the forecasting was often based on original data, which was not pre-processed. The volatility of and noise in unprocessed data seriously influence the forecasting accuracy and stability. The proposed hybrid system employs the ‘decomposition and ensemble’ strategy to effectively reduce noise in the wind speed time series signal. The results prove that eliminating the noise and uncertainty components from the original chaotic time series by pre-processing the raw data can remarkably improve the forecasting performance.
3. The forecasting performance of the fuzzy time series model is always influenced by the interval length, which in turn, depends on the discretization method. Therefore, to search for the most suitable discretization method for wind speed forecasting, four different interval partitioning methods of fuzzy time series have been discussed and compared. The results indicate that supervised discretization methods outperform unsupervised methods in most cases.
4. To obtain the best settings of the system, sensitivity analysis of the parameters of the hybrid system is performed, which demonstrates that by appropriately selecting the ensemble number, the white noise amplitude is found to increase forecasting accuracy.
5. The Diebold–Mariano (DM) test and forecasting effectiveness (FE) have been selected as testing methods, and the variance in the error is used to measure the stability of the forecasting results in addition to common evaluation metrics thereby enabling a more thorough evaluation of the proposed hybrid system.

3. Method

In this section, we describe all methods used in this study.

3.1. Data Pre-Processing Method—Ensemble Empirical Mode Decomposition

Wu and Wang [52] proposed the ensemble empirical mode decomposition in 2008, which was developed from the previous empirical mode decomposition with an intent to overcome the weakness of mode mixing. Empirical mode decomposition is a method to handle non-stationary signals, and was proposed in 1998 by Huang. Compared with wavelet analysis, empirical mode decomposition does not need to select the base function, and is a self-adaptive decomposition technique. Finite number of intrinsic mode functions can be obtained during the processing of raw signals. The intrinsic mode function time series can retain amplitude modulation information of the original signal sequence.

In addition, it must satisfy both conditions [53]—(1) in the entire sequence, the difference between the number of all maxima and minima and the number of zero-cross points is less than or equal to 1; and (2) the arithmetic mean of the upper envelope, obtained by the local maxima, and lower envelope, consisting of the local minima, is zero at each point.

However, the mode mixing phenomenon exists in empirical mode decomposition to represent either a single intrinsic mode function that includes components of various scales. On the contrary, a component of a similar scale may exist in disparate intrinsic mode functions. The ensemble empirical mode decomposition method eliminates the intermittent situation in the original time series by adding white noise, which not only improves the accuracy of the decomposed signal but also preserves the original information characteristics of the signal. The ensemble empirical mode decomposition is developed on the basis of auxiliary noise signal processing, and equalizes signals by adding small amplitude white noise effectively to overcome the mode mixing phenomenon of empirical mode decomposition [54]. The adaptive signal processing characteristics of ensemble empirical mode decomposition reduces the influence of human factors on the decomposition results. For the analysis of non-stationary and volatile time series, the ensemble empirical mode decomposition is especially applicable.

In line with the above description of the two methods, the sequence of steps followed during ensemble empirical mode decomposition are as follows [55]:

Step 1: Add the normal distribution white noise series to the signal that is to be decomposed.

Step 2: Decompose the signal with the added normal distribution white noise series into several intrinsic mode functions.

Step 3: Repeat *Step 1* and *Step 2*, and add a new white noise series each time.

Step 4: Regard the ensemble means of intrinsic mode functions that are obtained during decompositions as the final result.

It can be realized that the above algorithm depends on the amplitude of the added noise and ensemble times. When the amplitude of the added white noise is too low, the mode mixing problem cannot be suppressed, while if the amplitude is too high, more pseudo components will appear. In such a case, empirical mode decomposition is carried out causing the amount of calculation involved to increase greatly.

3.2. Forecasting Method—Weighted Fuzzy Time Series (FTS) Algorithm

The fuzzy time series algorithm is a common forecasting method owing to its easy calculations and great performance. Fuzzy time series are widely used in forecasting applications because of their capability of handling linguistic value datasets to obtain accurate forecasting. At present, it has been frequently and successfully used for forecasting nonlinear as well as dynamic datasets in various areas, including stock index [56], energy [57], course enrollment [58], green materials [59], load consumption [60], and so on. A fuzzy time series is defined by Song and Chissom [61] as follows.

Definition 1. $Y(t)$ ($t = 0, 1, 2, \dots$) is defined as a set of continuous numbers that is the universe of discourse and fuzzy sets $f_j(t)$ are constructed based on it. Then $F(t)$, a set of $f_1(t), f_2(t) \dots$, is regarded as the fuzzy time series which is defined on $Y(t)$.

Definition 2. $F(t)$ is assumed to be only caused by $F(t - 1)$. A forecasting model is described as $F(t) = F(t - 1) * R(t - 1, t)$, where $F(t - 1)$ and $F(t)$ are fuzzy sets and $R(t - 1, t)$ is the fuzzy logical relationship (FLR).

Definition 3. Let $F(t - 1) = A_i$ and $F(t) = A_j$. The fuzzy logical relationship (FLR) between two fuzzy values can be expressed as $A_i \rightarrow A_j$ where A_i and A_j represent the left-hand side (LHS) and right-hand side (RHS) of the FLR, respectively.

Definition 4. All single FLRs can be combined into several groups based on the same LHS of the FLR.

Then, the calculating steps of the weighted fuzzy time series can be described as in [62]:

Step 1: Determine the universe of discourse $U = [\min - a, \max + a]$, and then partition them into several intervals according to the interval partitioning methods mentioned above. From this, continuous data for further observations could be assigned linguistic values.

Step 2: Set a fuzzy membership function, and obtain the fuzzy set for actual continuous values. The fuzzy set A_i is defined based on intervals, as in [63].

$$\begin{aligned} A_1 &= 1/u_1 + 0.5/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0/u_9 + 0/u_{10} \\ A_2 &= 0.5/u_1 + 1/u_2 + 0.5/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0/u_9 + 0/u_{10} \\ &\vdots \\ A_{10} &= 0/u_1 + 0/u_2 + 0/u_3 + 0/u_4 + 0/u_5 + 0/u_6 + 0/u_7 + 0/u_8 + 0.5/u_9 + 1/u_{10} \end{aligned}$$

Step 3: Fuzzify observations. For example, the fuzzified result of one data is A_j when the maximum degree of membership of this data is in A_j .

Step 4: Determine the fuzzy logical relationships and group them. For example, if $A_i \rightarrow A_j, A_i \rightarrow A_k, A_i \rightarrow A_l$ can be grouped as $A_i \rightarrow A_j, A_k, A_l$.

Step 5: Establish weights. From step 4 above, the weight matrix can be obtained and further standardized. The defuzzified matrix can then be calculated by applying the centroid defuzzification method.

Step 6: Calculate forecasting results. Forecasting results can be calculated by multiplication of the defuzzified and standardized weighting matrices defined as follows:

$$\begin{aligned} W_s(t) &= (\hat{W}_1, \hat{W}_2, \dots, \hat{W}_k) \\ \hat{W}_i &= W_i / \sum_{i=1}^k W_i \end{aligned} \quad (1)$$

$$F(t) = D(t-1) * W_s(t-1) \quad (2)$$

Here, W_s is the standardized weighting matrix, D is the defuzzified matrix. W_i represents the unstandardized weighting matrix elements, while \hat{W}_i represents standardized ones, and $F(t)$ is the forecasting result.

Step 7: Lastly, forecasted values obtained above are amended by employing Equation (3) to obtain the ultimate forecasting result.

$$F_u(t) = y(t-1) + \alpha * (F(t) - y(t-1)) \quad (3)$$

where $y(t-1)$ is the actual value on time $t-1$, and F_s is the ultimate forecasting value.

3.3. Interval Partitioning Methods

The forecasting performance of the fuzzy time series model is influenced by interval length, and determination of the appropriate interval partitioning method is supposedly a challenging task [64]. However, interval partitioning methods, in turn, depend upon discretization methods and the selection of cut points [65].

Data discretization is a vital method that can reduce the actual demand of storage space for an obtained continuous data set by dividing it into finite number of intervals, which possess a high level of class coherence, and then assigning linguistic values to these intervals [66]. Data discretization comprises two main tasks—(1) determination of the number of disjoint intervals or cut points, which are generally obtained according to a heuristic rule; (2) finding boundaries of the intervals; that is, the interval range.

To date, various discretization methods have been developed owing to different needs, and these can be roughly classified into supervised and unsupervised methods. Supervised methods partition the

continuous data depending upon class information, while unsupervised methods need not follow the same methodology. Supervised discretization can be further divided into entropy or Chi-square-based discretization, while unsupervised discretization includes equal width and equal frequency interval discretization methods [67–70]. In the current fuzzy time series model, the equal width interval discretization method is frequently employed, and the supervised discretization methods are seldom used [71].

3.3.1. Equal Width Interval Algorithm

The equal-width (EW) interval algorithm is the simplest unsupervised discretization method. According to the number of intervals designated by the user, the range of the sorted numerical attributes denoted as (X_{\min}, X_{\max}) is divided into K equal sized intervals. Thus, the width of each interval is $(X_{\max} - X_{\min})/K$. However, when there exist points with considerable skewness, this method is not adaptive. The disadvantage of this method, caused by the uneven distribution of the time series, is that the data count in different intervals may vary significantly [72].

3.3.2. Equal Frequency Interval Algorithm

The equal frequency (EF) interval algorithm is similar to the equal width interval algorithm in that it also divides the sorted numerical attributes into K intervals. The difference, in this method, is that each interval now includes the same number (i.e., n/k) of objects with adjacent values, where n is the total data count [72]. In the equal frequency method, the same data point that occurs many times could be divided into different intervals. The method, known as the proportional k -interval discretization, attempts to avoid this restriction of the equal-width interval discretization. It separates the domain in intervals using similar data point distribution. The data points with the same value are assigned to the same interval. Therefore, some intervals may not always possess equal frequencies.

3.3.3. Entropy-Based Discretization Algorithm

The entropy-based discretization algorithm, proposed by Fayyad and Irani, relies on the class information of continuous numerical attributes, which is used for calculating and determining the cut points [73]. As it adopts a top-down splitting technique, this method partitions the interval into smaller intervals recursively until the stopping criterion, such as the Minimum Description Length Principle or Mutual Information Theory, is met [74].

The entropy-based method selects points for discretization depending on the class information entropy of candidate partitions. Information entropy is a measure of the degree of ordering of the system, and class information entropy measures the quantity of information that is required to determine which class a sample should belong to.

The steps of this algorithm can be described as follows:

Step 1: Define the entropy of intervals. For an object set T , the entropy function is calculated as under:

$$\mathbf{Entropy}(T) = -\sum_{i=1}^n p_i \cdot \log(p_i) \quad (4)$$

where n is the number of the data in set T and p_i is the probability of class i .

Step 2: Apply all possible cut points to divide the data into two parts, and from all possible cut methods, find the one with minimum entropy. For each cut point, the entropy of each split is defined as:

$$\mathbf{Entropy}(T|split) = p_{left} \cdot \mathbf{Entropy}(T_{left}) + p_{right} \cdot \mathbf{Entropy}(T_{right}) \quad (5)$$

where p_{left} and p_{right} represent probabilities of the left (T_{left}) and right (T_{right}) sets, respectively.

Step 3: Regard the two intervals obtained in step 2 as independent intervals and then repeat step 1.

Step 4: Run iterations, but stop the process when the set criterion is achieved.

3.3.4. Chi-Square-Based Discretization Algorithm

Chi-square (χ^2) is a discretization algorithm based on the value of Chi-square, which measures the relationship between a class and adjacent intervals. The Chi-square-based discretization algorithm splits the data set based on user-defined significance levels. This algorithm includes the top-down (Chi-split) and bottom-up (Chi-merge) methods, both of which are based on Chi-square. The top-down method regards the entire interval value as a discrete value and then split this interval into two adjacent sub-intervals. The process then runs into iterations and stops once a set criterion is achieved. When the Chi-square test is significant, the split must continue; otherwise, it should be stopped. Contrary to the top-down approach, the bottom-up method regards each attribute value as a discrete value and then repeatedly merges adjacent attribute values, if the two are statistically similar, until the stopping condition is met. The stopping criterion is determined by a Chi-square threshold defined by user to stop the merge operation when two adjacent intervals cannot be proven to be sufficiently similar [66].

Chi-square (χ^2) is a statistic to test the independence between row and column variables in a contingency table, as presented in Table 1. In the Chi-Square-based discretization algorithm, the formula to calculate χ^2 statistic at a cut point for two adjacent intervals is described in Equation (6) [75].

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (6)$$

c is the classes number.

O_{ij} is the example number in the i th interval and j th class.

E_{ij} is the expected frequency in the i th interval and j th class, computed by $E_{ij} = (R_i C_j)/N$.

R_i represents the example number in the i th interval.

C_j represents the examples number in the j th class.

When we apply the Chi-square to test the statistical independence of two variables, the confidence level is supposed to be artificially set. Too high confidence level will lead to excessive discretization, whereas it will lead to insufficient discretization. Moreover, a common deficiency of the Chi-merge approach is that it can only merge two adjacent intervals in each loop; thus, the discretization speed is slow when the number of samples is very large.

Table 1. The contingency table of Chi-square analysis.

	Class 1	Class 2	Class c	Sum
Interval 1	O_{11}	O_{12}	O_{1c}	R_1
Interval 2	O_{21}	O_{22}	O_{2c}	R_2
Sum	C_1	C_2	C_c	N

4. Data Description and Setup

To specifically evaluate and compare the ability and performance of the fuzzy time series models under different interval partitioning methods, three primary different wind speed time series datasets obtained from a wind farm located at Penglai in Shandong Province of China are selected. Shandong is surrounded by the sea on three sides, and is located in China's coastal wind belt, where wind resources are very rich. As such, prospects of wind power development in this region are extremely broad. The installed wind energy capacity of this region is about 67 million KW. Penglai, a part of Yantai, Shandong Province, located at $37^{\circ}48'$ N and $129^{\circ}45'$ E, belongs to the Northern temperate East Asian monsoon region continental climate and hilly area, which is south-high and north-low, possessing rich wind resources and many wind farms. The installed wind capacity of Yantai was 2104.15 MW in July 2016, and the wind power scale is the largest among power grids in the Shandong peninsula. Thus, it is crucial to accurately forecast the wind speed in this region. Accordingly, two thousand data

points with the sampling interval is 10 min and sampling frequency is 144 times per day were selected from each dataset recorded from 10:00, 1 January 2011 to 7:10, 15 January 2011 including training set (1500 samples) and the testing set (500 samples).

Features of the three wind speed datasets are listed in Table 2 and are visualized via the box and line charts in Figure 2. As described, all three datasets possess large fluctuations and are divided into training and testing samples. From the box chart, it is seen that Dataset III possesses the maximum degree of dispersion and the opposite is true for Dataset I. Table 2 presents numerical values of some statistical indicators; the standard deviations are approximately 2 m/s, and the interquartile ranges are mostly above 3 m/s. Both these values indicate significant fluctuations in the wind speed. This evident fluctuation in the wind speed datasets verifies the challenges involved in wind speed forecasting.

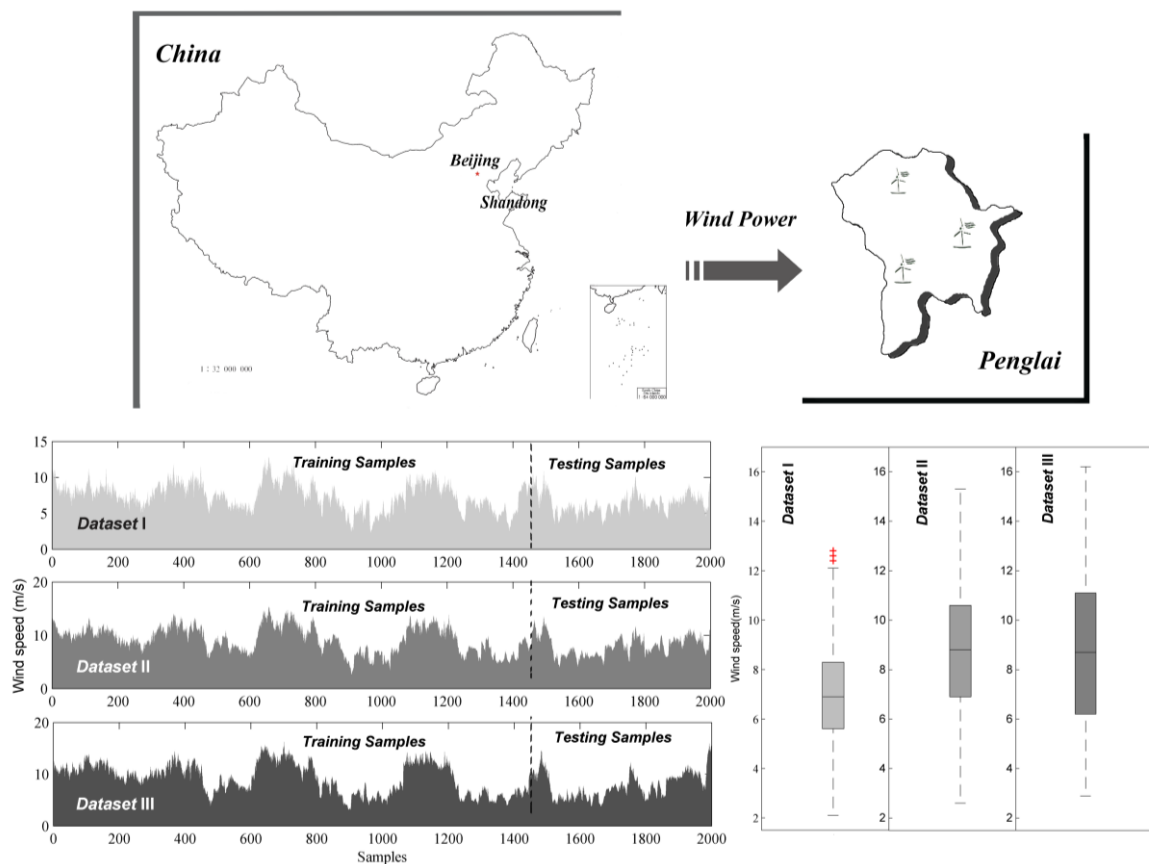


Figure 2. Data description of study sites in Penglai, Shandong Province of China.

For the fuzzy time series model and subsequent interval partitioning methods, the universe of discourse for wind data is defined as (2, 16.5). Wind-data intervals corresponding to four different interval partitioning methods are listed in Table 3.

The continuous values are transformed into 10 linguistic values A_1 – A_{10} . Taking the Chi-square-based discretization of Dataset III, the fuzzy relationship groups are summarized in Table 4. Each number in the matrix indicates the occurrence of a fuzzy logic relationship. Based on this matrix and Equation (1), the weight matrix can be calculated, as presented in Tables 4 and 5. Ultimately, forecasting values can be calculated by Equations (2) and (3). After repeated tests, the weight in Equation (3) was set as 0.5.

Table 2. Some statistical indicators of the Datasets.

Datasets		Numbers	Statistical Indicators				
			Maximum (m/s)	Minimum (m/s)	Mean (m/s)	Interquartile Range (m/s)	Std. (m/s)
<i>Equation</i>		-	-	-	$Mean = \frac{\sum_{i=1}^N x_i}{N}$	$Q_d = Q_U - Q_L$	$S = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$
<i>Dataset I</i>	All	2000	12.8	2.1	6.9815	2.7	1.8202
	Training	1500	12.8	2.1	7.2781	2.7	1.8852
	Testing	500	10.2	3.1	6.0918	1.5	1.2401
<i>Dataset II</i>	All	2000	15.3	2.6	8.7764	3.7	2.3675
	Training	1500	15.3	2.6	9.1389	3.9	2.4516
	Testing	500	12.2	3.9	7.6890	2.8	1.6787
<i>Dataset III</i>	All	2000	16.2	2.9	8.7374	4.9	2.8693
	Training	1500	16.2	2.9	9.1703	4.9	2.9067
	Testing	500	15.9	3.7	7.4384	3.5	2.312

Table 3. The intervals of the four interval partitioning methods.

Methods	Equal Width	Equal Frequency	Entropy Based	Chi-Square Based
u_1	(2.00, 3.45)	(2.00, 5.00)	(2.00, 3.90)	(2.00, 3.90)
u_2	(3.45, 4.90)	(5.00, 5.80)	(3.90, 5.00)	(3.90, 5.10)
u_3	(4.90, 6.35)	(5.80, 6.50)	(5.00, 6.40)	(5.10, 6.20)
u_4	(6.35, 7.80)	(6.50, 7.10)	(6.40, 7.40)	(6.20, 7.30)
u_5	(7.80, 9.25)	(7.10, 7.80)	(7.40, 8.90)	(7.30, 8.80)
u_6	(9.25, 10.70)	(7.80, 8.60)	(8.90, 9.80)	(8.80, 9.70)
u_7	(10.70, 12.15)	(8.60, 9.50)	(9.80, 10.90)	(9.70, 10.60)
u_8	(12.15, 13.60)	(9.50, 10.40)	(10.90, 12.80)	(10.60, 11.90)
u_9	(13.60, 15.05)	(10.40, 11.70)	(12.80, 15.30)	(11.90, 13.60)
u_{10}	(15.05, 16.50)	(11.70, 16.20)	(15.30, 16.20)	(13.60, 16.20)

Table 4. Fuzzy relationship groups and weight matrix before standardization.

P_{t-1}	P_t									
	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_1	10	7	0	0	0	0	0	0	0	0
A_2	6	114	18	3	0	0	0	0	0	0
A_3	1	20	119	29	3	0	0	0	0	0
A_4	0	0	32	97	27	1	0	1	0	0
A_5	0	0	3	29	106	26	7	0	0	0
A_6	0	0	0	0	31	60	35	8	0	0
A_7	0	0	0	0	4	34	50	52	5	0
A_8	0	0	0	0	0	13	50	137	56	2
A_9	0	0	0	0	0	0	4	56	145	27
A_{10}	0	0	0	0	0	0	0	4	25	42

Table 5. The standardized weight matrix.

	A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9	A_{10}
A_1	0.5882	0.4118	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
A_2	0.0426	0.8085	0.1277	0.0213	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
A_3	0.0058	0.1163	0.6919	0.1686	0.0174	0.0000	0.0000	0.0000	0.0000	0.0000
A_4	0.0000	0.0000	0.2025	0.6139	0.1709	0.0063	0.0000	0.0063	0.0000	0.0000
A_5	0.0000	0.0000	0.0175	0.1696	0.6199	0.1520	0.0409	0.0000	0.0000	0.0000
A_6	0.0000	0.0000	0.0000	0.0000	0.2313	0.4478	0.2612	0.0597	0.0000	0.0000
A_7	0.0000	0.0000	0.0000	0.0000	0.0276	0.2345	0.3448	0.3586	0.0345	0.0000
A_8	0.0000	0.0000	0.0000	0.0000	0.0000	0.0504	0.1938	0.5310	0.2171	0.0078
A_9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0172	0.2414	0.6250	0.1164
A_{10}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0563	0.3521	

5. Experimental Results for Datasets

For the simulation, wind speed data was recorded at 10-min intervals thereby obtaining three different datasets—Datasets I, II, and III. By considering Dataset I in our analysis, line charts of the fuzzy time series forecasted values, with different interval lengths, are shown in Figure 3.

- (1) The top half of Figure 3 presents forecasting results of the original data and that of data preprocessed via ensemble empirical mode decomposition employing fuzzy time series forecasting methods—entropy-based discretization, Chi-square-based discretization, equal frequency interval discretization, and equal width interval discretization. It is obvious that forecasting results obtained using fuzzy time series under supervised discretization methods tend to match actual values more closely compared to the unsupervised methods. The details of parts A and B in Figure 3 illustrate the local enlargement comparison of the different methods.
 - (a) As shown in Figure 3, compared to equal width interval discretization, forecasting curves of the entropy- and Chi-square-based discretization more closely follow the shape of the actual testing curve. Equal frequency interval discretization demonstrates the worst performance. Thus, supervised discretization methods are, in general, found to be superior to unsupervised methods.
 - (b) Better forecasting is achieved when the wind speed is steady without any sudden change. Evidently, the forecasting system perform better between sample numbers 130–170 and 300–350, and better follow the shape of the actual testing curve.
 - (c) Comparing the curves of the original and pre-pre-processing data, the degree of overlap of the curves in the second picture is evidently superior to that in the first. Thus, it can be seen that data pre-processing plays a vital role in wind speed forecasting.
 - (d) As shown in parts A and B in Figure 3, the degree of overlap of the curves near the local maximum forecasting value is better than that near the local minimum forecasting value. Near the local minimum forecasting value, the curve corresponding to equal frequency interval discretization, when compared to other curves, deviates considerably from the actual value curve.
- (2) The lower part of Figure 3 demonstrates the forecasting error (forecast value minus actual value) for the four different interval partitioning methods described in this paper.
 - (a) In terms of individual forecasting values, the forecasting error is notably large, such as that calculated for sample numbers 100, 250, and 300, wherein there exist large fluctuations in wind speed. It is conclude that the performance of forecasting methods is poor when large fluctuations are present in data.
 - (b) It is noteworthy that the forecasting error for pre-processed data is significantly less compared to original data. All points distribute around a zero-scale line. The points in the right image are also more concentrated than in the left one. It is to be noted that most points, which deviate from the zero-scale line, further belong to the equal frequency interval discretization method.

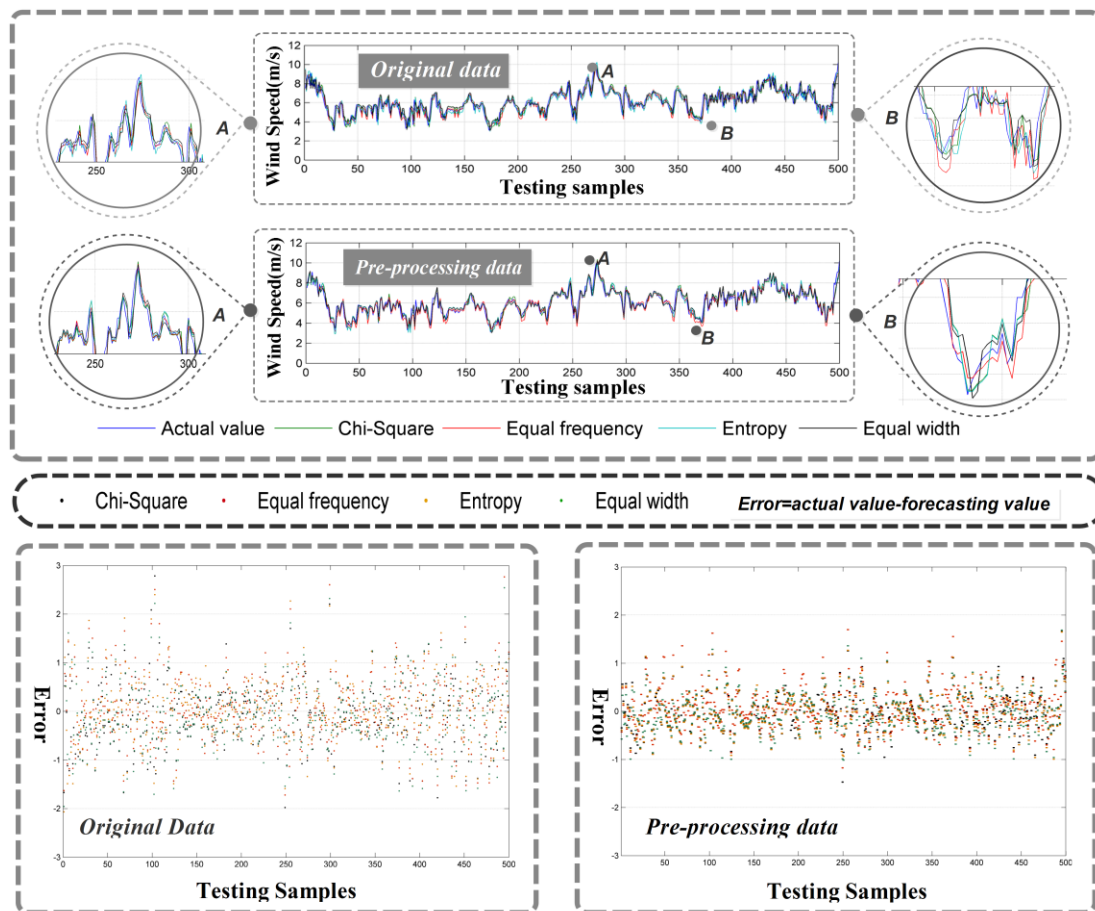


Figure 3. Forecasting results and error in fuzzy time series with different interval lengths using original and pre-processing data in Dataset I.

6. Analysis and Discussion

In this section, the performance of the different methods from computational aspect is discussed. Moreover, the frequency of data sampling plays a vital role in wind data. According to State Grid Dispatching scheduling and the energy industry standard NB/T31046-2013 which was formulated by National energy administration in China, 144 wind speed data should be obtained per day (24 h). And the wind energy measurement rule was set in 2013. The time interval of wind speed data obtained from wind farm is supposed to be no less than ten minute. Due to the non-storage of wind energy, short wind speed forecasting can warn dispatchers to carry out some necessary operation in a critical state to avoid economic losses and safety accidents as much as possible for the stable operation of power system. Accordingly, in this section, ten min wind speed data from three sites is selected to evaluate the performance of the models.

Several metrics have been employed by researchers in extant studies for error evaluation. However, there is no common standard to evaluate the forecasting performance of different methods. Therefore, various criteria are utilized to compare the forecasting performance. These criteria are defined in Table 6. MAE measures the difference between the forecasting values and observations; RMSE measures the deviation between observations and forecasted values, and it is more easily affected by extreme values than MAE; MAPE is the average of absolute percentage error to evaluate the forecasting accuracy in statistics; IA is a dimensionless index to compare different models and is selected as a substitutes for R or R^2 ; and VAR measures the stability of the methods. Furthermore, MAE, RMSE, MAPE, and VAR are negative indicators; i.e., the lower the better, while IA is a positive indicator.

Table 6. Specific definitions of error criteria.

Metric	Definition	Equation
MAE	The mean absolute error of forecasting results	$\mathbf{MAE} = \frac{1}{N} \sum_{i=1}^N y_i - \hat{y}_i $
RMSE	The root mean square value of the errors	$\mathbf{RMSE} = \sqrt{\frac{1}{N} \times \sum_{i=1}^N (y_i - \hat{y}_i)^2}$
MAPE	The average of absolute percentage error	$\mathbf{MAPE} = \frac{1}{N} \sum_{i=1}^N \left \frac{y_i - \hat{y}_i}{y_i} \right \times 100\%$
IA	The index of agreement of forecasting results	$\mathbf{IA} = 1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (\hat{y}_i - \bar{y} + y_i + \bar{y})^2}$
VAR	The variance of the forecasting error	$\mathbf{Var} = E(\hat{y} - E(\hat{y}))^2$

6.1. Experiment I: The Data Pre-Processing for Fuzzy Time Series Forecasting

The high volatility and instability of wind speed data undoubtedly increases the challenge in accurate forecasting. As a consequence, in the process of data analysis, it is necessary to process the original data according to specific analysis requirements. In this study, the ensemble empirical mode decomposition is utilized to pre-process original data thereby effectively reducing the influence of instability and noise. We set the ensemble number as 100 and noise amplitude as 0.2. As can be seen in Figure 4a, it is obvious that pre-processing data achieves better forecasting performance, and the variance in forecasting errors drops significantly. For a more direct and clear cognition, the improvement ratio of the indexes can be calculated using Equation (7):

$$\left| \frac{Index_{compared} - Index_{proposed}}{Index_{compared}} \right| \times 100\% \quad (7)$$

Table 7 quantitatively summarizes the improvement in forecasting performance through data pre-processing. In terms of MAE, RMSE, and MAPE, the average improvement ratio is about 30–40%, the highest being 38.86%, which is achieved under equal width interval discretization. In terms of IA, the average improvement ratio is relatively low—about 2% for Datasets II and III and 5% for Dataset I. This may be due to values of this index being large originally. Variance (VAR) demonstrates the highest average improvement ratio (about 60%) with the highest individual value being 62.43%. This proves that data pre-processing significantly improves the forecasting stability.

Remark 1: *The high volatility and instability of wind speed data affects the forecasting results significantly. Thus, suitable data pre-processing method can improve the forecasting performance greatly especially the stability of the forecasting results.*

Table 7. Improvement ratios of the different error criteria for the pre-processing strategy.

	MAE	RMSE	MAPE	IA	VAR
Dataset I					
FTS-Chi ²	36.82%	37.89%	37.14%	4.69%	61.33%
FTS-Entropy	36.68%	35.16%	35.63%	4.54%	58.92%
FTS-EF	35.46%	37.29%	35.66%	4.73%	60.59%
FTS-EW	37.65%	38.86%	37.90%	5.30%	62.43%
Dataset II					
FTS-Chi ²	31.81%	33.64%	31.17%	1.79%	55.94%
FTS-Entropy	33.13%	34.61%	31.89%	1.79%	57.23%
FTS-EF	29.62%	31.38%	29.09%	1.66%	52.89%
FTS-EW	28.27%	29.99%	26.99%	1.76%	50.86%
Dataset III					
FTS-Chi ²	32.09%	33.65%	29.93%	1.12%	55.97%
FTS-Entropy	34.54%	35.82%	31.22%	1.28%	59.58%
FTS-EF	32.25%	33.15%	30.72%	1.15%	55.31%
FTS-EW	32.08%	33.45%	31.17%	1.16%	55.56%

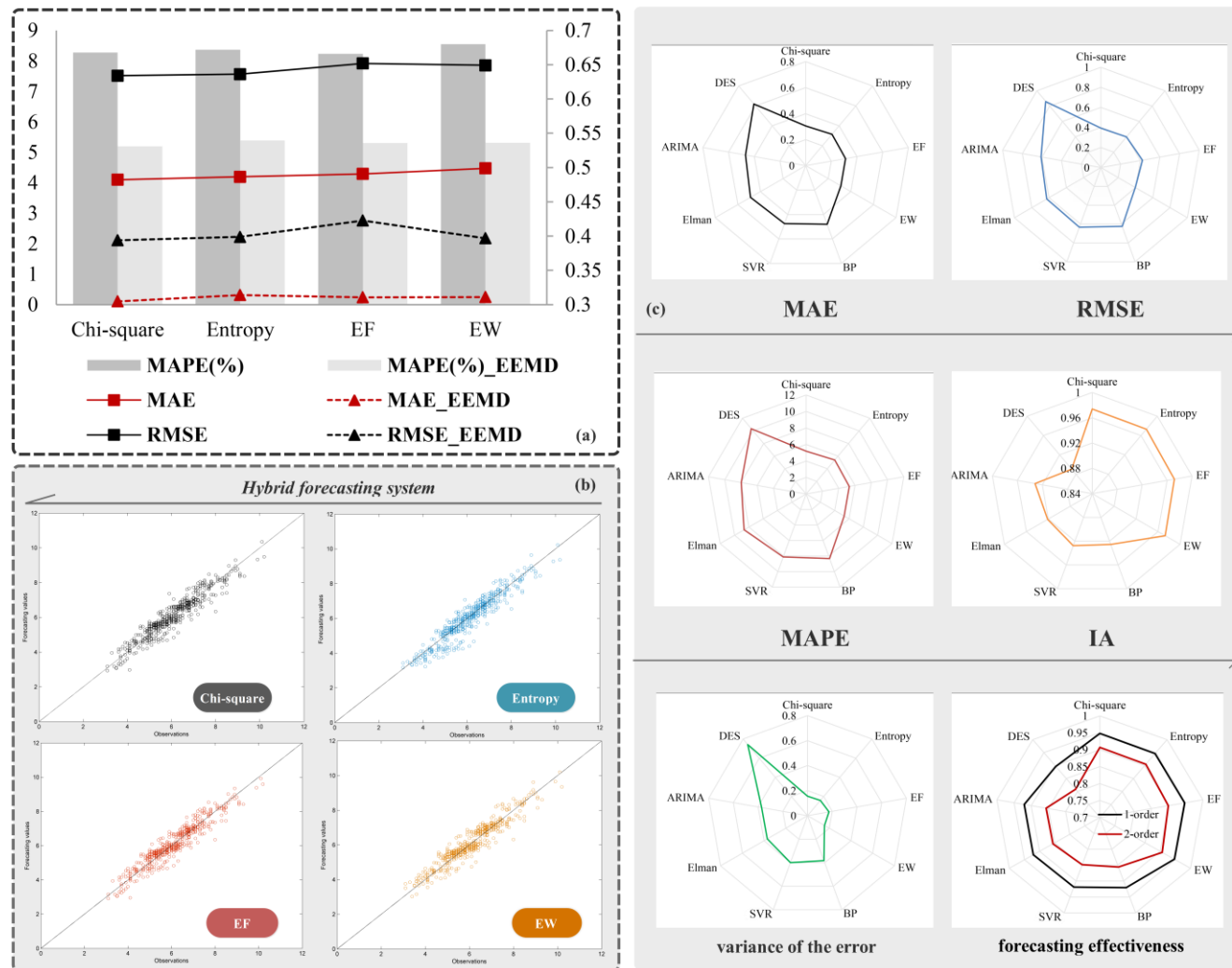


Figure 4. Comparison of forecasting results obtained using different models for Dataset I. (a) Comparison of the forecasting results obtained from original and pre-processing data; (b) Comparison of actual and forecasting values of hybrid forecasting system; (c) Comparison of forecasting performance for different models

6.2. Experiment II: The Comparison of Fuzzy Time Series, Artificial Neural Network, Statistical Models and Support Vector Regression

Owing to the widespread popularity of artificial intelligence, statistical models, and Support Vector Regression (SVR), this experiment was designed to compare the performance of the proposed hybrid forecasting system against artificial intelligence (Back Propagation Neural Network (BPNN), Extreme Learning Machine (ELM), and Elman) and statistical (Double Exponential Smoothing (DES) and Autoregressive Integrated Moving Average (ARIMA) models. In all artificial intelligence models, the node-point numbers of input and output layers are set as 5 and 1, respectively. For hidden layers in BPNN, ELM, and Elman, the node-point numbers are, respectively, set as 2, 20, and 14. For the ARIMA (p, d, q) model, values of p, d , and q are set as 4, 1, and 5, respectively, in confirmation with the A-Information Criterion (AIC) and the stationary test. In SVR, the radial basis function (RBF) is selected as kernel function. The precise parameter settings are listed in Table 8 and other parameters use the default setting.

Table 8. Experimental parameter values in different models.

Model	Experimental Parameter	Value
BPNN	Maximum number of iteration times	1000
	Learning rate	0.01
	Training accuracy goal	0.00001
	Node-point number of input layer	5
	Node-point number of hidden layer	2
	Node-point number of output layer	1
ELM	Node-point number of input layer	5
	Node-point number of hidden layer	20
	Node-point number of output layer	1
Elman	Node-point number of input layer	5
	Node-point number of hidden layer	14
	Node-point number of output layer	1
	Iteration number of display once in an image	20
	Maximum number of iteration times	1000
SVR	Node point number of input layer	5
	Node point number of output layer	1
	Type of SVR model	epsilon-SVR
	Type of kernel function	RBF
	Parameter of epsilon-SVR	4
ARIMA (p, d, q)	Autoregressive term (p)	4
	Moving average number (q)	5
	Difference times (d)	1
DES	Smoothing coefficient	0.9

Results of the abovementioned comparison are presented in Table 9. Considering Dataset I, the proposed hybrid forecasting system achieves the optimum MAPE value amongst the models compared. As shown in Figure 4c, we can easily see that DES demonstrates the worst performance and its corresponding MAPE increases by about 5% when compared to the proposed hybrid forecasting system. The proposed system betters the performance of all models in terms of other indexes too. Amongst artificial neural networks, ELM achieves better forecasting accuracy and stability, while Elman performs relatively poorly. DES also exhibits the largest variance of the forecasting error indicating that the forecasting accuracy of the DES is unstable when compared to, both, the proposed forecasting system as well as artificial neural networks.

Table 9. Comparison of the hybrid forecasting system against artificial intelligence, statistical, and persistence model.

<i>Dataset I</i>	Hybrid Forecasting System				Artificial Neural Network			SVR	Statistical		Persistence Model
	Chi ²	Entropy	EF	EW	BPNN	ELM	Elman		ARIMA (4,1,5)	DES	
MAE	0.304745	0.314015	0.310712	0.311066	0.480500	0.464689	0.488986	0.474854	0.468817	0.618175	0.458800
RMSE	0.393810	0.398935	0.422710	0.396898	0.624371	0.612871	0.626326	0.633335	0.608993	0.860441	0.617997
MAPE (%)	5.199317	5.384246	5.300246	5.310732	8.313917	7.834408	8.664982	8.114452	7.966683	10.328762	7.749842
IA	0.973831	0.973006	0.971658	0.973076	0.925417	0.930872	0.921294	0.927558	0.931646	0.890443	0.934266
VAR	0.155138	0.159302	0.173445	0.156431	0.382100	0.374668	0.370987	0.399316	0.371603	0.741841	0.382683
<i>Dataset II</i>	Hybrid Forecasting System				Artificial Neural Network			SVR	Statistical		Persistence Model
	Chi ²	Entropy	EF	EW	BPNN	ELM	Elman		ARIMA (4,1,5)	DES	
MAE	0.303159	0.305047	0.292344	0.327601	0.429400	0.419508	0.485842	0.437043	0.422248	0.531832	0.411800
RMSE	0.385926	0.394417	0.382119	0.420453	0.572332	0.559468	0.623497	0.598708	0.563047	0.722019	0.554923
MAPE (%)	4.040435	4.063754	3.933468	4.372152	5.727318	5.513481	8.556761	5.842899	5.607781	6.913070	5.417698
IA	0.986645	0.986254	0.987135	0.984222	0.969177	0.970732	0.923628	0.966562	0.969659	0.955681	0.971915
VAR	0.149199	0.155871	0.146307	0.177116	0.325634	0.312639	0.374678	0.356416	0.317654	0.522351	0.308557
<i>Dataset III</i>	Hybrid forecasting System				Artificial Neural Network			SVR	Statistical		Persistence Model
	Chi ²	Entropy	EF	EW	BPNN	ELM	Elman		ARIMA (4,1,5)	DES	
MAE	0.319252	0.326514	0.336218	0.334258	0.465963	0.452667	0.521913	0.489951	0.465122	0.601174	0.456400
RMSE	0.419349	0.431265	0.435344	0.425761	0.628820	0.613350	0.695743	0.650611	0.623680	0.829816	0.618935
MAPE (%)	4.655949	4.692781	5.037214	4.771814	6.583105	6.338355	7.535347	6.924305	6.555370	8.308627	6.386741
IA	0.991649	0.991099	0.991225	0.991315	0.980192	0.981376	0.974440	0.979072	0.971571	0.968741	0.981649
VAR	0.176199	0.186327	0.184881	0.181607	0.396132	0.376780	0.481291	0.423972	0.389664	0.689968	0.383807

In real world forecasting applications, the conventional statistical model may not be suitable owing to its inherent nonlinearity and instability. The use of artificial neural networks usually requires setting of many parameter values which significantly affects the forecasting performance; also, the forecasting results are different for several experiments conducted using the same sample. Additionally, in certain complex networks, the response time of the model substantially long. This may be considered as a drawback, since the timeliness of forecasting results is of critical importance in modern economic and industrial applications, especially in the energy sector.

To further demonstrate the performance of the proposed forecasting system, the persistence model, one of the most popular and frequently utilized benchmark methods, has been used as the benchmark test in our study. The persistence model simply assumes that forecasted value at any time t is identical to the last observation. The model does not require any parameter setting nor does it involve exogenous variables. Nonetheless, it usually demonstrates great performance [76,77]. Comparison results presented in Table 9 indicate that the proposed hybrid forecasting system demonstrates better forecasting performance in terms of all five model evaluation criteria. It can, thus, be concluded that the proposed hybrid forecasting system performs better than the benchmark persistence model.

Remark 2: *Comparing with the artificial neural network, statistical models, Support Vector Regression and persistence model, the proposed hybrid forecasting system possesses the better forecasting accuracy and stability than others. Moreover, unlike the traditional time series models which need a large amount of historical data and have restrictions of linear or normality postulates assumptions, and artificial neural network which have many parameters and complex structure, the proposed hybrid forecasting system has the advantage of the simple calculation and stable result ensuring the timeliness and reliability of the forecasting results.*

6.3. Experiment III: Forecasting Performance of the Fuzzy Time Series with Different Interval Partitioning Methods

Table 10 enlists the forecasting results in terms of MAE, RMSE, MAPE, VAR, and IA for original as well as pre-processed data using the four previously described discretization algorithms—Chi-square-based discretization (χ^2), entropy based discretization, equal frequency interval discretization, and equal width interval discretization. Most of the metrics indicate that the Chi-square-based discretization performs the best for Datasets I and III. For dataset II, the entropy-based discretization method demonstrates the best forecasting performance for original data, while the equal frequency interval discretization rules the roost in handling pre-processed data. Figure 4 shows the forecasting results graphic of the three datasets. From Table 10 and Figure 4a, it can be concluded that supervised discretization methods possess better stability and forecasting accuracy compared to unsupervised methods. In Figure 4b, scatter plot of the observations and values forecasted by the proposed hybrid forecasting system indicates that the proposed system demonstrates great performance.

Remark 3: *The forecasting results of the fuzzy time series with four different interval partitioning methods do not have large difference but the supervised discretization methods outperform than unsupervised discretization methods and the equal frequency interval discretization has the worst performance in general.*

Table 10. Comparison of fuzzy time series using different interval partitioning methods.

<i>Dataset I</i>	Original Data				Hybrid Forecasting System			
	FTS-Chi ²	FTS-Entropy	FTS-EF	FTS-EW	EEMD-FTS-Chi ²	EEMD-FTS-Entropy	EEMD-FTS-EF	EEMD-FTS-EW
MAE	0.482308	0.486552	0.490665	0.498894	0.304745	0.314015	0.310712	0.311066
RMSE	0.634074	0.651953	0.636203	0.649179	0.39381	0.42271	0.398935	0.396898
MAPE (%)	8.270632	8.36813	8.234151	8.551709	5.199317	5.384246	5.300246	5.310732
IA	0.930179	0.929037	0.929454	0.924069	0.973831	0.973006	0.971658	0.973076
VAR	0.401151	0.404214	0.422243	0.416326	0.155138	0.159302	0.173445	0.156431
<i>Dataset II</i>	Original Data				Hybrid Forecasting System			
	FTS-Chi ²	FTS-Entropy	FTS-EF	FTS-EW	EEMD-FTS-Chi ²	EEMD-FTS-Entropy	EEMD-FTS-EF	EEMD-FTS-EW
MAE	0.444548	0.433445	0.437201	0.456687	0.303159	0.305047	0.292344	0.327601
RMSE	0.581522	0.584344	0.57475	0.600577	0.385926	0.382119	0.394417	0.420453
MAPE (%)	5.870509	5.730971	5.774887	5.988826	4.040435	4.063754	3.933468	4.372152
IA	0.969249	0.970143	0.96978	0.967156	0.986645	0.986254	0.987135	0.984222
VAR	0.338608	0.330869	0.34208	0.360426	0.149199	0.155871	0.146307	0.177116
<i>Dataset III</i>	Original Data				Hybrid Forecasting System			
	FTS-Chi ²	FTS-Entropy	FTS-EF	FTS-EW	EEMD-FTS-Chi ²	EEMD-FTS-Entropy	EEMD-FTS-EF	EEMD-FTS-EW
MAE	0.470124	0.481921	0.513594	0.492126	0.319252	0.326514	0.336218	0.334258
RMSE	0.632072	0.678365	0.645095	0.639762	0.419349	0.435344	0.431265	0.425761
MAPE (%)	6.644708	6.774126	7.323562	6.932627	4.655949	4.692781	5.037214	4.771814
IA	0.980658	0.979836	0.97873	0.979918	0.991649	0.991099	0.991225	0.991315
VAR	0.400191	0.416966	0.457436	0.408696	0.176199	0.186327	0.184881	0.181607

6.4. Experiment IV: Testing Based on the DM Test and Forecasting Effectiveness

Although the evaluation metrics presented in experiment II have been well compared to evaluate the forecasting performance of the different forecasting models, the performance of these models has been further studied using statistical testing methods based on the DM test and forecasting effectiveness (FE). This section discusses these methods thereby enabling a more comprehensive test and comparison of the models' performance.

6.4.1. DM Test

The Diebold–Mariano test, which focuses on forecasting accuracy, is used to test the difference between the proposed system's forecasting accuracy and that of other methods [78].

The test is described as follows:

$$\begin{aligned} H_0 &: E(d_h) = 0, \forall n \\ H_1 &: E(d_h) \neq 0, \exists n \end{aligned} \quad (8)$$

Statistic values of the DM test are described by:

$$DM = \frac{\sum_{h=1}^k \left(L(\varepsilon_{t+h}^{(i)}) - L(\varepsilon_{t+h}^{(j)}) \right) / k}{\sqrt{S^2/k}} s^2 \quad (9)$$

ε_{t+h} denotes the forecasting error

S^2 denotes the estimation value for the variance of $d_h = L(\varepsilon_{t+h}^{(i)}) - L(\varepsilon_{t+h}^{(j)})$

L denotes a loss function that is utilized to represent the forecasting accuracy of the model.

Absolute deviation error loss and square error loss are two popular loss functions, which are widely employed.

Absolute deviation loss:

$$L(\varepsilon_{t+h}^{(i)}) = \left| \varepsilon_{t+h}^{(i)} \right| \quad (10)$$

Square error loss:

$$L(\varepsilon_{t+h}^{(i)}) = \left(\varepsilon_{t+h}^{(i)} \right)^2 \quad (11)$$

When there is no significant difference between forecasting performance of the compared models, we will reject the null hypothesis given by

$$|DM| > z_{\alpha/2} \quad (12)$$

where $Z_{\alpha/2}$ is the critical value of the standard normal distribution when the significance level is α .

In our analysis, we used the DM test to investigate significant differences in performance between the proposed hybrid system and traditional models. The results of the DM test on the basis of the square error loss function are presented in Table 11, which indicate that the DM statistical values for all models far exceed the critical value at 1% significance level. As obvious, the proposed hybrid system performs differently when compared to the traditional models at 1% significance level. Combining this with the evaluation criteria in Experiment II, the proposed hybrid system is outright better than the traditional models and potentially meets the requirements of wind speed forecasting.

6.4.2. Forecasting Effectiveness

In this section, forecasting effectiveness is introduced, which evaluates the performance of models by using the sum of the squared errors as well as the mean and mean squared deviation of the

forecasting accuracy. Furthermore, the skewness and kurtosis of the forecasting accuracy distribution need to be considered in practical circumstances. The general form of forecasting effectiveness is described as follows [79].

The k th-order forecasting effectiveness unit is described as:

$$m^k = \sum_{n=1}^N Q_n A_n^k$$

$$\sum_{n=1}^N Q_n = 1 \quad (13)$$

where Q_n denotes the discrete probability distribution at time n . As any prior information of the discrete probability distribution is unknown, Q_n is defined as $1/N$. A_n is the forecasting accuracy defined as:

$$A_n = 1 - |\varepsilon_n| \quad (14)$$

$$\varepsilon_n = \begin{cases} -1, & (y_n - \hat{y}_n)/y_n < -1 \\ (y_n - \hat{y}_n)/y_n, & -1 \leq (y_n - \hat{y}_n)/y_n < 1 \\ 1, & (y_n - \hat{y}_n)/y_n > 1 \end{cases} \quad (15)$$

The k -order forecasting effectiveness is defined as:

$$H(m^1, m^2, \dots, m^k) \quad (16)$$

When $H(x) = x$ is a continuous function in one-variable, the first-order forecasting effectiveness is the expected forecasting accuracy sequence defined as $H(m^1) = m^1$. Similarly, when $H(x, y) = x(1 - \sqrt{y - x^2})$ is a continuous function in two variables, the second-order forecasting effectiveness is the difference between the standard deviation and expectation, which can be described as

$$H(m^1, m^2) = m^1 \left(1 - \sqrt{m^2 - (m^1)^2} \right) \quad (17)$$

In this study, forecasting effectiveness was also used to evaluate the performance of different models. The model which possesses greater forecasting effectiveness is said to perform better. The first-order forecasting effectiveness is based on the expected value of the forecasting accuracy sequence, while the second-order forecasting effectiveness is related to the difference between the standard deviation and expectation of the forecasting accuracy sequence. Detailed results of the first- and second-order forecasting effectiveness are presented in Table 12. It can be easily seen that the proposed hybrid forecasting system outperforms the other models, for the value of the forecasting effectiveness of the proposed system far exceeds that corresponding to other models in all cases. Take dataset I for example, the first-order forecasting effectiveness of BPNN, ELM, Elman, SVR, ARIMA, and DES models are, respectively, 0.9209, 0.922, 0.9205, 0.9189, 0.9203, and 0.8967. At the same time, corresponding values of the proposed hybrid forecasting system with four different discretization methods are 0.9480, 0.9462, 0.9470, and 0.9469. Further, the second-order forecasting effectiveness values for the above methods and the proposed hybrid system are 0.8558, 0.8563, 0.8557, 0.8487, 0.8565, and 0.8086 and 0.9069, 0.9049, 0.8994, 0.9063, respectively.

Remark 4: The results obtained from the DM test and forecasting effectiveness indicate that the forecasting accuracy of the proposed system is remarkably higher than the BPNN, ELM, Elman, SVR, ARIMA, and DES models, and the developed hybrid forecasting system is more viable and significantly superior to the traditional forecasting models.

Table 11. DM test results of different models for the three datasets.

Datasets	Models	BPNN	ELM	Elman	SVR	ARIMA	DES
Dataset I	Hybrid system1	9.5759	6.9282	9.6694	8.6703	9.6704	9.4034
Dataset II		8.1057	5.2140	14.5774	6.4442	8.0502	8.9632
Dataset III		8.5758	6.8089	12.069	9.5842	8.5689	9.5542
Dataset I	Hybrid system2	8.2046	9.1922	8.2829	7.4465	8.3739	9.2545
Dataset II		8.5994	7.8156	14.7021	6.6676	8.5149	8.9278
Dataset III		7.4355	8.0469	11.9691	8.4399	7.3988	8.9997
Dataset I	Hybrid system3	9.2870	7.9969	9.3517	8.4294	9.3695	9.3582
Dataset II		7.9956	8.2683	14.2859	6.2895	7.8031	8.8359
Dataset III		8.3252	6.9085	12.1188	9.0679	8.2286	9.3185
Dataset I	Hybrid system4	9.5094	8.9113	9.5849	8.5284	9.6266	9.3763
Dataset II		7.4392	7.5030	13.8997	5.7378	7.1529	8.2366
Dataset III		7.7251	7.6462	11.9581	8.8024	7.6623	9.1974

Table 12. Forecasting effectiveness of different models for the three datasets.

Models		Dataset I		Dataset II		Data III	
		First-Order	Second-Order	First-Order	Second-Order	First-Order	Second-Order
Compared Models	BPNN	0.9209	0.8558	0.9429	0.8943	0.9338	0.8789
	ELM	0.922	0.8563	0.9442	0.8986	0.9362	0.8825
	Elman	0.9205	0.8557	0.8908	0.8028	0.8736	0.7760
	SVR	0.9189	0.8487	0.9416	0.8868	0.9308	0.8740
	ARIMA	0.9203	0.8565	0.9439	0.8969	0.9344	0.8803
	DES	0.8967	0.8086	0.9309	0.8741	0.9169	0.8463
Hybrid Forecasting System	Chi	0.9480	0.9069	0.9596	0.9284	0.9534	0.9151
	Entropy	0.9462	0.9049	0.9594	0.9267	0.9531	0.9143
	EF	0.9470	0.8994	0.9607	0.9272	0.9496	0.9058
	EW	0.9469	0.9063	0.9563	0.9219	0.9523	0.9151

7. Sensitivity Analysis of Parameters in the Proposed Hybrid Forecasting System

The proposed hybrid forecasting system involves two parameters—ensemble number and noise amplitude—that need to be predefined [80]. To investigate the sensitivity of these parameters, Dataset I was processed using the proposed hybrid forecasting system with the Chi-square-based discretization algorithm.

7.1. Setting the Ensemble Number for Ensemble Empirical Mode Decomposition

In this case, the noise amplitude is maintained constant, and the number of ensembles is varied. However, there is no unified standard for the size of these parameters. By referring to several experiments and literature [4,81,82], we set the amplitude of white noise as 0.2 and the ensemble number as 50, 100, and 200. Table 13 compares the forecasting results obtained with the use of different ensemble numbers. The results indicate that when the ensemble number is 100, the system demonstrates the best forecasting performance. The forecasting accuracy decreases as we go above or below this value. As an illustration of this fact, MAPE values corresponding to ensemble numbers of 50, 100, and 200 were found to be 5.7744%, 5.1993%, and 5.7811%, respectively.

7.2. Setting Amplitude of Added Noise

The influence of added white noise amplitude on the forecasting performance is explored in this section. Here, the ensemble number is kept constant, and the amplitude of added noise is varied. By referring to literature [80], we set the amplitudes of the added white noise as 0.1, 0.2, and 0.5, while ensemble number was maintained at 100. Table 13 represents the forecasting results obtained using proposed system with different values of the added noise amplitude. In terms of the criteria

mentioned in Section 6, best forecasting results are achieved when the amplitude of added noise is maintained as 0.2. The results in Table 13 indicate that a change in amplitude of the added noise influences the forecasting accuracy. If too small amplitude is selected for the added noise, a series of smooth and stable data may not be introduced. On the other hand, if we select too large a noise amplitude, some frequency information could be lost in the noise, and the forecasting accuracy will decrease.

Table 13. Results of sensitivity analysis of parameters in the proposed hybrid forecasting system.

The Value of the Ensemble Number Is 200		MAE	RMSE	MAPE (%)	IA	VAR
The amplitude of added noise	0.1	0.356060	0.469598	6.013668	0.961840	0.220853
	0.2	0.304745	0.393810	5.199317	0.973831	0.155138
	0.5	0.335544	0.432928	5.720263	0.967039	0.187473
White noise is 0.5		MAE	RMSE	MAPE (%)	IA	VAR
The value of ensemble number	50	0.340148	0.438051	5.774439	0.966492	0.192129
	100	0.304745	0.393810	5.199317	0.973831	0.155138
	200	0.342039	0.441753	5.781073	0.966076	0.195446

8. Further Experiments for Hourly Time Horizon

In order to support the merits of the proposed hybrid system in comparison to other forecasting models, we performed a further experiment comprising the hourly time-horizon wind speed forecasting. The results of this experiment, in terms of evaluation criteria, are presented in Table 14, and the results of the DM test and forecasting effectiveness are listed in Tables 15 and 16, respectively. It is easily recognized that MAPE of the proposed system is about 7%, while for the compared models, this value varies in the range of 15–20%. Corresponding VAR values are about 0.3 and above 1, respectively, indicating that forecasting results of the proposed system have better accuracy and stability. The performance of artificial neural networks is only slightly different from each other, while DES is evidently poor compared to ARIMA amongst statistical models.

The DM statistical values of all models are about 5, which is higher than the critical value at the 1% significance level. We can, thus, conclude that the proposed hybrid system is obviously different and performs better compared to other models at the 1% significance level. Combining this with the results based on evaluation criteria, the proposed hybrid system can be seen to outperform traditional models.

It can be inferred from Table 16 that the forecasting effectiveness of the proposed system exceeds that of the compared models under all cases. The first-order forecasting effectiveness offered by BPNN, ELM, Elman, SVR, ARIMA, and DES is about 0.85, while that corresponding to the proposed hybrid forecasting system with four different interval partitioning methods is about 0.93. The respective second-order values are about 0.88 and 0.75. Amongst the models being compared, DES has the worst performance with respective first- and second-order forecasting effectiveness values of 0.799 and 0.6614.

Remark 5: As for the hourly time-horizon wind speed forecasting, the evaluation criteria and testing results which are obtained by DM test the forecasting effectiveness all show that the level of forecasting accuracy of the proposed system is remarkably higher than the compared model. But, the forecasting performance for the 10 min-horizon wind speed are overall superior to the hourly time-horizon wind speed for the same model. Based on the above analysis, we can conclude that the proposed system has general applicability and great performance.

Table 14. Comparison of different models for the hourly time horizon wind speed forecasting.

MODELS		MAE	RMSE	MAPE	IA	VAR
Hybrid forecasting system	Chi ²	0.390194	0.055946	6.411913	0.953606	0.260964
	Entropy	0.416678	0.058084	6.928242	0.950354	0.264313
	EF	0.437427	0.061288	7.264912	0.949147	0.312299
	EW	0.432544	0.061552	7.003766	0.943157	0.315089
Artificial Neural Network	BPNN	0.825827	1.068416	14.24583	0.718107	1.154761
	ELM	0.859881	1.100304	14.36884	0.714523	1.221836
	Elman	0.843463	1.083612	14.61775	0.72943	1.177776
Statistical	ARIMA	0.791711	1.024048	13.37103	0.727751	1.061284
	DES	1.205944	1.563405	20.08901	0.68189	2.472686
SVR		0.948013	1.225591	16.6055	0.646858	1.499271

Table 15. DM test results of different models for hourly time horizon wind speed forecasting.

DM Test		BP	ELM	Elman	SVR	ARIMA	DES
Hybrid system 1	Chi ²	4.877626	4.950937	4.90825	4.686209	4.909181	5.366497
Hybrid system 2	Entropy	4.737063	4.798341	4.770708	4.573673	4.735968	5.327807
Hybrid system 3	EF	4.475986	4.527564	4.495222	4.397966	4.445831	5.258996
Hybrid system 4	EW	4.571886	4.634924	4.605242	4.467959	4.522839	5.257129

Table 16. Forecasting effectiveness of different forecasting models for hourly time horizon wind speed forecasting.

Forecasting Effectiveness	Chi ²	Entropy	EF	EW	BPNN
first-order	0.93588	0.930718	0.927351	0.929962	0.855717
second-order	0.88826	0.88145	0.870173	0.878382	0.747753
<i>Forecasting effectiveness</i>	ELM	Elman	SVR	ARIMA	DES
first-order	0.857697	0.853565	0.833945	0.862438	0.79911
second-order	0.749865	0.745601	0.713875	0.758843	0.661372

9. Conclusions

Data pre-processing and future forecasting are crucial tasks in modern national and regional economic development, especially in the energy sector. Poor energy forecasting may lead to wastage of the already scarce energy sources. As such, both accuracy and stability are important objectives to be achieved in energy forecasting. Nevertheless, accurate energy forecasting is considered to be a challenging task because of various influencing factors, such as noise and high data volatility. Conventional statistical models require a large amount of historical data and face restrictions, such as linear or normality postulates. On the other hand, use of artificial neural networks involves several parameters and requires substantial response time. To overcome the limitations and challenges in these methods, we proposed the hybrid forecasting system with four different interval partitioning methods.

By comparing the forecasting accuracy, stability, and effectiveness of the proposed system against conventional statistical models and artificial neural networks via the data from three sites, it is concluded that the proposed system significantly outperforms the other models. Especially, the variance criterion (VAR) for the DES model is significantly larger compared to that for the proposed hybrid forecasting system thereby reducing the stability and reliability of DES forecasting results. Also, because the proposed system involves simple calculations and results do not change with time for the same sample, the forecasting efficiency and stability is evidently improved.

The volatility and instability of raw data increase the difficulties involved in wind speed forecasting; thus, the pre-processing the data prior to forecasting is essential. Experiments performed in this study indicate that the ‘decomposition and ensemble’ strategy for raw data remarkably improves

the forecasting performance. The comparison of forecasting results obtained using four different interval partitioning methods indicate that although forecasting accuracy does vary significantly between them, the supervised discretization methods are superior to unsupervised methods.

Additionally, sensitivity analysis of parameters used in the proposed forecasting system indicates that by appropriately setting the ensemble number and white noise amplitude, the forecasting accuracy can be greatly improved. In order to prove the superiority of the proposed hybrid system over other forecasting models, the hourly time-horizon wind speed was further simulated. Results of this simulation indicate that the proposed system has better performance compared to all other models for different time-horizon datasets. Further, forecasting performance of the proposed system for the 10 min-horizon wind speed is superior to the forecasting performance for the hourly time-horizon wind speed. In conclusion, the proposed hybrid forecasting system demonstrates better forecasting accuracy, effectiveness, and stability while handling noisy and insufficient datasets in the wind energy system.

Acknowledgments: This research was supported by the National Natural Science Foundation of China (Grant No. 71573034).

Author Contributions: Hufang Yang proposed the concept of this research and provided overall guidance; Zaiping Jiang wrote the whole manuscript. Hufang Yang carried out the data analysis; Haiyan Lu polished the manuscript and supported in part the data processing.

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

Abbreviations

Abbreviations in this manuscript are summed up as follows:

AIC	A-Information Criterion
ARIMA	Autoregressive Integrated Moving Average
BPNN	Back Propagation Neural Network
Chi ²	Chi-square
CRO	Coral Reefs Optimization
DES	Double Exponential Smoothing
DM	Diebold–Mariano
EF	equal frequency
ELM	Extreme Learning Machine
EW	equal width
FE	forecasting effectiveness
FLR	Fuzzy Logical Relationship
FSP	Feature Selection Problem
FTS	Fuzzy time series algorithm
HS	Harmony Search
IA	Index of agreement of forecasting results
LHS	Left-hand side
MAE	Mean absolute error
MAPE	Mean Absolute Percentage Error
R	Correlation coefficient
RBFNN	Radial Basis Function Neural Network
RHS	Right-hand side
RMSE	Root Mean Square Error
SAM	Seasonal Adjustment Method
SVR	Support Vector Regression
VAR	Variance of the error

References

1. Baños, R.; Manzano-Agugliaro, F.; Montoya, F.G.; Gil, C.; Alcayde, A.; Gómez, J. Optimization methods applied to renewable and sustainable energy: A review. *Renew. Sustain. Energy Rev.* **2011**, *15*, 1753–1766. [[CrossRef](#)]
2. Harmsen, J.H.M.; Roes, A.L.; Patel, M.K. The impact of copper scarcity on the efficiency of 2050 global renewable energy scenarios. *Energy* **2013**, *50*, 62–73. [[CrossRef](#)]
3. Yesilbudak, M.; Sagioglu, S.; Colak, I. A new approach to very short term wind speed prediction using k-nearest neighbor classification. *Energy Convers. Manag.* **2013**, *69*, 77–86. [[CrossRef](#)]
4. Wang, J.; Jiang, H.; Zhou, Q.; Wu, J.; Qin, S. China's natural gas production and consumption analysis based on the multicycle Hubbert model and rolling Grey model. *Renew. Sustain. Energy Rev.* **2016**, *53*, 1149–1167. [[CrossRef](#)]
5. Hernández-Escobedo, Q.; Saldaña-Flores, R.; Rodríguez-García, E.R.; Manzano-Agugliaro, F. Wind energy resource in Northern Mexico. *Renew. Sustain. Energy Rev.* **2014**, *32*, 890–914. [[CrossRef](#)]
6. Oh, K.Y.; Kim, J.Y.; Lee, J.K.; Ryu, M.S.; Lee, J.S. An assessment of wind energy potential at the demonstration offshore wind farm in Korea. *Energy* **2012**, *46*, 555–563. [[CrossRef](#)]
7. Montoya, F.G.; Manzano-Agugliaro, F. Wind turbine selection for wind farm layout using multi-objective evolutionary algorithms. *Expert Syst. Appl.* **2014**, *41*, 6585–6595. [[CrossRef](#)]
8. Manzano-Agugliaro, F.; Alcayde, A.; Montoya, F.G.; Zapata-Sierra, A.; Gil, C. Scientific production of renewable energies worldwide: An overview. *Renew. Sustain. Energy Rev.* **2013**, *18*, 134–143. [[CrossRef](#)]
9. World Wind Energy Association. Available online: <http://www.wwindea.org/11961-2/> (accessed on 24 July 2017).
10. Ma, X.; Jin, Y.; Dong, Q. A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. *Appl. Soft Comput. J.* **2017**, *54*, 296–312. [[CrossRef](#)]
11. State Grid. *Nb/T 31046 Function Specification of Wind Power Prediction*; China Electric Power Press: Beijing, China, 2013.
12. Chunyan, Y. *Research on Wind Speed and Wind Power Forecasting Related Issue*; Huazhong University of Science and Technology: Wuhan, China, 2013.
13. Hernandez-escobedo, Q.; Manzano-agugliaro, F.; Gazquez-parra, J.A.; Zapata-sierra, A. Is the wind a periodical phenomenon? The case of Mexico. *Renew. Sustain. Energy Rev.* **2011**, *15*, 721–728. [[CrossRef](#)]
14. Ackermann, T.; Söder, L. Wind energy technology and current status: A review. *Renew. Sustain. Energy Rev.* **2011**, *4*, 315–374. [[CrossRef](#)]
15. Chang, P.C.; Yang, R.Y.; Lai, C.M. Potential of Offshore Wind Energy and Extreme Wind Speed Forecasting on the West Coast of Taiwan. *Energies* **2015**, *8*, 1685–1700. [[CrossRef](#)]
16. Safat, A. A Physical Approach to Wind Speed Prediction for Wind Energy Forecasting. Available online: <http://www.iawe.org/Proceedings/CWE2006/MC4-01.pdf> (accessed on 24 July 2017).
17. Yamaguchi, A.; Enoki, K.; Ishihara, T.; Fukumoto, Y.; Okino, M.; Iba, S.; Ohya, Y.; Karasudani, T.; Watanabe, K.; Noda, M.; et al. Wind Power Forecasting with Physical Model and Multi Time Scale Model. *J. Wind Eng.* **2010**, *2007*, 251–264. [[CrossRef](#)]
18. Filik, T. Improved Spatio-Temporal Linear Models for Very Short-Term Wind Speed Forecasting. *Energies* **2016**, *9*, 168. [[CrossRef](#)]
19. Lei, M.; Luan, S.; Jiang, C.; Liu, H.; Yan, Z. A review on the forecasting of wind speed and generated power. *Renew. Sustain. Energy Rev.* **2009**, *13*, 915–920. [[CrossRef](#)]
20. Shukur, O.B.; Lee, M.H. Daily wind speed forecasting through hybrid AR-ANN and AR-KF models. *J. Teknol.* **2015**, *72*, 89–95. [[CrossRef](#)]
21. Zhang, C.L. The Wind Speed Prediction Based on AR Model and BP Neural Network. *Adv. Mater. Res.* **2012**, *450–451*, 1593–1596.
22. Torres, J.L.; García, A.; De Blas, M.; De Francisco, A. Forecast of hourly average wind speed with ARMA models in Navarre (Spain). *Sol. Energy* **2005**, *79*, 65–77. [[CrossRef](#)]
23. Cadenas, E.; Rivera, W.; Campos-Amezcuca, R.; Heard, C. Wind Speed Prediction Using a Univariate ARIMA Model and a Multivariate NARX Model. *Energies* **2016**, *9*, 109. [[CrossRef](#)]

24. Wang, G.Q.; Wang, S.; Liu, H.Y.; Xue, Y.D.; Ping, Z.; Amp, E. Self-adaptive and dynamic cubic ES method for wind speed forecasting. *Power Syst. Prot. Control* **2014**, *42*, 117–122.
25. Booth, D.E. Time Series (3rd ed.). *J. Technometrics* **1992**, *34*, 118–119.
26. Li, G.; Shi, J. On comparing three artificial neural networks for wind speed forecasting. *Appl. Energy* **2010**, *87*, 2313–2320. [[CrossRef](#)]
27. Hyperbolic tangent basis function neural networks training by hybrid evolutionary programming for accurate short-term wind speed prediction. In Proceedings of the Ninth Intelligent Systems Design and Applications Conference (ISDA'09), Pisa, Italy, 30 November–2 December 2009.
28. Salcedo-sanz, S.; Pastor-sánchez, A.; Prieto, L.; Blanco-aguilera, A.; García-herrera, R. Feature selection in wind speed prediction systems based on a hybrid coral reefs optimization—Extreme learning machine approach. *Energy Convers. Manag.* **2014**, *87*, 10–18. [[CrossRef](#)]
29. Salcedo-Sanz, S.; Pastor-Sánchez, A.; Ser J, D.e.l.; Prieto, L.; Geem, Z.W. A Coral Reefs Optimization algorithm with Harmony Search operators for accurate wind speed prediction. *Renew. Energy* **2015**, *75*, 93–101. [[CrossRef](#)]
30. Zhang, Q.; Lai, K.K.; Niu, D.; Wang, Q.; Zhang, X. A Fuzzy Group Forecasting Model Based on Least Squares Support Vector Machine (LS-SVM) for Short-Term Wind Power. *Energies* **2012**, *5*, 3329–3346. [[CrossRef](#)]
31. Salcedo-Sanz, S.; Ortiz-García, E.G.; Pérez-Bellido, A.M.; Portilla-Figueras, E.; Prieto, L.; Paredes, D.; Correoso, F. Performance comparison of Multilayer Perceptrons and Support vector Machines in a Short-term Wind speed Prediction Problem. *Neural Netw. World* **2009**, *19*, 37–51.
32. Ortiz-García, E.G.; Salcedo-Sanz, S.; Pérez-Bellido, Á.M.; Gascón-Moreno, J.; Portilla-Figueras, J.A.; Prieto, L. Short-term wind speed prediction in wind farms based on banks of support vector machines. *Wind Energy* **2011**, *14*, 193–207. [[CrossRef](#)]
33. Salcedo-Sanz, S.; Ortiz-Garci, E.G.; Pérez-Bellido, Á.M.; Portilla-Figueras, A.; Prieto, L. Short term wind speed prediction based on evolutionary support vector regression algorithms. *Expert Syst. Appl.* **2011**, *38*, 4052–4057. [[CrossRef](#)]
34. Jiang, Y.; Song, Z.; Kusiak, A. Very short-term wind speed forecasting with Bayesian structural break model. *Renew. Energy* **2013**, *50*, 637–647. [[CrossRef](#)]
35. Troncoso, A.; Salcedo-sanz, S.; Casanova-mateo, C.; Riquelme, J.C.; Prieto, L. Local models-based regression trees for very short-term wind speed prediction. *Renew. Energy* **2015**, *81*, 589–598. [[CrossRef](#)]
36. Pourmousavi Kani, S.A.; Ardehali, M.M. Very short-term wind speed prediction: A new artificial neural network-Markov chain model. *Energy Convers. Manag.* **2011**, *52*, 738–745. [[CrossRef](#)]
37. Khashei, M.; Bijari, M.; Ardali, G.A.R. Improvement of Auto-Regressive Integrated Moving Average models using Fuzzy logic and Artificial Neural Networks (ANNs). *Neurocomputing* **2009**, *72*, 956–967. [[CrossRef](#)]
38. Salcedo-Sanz, S.; Prieto, L.; Prieto, L.; Correoso, F. Letters: Accurate short-term wind speed prediction by exploiting diversity in input data using banks of artificial neural networks. *Neurocomputing* **2009**, *72*, 1336–1341. [[CrossRef](#)]
39. Chang, W.Y. Short-Term Wind Power Forecasting Using the Enhanced Particle Swarm Optimization Based Hybrid Method. *Energies* **2013**, *6*, 4879–4896. [[CrossRef](#)]
40. Salcedo-Sanz, S.; Pérez-Bellido, Á.M.; Ortiz-García, E.G.; Portilla-Figueras, A.; Prieto, L.; Paredes, D. Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction. *Renew. Energy* **2009**, *34*, 1451–1457. [[CrossRef](#)]
41. Sanz, S.S.; Prieto, L.; Paredes, D.; Correoso, F. Short-term wind speed prediction by hybridizing global and mesoscale forecasting models with artificial neural networks. In Proceedings of the Eighth International Conference on Hybrid Intelligent Systems (HIS'08), Barcelona, Spain, 10–12 September 2008.
42. Hervás-Martínez, C.; Salcedo-Sanz, S.; Gutiérrez, P.A.; Ortiz-García, E.G.; Prieto, L. Evolutionary product unit neural networks for short-term wind speed forecasting in wind farms. *Neural Comput. Appl.* **2012**, *21*, 993–1005. [[CrossRef](#)]
43. Zhang, W.; Wang, J.; Wang, J.; Zhao, Z.; Tian, M. Short-term wind speed forecasting based on a hybrid model. *Appl. Soft Comput. J.* **2013**, *13*, 3225–3233. [[CrossRef](#)]
44. Hong, Y.Y.; Yu, T.H.; Liu, C.Y. Hour-Ahead Wind Speed and Power Forecasting Using Empirical Mode Decomposition. *Energies* **2013**, *6*, 6137–6152. [[CrossRef](#)]
45. Liu, H.; Chen, C.; Tian, H.Q.; Li, Y.F. A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks. *Renew. Energy* **2012**, *48*, 545–556. [[CrossRef](#)]

46. Liu, H.; Tian, H.Q.; Liang, X.F.; Li, Y.F. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. *Appl. Energy* **2015**, *157*, 183–194. [[CrossRef](#)]
47. Liu, H.; Tian, H.; Liang, X.; Li, Y. New wind speed forecasting approaches using fast ensemble empirical model decomposition, genetic algorithm, Mind Evolutionary Algorithm and Artificial Neural Networks. *Renew. Energy* **2015**, *83*, 1066–1075. [[CrossRef](#)]
48. Tascikaraoglu, A.; Uzunoglu, M. A review of combined approaches for prediction of short-term wind speed and power. *Renew. Sustain. Energy Rev.* **2014**, *34*, 243–254. [[CrossRef](#)]
49. Jilani, T.A.; Burney, S.M.A. M-Factor High Order Fuzzy Time Series Forecasting for Road Accident Data. *Adv. Soft Comput.* **2007**, *41*, 246–254.
50. Jiang, P.; Wang, Y.; Wang, J. Short-term wind speed forecasting using a hybrid model. *Energy* **2016**, *119*, 561–577. [[CrossRef](#)]
51. Masrur, H.; Nimol, M. Short Term Wind Speed Forecasting Using Artificial Neural Network: A Case Study. In Proceedings of the International Conference on Innovations in Science, Engineering and Technology (ICISSET), Dhaka, Bangladesh, 28–29 October 2016.
52. Niazy, R.K.; Beckmann, C.F.; Brady, J.M.; Smith, S.M. Performance Evaluation of Ensemble Empirical Mode Decomposition. *Adv. Adapt. Data Anal.* **2009**, *1*, 231–242. [[CrossRef](#)]
53. Zhu, B. A Novel Multiscale Ensemble Carbon Price Prediction Model Integrating Empirical Mode Decomposition, Genetic Algorithm and Artificial Neural Network. *Energies* **2012**, *5*, 163–170. [[CrossRef](#)]
54. Zhaohua, W.U.; Huang, N.E. Ensemble Empirical Mode Decomposition: A Noise-Assisted Data Analysis Method. *Adv. Adapt. Data Anal.* **2011**, *1*, 1–41.
55. Yu, L.; Wang, Z.; Tang, L. A decomposition—Ensemble model with data-characteristic-driven reconstruction for crude oil price forecasting. *Appl. Energy* **2015**, *156*, 251–267. [[CrossRef](#)]
56. Chen, Y.S.; Cheng, C.H.; Tsai, W.L. Modeling fitting-function-based fuzzy time series patterns for evolving stock index forecasting. *Appl. Intell.* **2014**, *41*, 327–347. [[CrossRef](#)]
57. Wang, J.; Xiong, S. A hybrid forecasting model based on outlier detection and fuzzy time series—A case study on Hainan wind farm of China. *Energy* **2014**, *76*, 526–541. [[CrossRef](#)]
58. Li, S.T.; Cheng, Y.C. Deterministic fuzzy time series model for forecasting enrollments. *Comput. Math. Appl.* **2007**, *53*, 1904–1920. [[CrossRef](#)]
59. Lee, Y.C.; Wu, C.H.; Tsai, S.B. Grey system theory and fuzzy time series forecasting for the growth of green electronic materials. *Int. J. Prod. Res.* **2014**, *52*, 2931–2945. [[CrossRef](#)]
60. Sadaei, H.J.; Guimarães, F.G.; Da Silva, C.J.; Lee, M.H.; Eslami, T. Short-term load forecasting method based on fuzzy time series, seasonality and long memory process. *Int. J. Approx. Reason.* **2017**, *83*, 196–217. [[CrossRef](#)]
61. Song, Q.; Chissom, B.S. *Fuzzy Time Series and Its Models*; Elsevier North-Holland, Inc.: Amsterdam, The Netherlands, 1993.
62. Yu, H.K. Weighted fuzzy time series models for TAIEX forecasting. *Phys. A Stat. Mech. Appl.* **2012**, *349*, 609–624. [[CrossRef](#)]
63. Abdullah, L.; Taib, I. High order fuzzy time series for exchange rates forecasting. In Proceedings of the 2011 3rd Conference on Data Mining and Optimization (DMO), Putrajaya, Malaysia, 28–29 June 2011; pp. 1–5.
64. Chen, M.; Chen, B. A hybrid fuzzy time series model based on granular computing for stock price forecasting. *Inf. Sci.* **2015**, *294*, 227–241. [[CrossRef](#)]
65. Lu, W.; Chen, X.; Pedrycz, W.; Liu, X.; Yang, J. Using interval information granules to improve forecasting in fuzzy time series. *Int. J. Approx. Reason.* **2015**, *57*, 1–18. [[CrossRef](#)]
66. Dash, R.; Paramguru, R.L.; Dash, R. Comparative Analysis of Supervised and Unsupervised Discretization Techniques. *Int. J. Adv. Sci. Technol.* **2011**, *2*, 29–37.
67. Duda, J. Supervised and Unsupervised Discretization of Continuous Features. In Proceedings of the Twelfth International Conference on Machine Learning, Tahoe, CA, USA, 9–12 July 1995; Volume 12, pp. 194–202.
68. Peng, L.; Wang, Q.; Yujia, G. Study on Comparison of Discretization Methods. In Proceedings of the International Conference on Artificial Intelligence and Computational Intelligence, 2009 (AICI'09), Shanghai, China, 7–8 November 2009; pp. 380–384.
69. Hua, H.; Zhao, H. *A Discretization Algorithm of Continuous Attributes Based on Supervised Clustering*; Photoelectric Information Technology Research Room: Liaoning, China, 2009; pp. 1–5.

70. Joița, D. *Unsupervised Static Discretization Methods in Data Mining*; Titu Maiorescu University: Bucharest, Romania, 2010.
71. Schmidberger, G.; Frank, E. *Unsupervised Discretization Using Tree-Based Density Estimation*; Springer: Berlin/Heidelberg, Germany, 2005.
72. Wu, C.H.; Kao, S.C.; Okuhara, K. Examination and comparison of conflicting data in granulated datasets: Equal width interval vs. equal frequency interval. *Inf. Sci.* **2013**, *239*, 154–164. [[CrossRef](#)]
73. Fayyad, U.; Irani, K. Multi-Interval Discretization of Continuous-Valued Attributes for Classification Learning. In Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, Chambéry, France, 28 August–3 September 1993; pp. 1022–1027.
74. Soares, C.; Knobbe, A. Entropy-based discretization methods for ranking data. *Inf. Sci.* **2016**, *329*, 921–936.
75. Boulle, M. Khiops: A Statistical Discretization Method of Continuous Attributes. *Mach. Learn.* **2004**, *55*, 53–69. [[CrossRef](#)]
76. Renani, E.T.; 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](#)]
77. Du, P.; Wang, J.; Guo, Z.; Yang, W. Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting. *Energy Convers. Manag.* **2017**, *150*, 90–107. [[CrossRef](#)]
78. Xu, Y.; Yang, W.; Wang, J. Air quality early-warning system for cities in China. *Atmos. Environ.* **2017**, *148*, 239–257. [[CrossRef](#)]
79. Xiao, L.; Shao, W.; Wang, C.; Zhang, K.; Lu, H. Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting. *Appl. Energy* **2016**, *180*, 213–233. [[CrossRef](#)]
80. Wang, S.; Zhang, N.; Wu, L.; Wang, Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* **2016**, *94*, 629–636. [[CrossRef](#)]
81. Wang, Y.H.; Yeh, C.H.; Young, H.W.V.; Hu, K.; Lo, M.T. On the computational complexity of the empirical mode decomposition algorithm. *Phys. A Stat. Mech. Appl.* **2014**, *400*, 159–167. [[CrossRef](#)]
82. Ma, X.; Liu, D. Comparative Study of Hybrid Models Based on a Series of Optimization Algorithms and Their Application in Energy System Forecasting. *Energies* **2016**, *9*, 640. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).