Renewable Energy 60 (2013) 185–194

Contents lists available at SciVerse ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

A hybrid forecasting approach applied to wind speed time series

Jianming Hu*, Jianzhou Wang, Guowei Zeng

School of Mathematics & Statistics, Lanzhou University, Lanzhou 730000, China

ARTICLE INFO

Article history:
Received 23 January 2013
Accepted 6 May 2013
Available online

Keywords:
Wind farm
Wind speed forecasting
Ensemble Empirical Mode Decomposition (EEMD)
Support Vector Machine (SVM)

ABSTRACT

In this paper, a hybrid forecasting approach, which combines the Ensemble Empirical Mode Decomposition (EEMD) and the Support Vector Machine (SVM), is proposed to improve the quality of wind speed forecasting. The essence of the methodology incorporates three phases. First, the original data of wind speed are decomposed into a number of independent Intrinsic Mode Functions (IMFs) and one residual series by EEMD using the principle of decomposition. In order to forecast these IMFs, excepting the highest frequency acquired by EEMD, the respective estimates are yielded using the SVM algorithm. Finally, these respective estimates are combined into the final wind speed forecasts using the principle of ensemble. The proposed hybrid method is examined by forecasting the mean monthly wind speed of three wind farms located in northwest China. The obtained results confirm an observable improvement for the forecasting validity of the proposed hybrid approach. This tool shows great promise for the forecasting of intricate time series which are intrinsically highly volatile and irregular.

1. Introduction

With the rapid development of society, economy and industry, the demand for electricity, as a fundamental and indispensable power, is growing dramatically. However, natural resources, especially fossil fuels, are being extensively exploited. This encourages the exploration of renewable, clean and free energy sources [1]. Wind power is a required energy source. Wind energy projects have environmental, economical, and social benefits, and have been successfully introduced in many countries. To a large extent, wind energy is one of the predominant alternative sources of energy, e.g. wind energy represents about 10% of energy consumption in Europe, even over 15% in countries such as Spain and Germany [2]. Similarly, in China, the burgeoning wind energy industry is substantially supported by government. The installed wind power capacity was 44733.29 MW in 2010, with an average annual growth rate of 73.3%, ranking first in the world [3]. In 2011, the installed wind power capacity was 62364.2 MW, with an average annual growth rate of 39.4% [4].

It is well-known that one of the most significant factors of wind power generation is wind speed. Unfortunately, wind speed, viewed as one of most difficult meteorological parameters, is not an easy factor to forecast. A series of meteorological factors such as pressure and temperature differences, the rotation of the earth and local characteristics of the surface influence wind speed; additionally, the complex interactions between these meteorological factors made wind speed even more difficult to forecast [5]. However, in a liberalized electricity market, the excellent performance of wind speed forecasting will help boost competitiveness of wind energy compared to other forms of energy. Considering the operation of electric utilities which integrate wind energy, it is becoming increasingly important and pertinent to forecast wind speed. For example, a 10% deviation of the expected wind speed results approximately in a 30% deviation in the expected wind power generation [6].

In recent years, wind speed forecasting has attracted many researchers to do an array of studies in this field, which can be divided into short-term forecasting and long-term forecasting according to the time scales of wind speed data and the purpose of the forecast. While accurate short-term forecasts of wind speed minimize scheduling errors which has made a great impact on grid reliability and market based ancillary service costs [7], long-term forecasts provide important references for site location, planning of windmills and the selection of an optimal size of the wind machine for a particular site [8]. According to the data used by models, these models can be classified by two catalogs. One catalog belongs to meteorological methods. These models, developed by meteorologists for large scale area weather prediction, using physical data such as temperature, pressure and topographical information, are used to forecast the future wind speed [9–11]. These meteorological models do not give accurate results. Other kinds of models are employed to forecast wind speed using past wind data which may

* Corresponding author. Tel.: +86 18993173569; fax: +86 931 8912481.
E-mail address: hujm11@lzu.edu.cn (J. Hu).

0960-1481/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.renene.2013.05.012

be analyzed with different statistical methods, including time series models such as the Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA) models [12,13] and artificial intelligence techniques including Support Vector Machines (SVM) [8] and Artificial Neural Networks (ANNs). ANNs are some of the most widely used models in the last decade for wind speed forecasting and other disciplines where time series are used. They include but are not limited to training algorithms including the Back-Propagation (BP) algorithm and the Levenberg Marquardt (LM) algorithm [4], the control algorithm based on the ANN model using the back-propagation method [14], a new strategy based on Fuzzy Logic and Artificial Neural Networks [15] and various other forms of neural networks such as Multi-layer Feed-forward Neural Networks [1,14,16–19] and Recurrent Neural Networks [20–22].

Relatively, these models using past wind data can provide moderately good results for wind speed. However, most of them are subject to shortcomings, mainly in two aspects. First, traditional time series techniques can obtain good prediction achievements under the assumption of conditions that time series present linearity and stationarity. Unfortunately, in reality, wind speed series can rarely adhere to this due to their intrinsic complexity and volatility. Secondly, despite showing superiority over traditional statistical techniques, artificial intelligence models possess their own defects and drawbacks, e.g. ANN models sometimes fall into dilemma of local minima as well as over-fitting, and also are sensitive to parameter selection. In view of the limitations of traditional techniques and artificial intelligence techniques, a novel approach is required in order to remedy the shortcomings of those models. Considering the distinction of the mean monthly wind speed data—non-constant mean and variance, high volatility and irregularity—the thought of taking a deep insight into the original data and taking advantage of artificial intelligence techniques should be adopted besides improving the models themselves.

Removing noisy data, an important part of data cleaning is significant and meaningful prior to the operation of forecasting wind speed. The methods such as Wavelet Decomposition [23] and Empirical Mode Decomposition (EMD) [19] can be applied to eliminate noisy data. However, the wavelet de-noising technology is sensitive to the selection of threshold and empirical mode decomposition suffers from an intrinsic drawback—the frequent appearance of mode mixing. Fortunately, there exists an improved method named Ensemble Empirical Mode Decomposition (EEMD) which makes up for the deficiency of EMD. The EEMD is different from other traditional decomposition methodologies such as the Fourier decomposition and wavelet decomposition. It is an empirical, intuitive, direct and self-adaptive data processing method created especially for non-linear and non-stationary signal sequences.

In an attempt to more precisely appraise the wind energy reserves in Zhangye, Jiuquan and Mazong Mountain, all located in northwest China, a hybrid approach is proposed in this paper for such tough forecasting tasks. This approach is developed through combining the EEMD, Partial Autocorrelation Function (PACF) and Support Vector Machine (SVM). The EEMD is firstly employed to disassemble the original wind speed time series into a number of independent Intrinsic Mode Functions (IMFs), thereby better understanding wind speed data structure. The PACF is used to determine the correlation between the data embedded IMF in the same frequency band and identify the lag orders, making preparation for prediction. The SVM algorithm which is a useful methodology and also a new kind of intelligent machine is applied to forecast future numerical values using the data of IMFs available in different frequencies. The simulation process and results show that this hybrid forecasting model is simple and quite efficient to forecast wind speed with the feature of high volatility and irregularity.

The rest of the paper is organized as follows. Section 2 describes acquired tools and the proposed hybrid approach in detail. Forecasting results and the effectiveness of the proposed methodology are discussed in Section 3. Finally, Section 4 concludes the paper.

2. Methodology

Owing to the inherent characteristics of nonlinearity, non-stationarity, high fluctuation and irregularity of wind speed series and the intermittent and stochastic nature of the wind source, a strategy which contains the ideology of decomposition, de-noise, single forecast and ensemble is introduced. According to the strategy, the forecasting wind speed process can be divided into three primary stages that are decomposition, de-noise plus single forecast and ensemble, respectively. First, with the help of some popular decomposition methods, the original wind speed series can be divided into a number of components depicting relatively simple but meaningful local time scales. Then, a useful prediction method is employed to forecast all components respectively. Finally, these estimates are aggregated into the final estimate as the forecast result. This type of strategy can provide both thorough components of the original data and an excellent forecasting result.

In terms of this strategy, a hybrid method integrating EEMD, PACF and the SVM is proposed to enhance the quality of wind speed forecast of three wind farms located in northwest China. This paper will demonstrate the effectiveness of the hybrid model for forecasting wind speed. Before starting to use the hybrid method, it is necessary to describe the theory of the acquired tools in the proposed approach. First, the decomposition techniques of EMD and EEMD and the theory of PACF are briefly introduced, and the principle of SVM algorithm is presented. Then the EEMD, PACF and SVM are combined into the developed approach.

2.1. Empirical Mode Decomposition (EMD)

The EMD has been widely accepted as a method of dealing with non-linear and non-stationary data. The basic idea of EMD is to identify the intrinsic oscillatory modes and to decompose original time series data into a finite and small number of oscillatory modes based on the local characteristic time scale by itself [24]. The decomposition is based on the following assumptions: (1) the signal has at least two extrema – one maximum and one minimum; (2) the characteristic time scale is defined by the time lapse between the extrema; and (3) if the data are totally devoid of extrema but contain only inflection points, then they can be differentiated one or more times to reveal the extrema. Final results can be obtained by integration(s) of the components. In the EMD method, the Intrinsic Mode Functions (IMFs) concept is introduced as one of the key innovations; it satisfies the following two properties: (a) In the whole data series, the number of extrema and the number of zero crossing in a whole sampled data set must either equal or differ at most by one; (b) At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. With the assumptions of decomposition and the above definition for the IMF, an original data series $X(t)(t = 1, 2\cdots T)$ can be decomposed in terms of the following sifting procedure. Fig. 1 demonstrates the detailed process of the EMD algorithm.

The above sifting procedure may be repeated many times. In the procedure, the IMF series must retain enough physical sense of amplitude and frequency modulations. To serve this purpose, the threshold value mentioned in step (2) for the sifting process to stop needs to be determined; a standard deviation S_d is defined to accomplish this. The S_d is computed from the two consecutive sifting results as