Applied Energy 197 (2017) 151–162

Contents lists available at ScienceDirect

Applied Energy

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

A new dynamic integrated approach for wind speed forecasting

Shaolong Sun [a], Han Qiao [b,c], Yunjie Wei [a,d,e], Shouyang Wang [a,b,e,*]

[a] Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
[b] School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China
[c] The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA
[d] Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong
[e] Center for Forecasting Science, Chinese Academy of Sciences, Beijing 100190, China

HIGHLIGHTS

- A new dynamic integrated approach is proposed for wind speed forecasting.
- Two steps are involved: data analysis and forecast modeling.
- Optimized core vector regression is firstly employed to forecast wind speed.
- Empirical results statistically verify the performance of our new approach.

ARTICLE INFO

Article history:
Received 3 December 2016
Received in revised form 26 March 2017
Accepted 1 April 2017

Keywords:
Wind speed forecasting
Core vector machine
Phase space reconstruction
Kernel principal component analysis
Competition over resource algorithm

ABSTRACT

Wind energy is considered as one of the most promising and economical renewable energy. In order to insure maximum yield of wind energy, it is vital to evaluate wind energy potential of the wind farms. Since wind energy is proportional to the cube of wind speed, the evaluation of wind energy potential assessment comes down to the wind speed forecasting. In this paper, the wind speed is predicted by utilizing a new dynamic integrated approach. The novelties of this method mainly include: firstly, the Phase Space Reconstruction (PSR) is employed to dynamically choose the input vectors of the forecasting model; secondly, the data preprocessing approach, named the Kernel Principal Component Analysis (KPCA), is proposed to efficiently extract the nonlinear characteristics of the high-dimensional feature space reconstructed by the PSR; thirdly, Core Vector Regression (CVR) model, whose parameters are determined by the Competition Over Resource (COR) heuristic algorithm, is adopted to the model for quick computational speed; finally, the Grey Relational Analysis, Diebold-Mariano and Pesaran-Timmermann statistic are treated as evaluation tools to assess the forecasting effectiveness of this approach. The empirical results show that this integrated approach can significantly improve forecasting effectiveness and statistically outperform some other benchmark methods in terms of the directional forecasting and level forecasting.

1. Introduction

Energy is the material basis of human survival and economic development. However, with the continuous and rapid development of the world economy, energy shortage, environmental pollution, ecological deterioration and other issues gradually deepened. The contradiction between energy supply and demand is becoming increasingly prominent and the global energy crisis is gradually becoming significant [1]. At present, wind power is widely used not only in the developed countries but also in many developing countries. In some developed countries, wind power has partly displaced the conventional power generation modes and is providing the basic driving force of economic development. Although it is not a long time for China to develop and utilize the new energy and the renewable energy, in the recent years, the growth rate of utilizing the new energy is more than 25% per year [2,3].

According to the world wind energy association, worldwide wind capacity has reached 456 GW by the end of June 2016 and 500 GW expected for the full year, the global wind capacity grew by 5% in the first half year and by 16.1% on an annual basis [4]. Presently, in the U.S., wind energy provides nearly 5% of the nation's total electricity generation. With a 65 GW power capacity

* Corresponding author at: Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China.
E-mail address: sywang@amss.ac.cn (S. Wang).

http://dx.doi.org/10.1016/j.apenergy.2017.04.008
0306-2619/

deployment, utility-scale installations in 39 states for wind power generation exceed 12% in 11 of those states [5].

Meanwhile, in order to guarantee maximum yield of the wind energy, it is essential to assess wind energy potential of the possible installation sites. Because wind energy is proportional to the cube of wind speed, it is important to obtain the characteristics of the wind speed. Hence, the wind energy potential assessment plays an important role in the wind farm design, and the wind energy evaluation comes down to the wind speed forecasting. Owing to the nonlinearity and uncertainty of the wind speed, the accuracy of wind speed forecasting plays a vital role in the wind power scheduling. At present, the wind speed forecasting errors are approximately in the range of 20–40% [6,7]. According to the requirement of wind power operation, forecasting can be classified into four scales: very short-term, short-term, mid-term and long-term. Very short-term and short-term forecasting is chiefly served for load tracking and pre-load sharing. The wind turbines maintenance scheduling and the power system management are mainly classified to the long-term and mid-term forecasting methods. Those proposed forecasting models can be divided into three categories: numerical weather models [8], statistical models [9–11] and machine learning models [12–16]. Nevertheless, based on the forecast horizons, it can be classified into the one-step-ahead forecasting and multi-step-ahead forecasting.

Through a review of the previous literature, the methods in those papers conclude data preprocessing, forecast modeling and optimization algorithms for determine parameters. The data preprocessing includes primarily the outlier detection, de-noising and data reduction. Wang and Xiong [17] proposed a hybrid model including the outlier detection, ARMA, and the bivariate fuzzy time series model. By evaluating the performance of the proposed method, the model significantly improves the precision of one-step-ahead wind speed forecasting on a daily scale compared with the individual methods. Wang et al. [18] utilized the SVM technique to detect the outlier and combined the seasonal index adjustment and Elman recurrent neural network method to construct the hybrid model. However, the data preprocessing technique of the hybrid model is vital, Hu et al. [19] used the Empirical Wavelet Transform technique to decompose the primal wind speed series, and Liu et al. [20] presented two novel hybrid models for the wind speed multi-step-ahead forecasting, which combine the fast ensemble empirical model decomposition (FEEDM), mind evolutionary algorithm, genetic algorithm and multi-layer perceptron neural network. The empirical results show that the FEEDM is efficient decomposition technique. Liu et al. [21] proposed four different hybrid models by integrating the four different signal decomposition techniques, including the wavelet decomposition, the wavelet packet decomposition, the empirical model decomposition and the fast ensemble empirical mode decomposition. The results of the experimentation reveal that the fast ensemble empirical mode decomposition and the wavelet packet transform outperform the wavelet transform and the empirical mode decomposition every time. The most prevalent techniques for wind speed and wind energy decomposing are the wavelet transform, the wavelet packet transform, the Fourier transform, the empirical model decomposition and its derivatization. However, there are many factors that can affect the wind speed, such as relative humidity, air pressure, air temperature and so on. Moreover, multicollinearity exists in those variables, which cannot be modeled directly. It is essential to compress the multivariable data. Liu et al. [22] utilized principal component analysis (PCA) to compress the number of input variables and the redundant information of the input variables can be effectively reduced, and the results show that the PCA is efficient.

It is known that the statistical method performs well in short-term, mid-term and long-term forecasting [8]. Presently, along with the blooming development and prevalence of the big data, machine learning and deep learning technique are becoming more and more potential and promising. Therefore, those techniques are widely applied in the wind speed forecasting, including fuzzy logic models [23], artificial neural network [24], and support vector machine [25]. However, with the rapid development of the intelligent algorithms, such as the simulated annealing (SA) [26], particle swarm optimization algorithm (PSO) [23] and gravitational search algorithm (GSA) [27] are widely utilized to optimize the parameters of the forecasting approach. In addition, those literatures have proved that the intelligent optimization algorithms improve the accuracy and stability of the forecasting models.

The main contribution of this paper is to propose a new dynamic integrated approach integrating kernel principal component analysis (KPCA) and optimized core vector regression (OCVR) model based on phase space reconstruction (PSR) to improve the performance of wind speed forecasting in terms of a multi-step-ahead forecasting scheme, forecasting accuracy, time saving, and robustness, and to compare its forecasting performance with some other popular existing forecasting techniques including typical integrated approach without any intelligent optimization scheme and similar integrated approach with an intelligent optimization scheme.

The remainder of this paper is organized as follows: Some related methodologies and a new dynamic integrated approach are introduced in Section 2. The empirical results based on the three wind farms are given in Sections 3 and 4 provides some conclusions.

2. Methodology

Before presenting our new approach, we introduce some methods which will be utilized in the approach.

2.1. Phase space reconstruction

The phase space reconstruction (PSR) was first proposed by Takens [28]. The main idea of this method is to use delay coordinates method to reconstruct the phase space of chaotic time series $x = \{x_i | i = 1, 2, \cdots, n\}$,

$$X = \left\{X_i | X_i = \left[x_i, x_{i+\tau}, \cdots, x_{i+(m-1)\tau}\right]^T, \quad i = 1, 2, \cdots, N\right\} \tag{1}$$

where m and τ are the embedding dimension and delay time respectively, and $N = n - (m-1)\tau$ is the number of point-phase space.

In the method, the selection of embedding dimension m and delay time τ is very important, and the appropriate parameters will improve the accuracy of forecasting. In this paper, the C-C method is employed to compute those two parameters of PSR. In C-C method the embedding dimension and the delay time can be simultaneously estimated by utilizing the correlation integral. For further details on this method, please refer to [29].

2.2. Kernel principal component analysis

Kernel Principal Component Analysis (KPCA) is an improved Principal Component Analysis (PCA), first proposed by Schölkopf et al. [30]. KPCA can be used to generalize linear PCA into nonlinear situation by means of nonlinear kernel function. A key ideal behind KPCA method is to transform the input data into a high dimensional feature space G in which the PCA method can be carried out, and the implicit feature vector in G does not need to be computed explicitly, while it is just done by computing the inner product of two vectors in G with a kernel function.