Energy 76 (2014) 526–541

Contents lists available at ScienceDirect

Energy

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

# A hybrid forecasting model based on outlier detection and fuzzy time series – A case study on Hainan wind farm of China

Jianzhou Wang [a], Shenghua Xiong [b,*]

[a] School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China
[b] School of Mathematics and Statistics, Lanzhou University, Tianshui Road 222, Lanzhou 730000, Gansu Province, China

ARTICLE INFO

*Article history:*
Received 6 May 2014
Received in revised form
31 July 2014
Accepted 11 August 2014
Available online 16 September 2014

*Keywords:*
Outlier detection
ARMA
BPANN
Bivariate fuzzy time series

ABSTRACT

Wind energy is regarded as a worldwide renewable and alternative energy that can relieve the energy shortage, reduce environmental pollution, and provide a significant potential economic benefit. In this paper, a hybrid method is developed to properly and efficiently forecast the daily wind speed in Hainan Province, China. The proposed hybrid forecasting model consists of outlier detection and a bivariate fuzzy time series, which provides a more powerful forecasting capacity of daily wind speed than that of traditional single forecasting models. To verify the developed approach, daily wind speed data from January 2008 to December 2012 in Hainan Province, China, are used for model construction and testing. The results show that the developed hybrid forecasting model achieves high forecasting accuracy and is suitable for forecasting the wind energy of China's large wind farms.



## 1. Introduction

Recent population increases, rapid economy growth, and fossil resource depletion have resulted in energy shortages and severe environment contamination, which have encouraged the use of alternative clean energy resources [1]. Wind power is a type of renewable energy that is globally recognized as an effective method for mitigating climate change, improving energy security, and supporting low-carbon industrial and economic growth [2]. The availability of efficient wind energy forecasts allows for the design of connections or disconnections between wind turbines and conventional generators to attain low-spinning reserves and optimal operational costs [3], thus reducing fluctuation and instability [4].

Recently, numerous forecasting models have been proposed to improve the accuracy of wind speed forecasts; these models can be classified into three categories: (1) physical models, (2) statistical models, and (3) artificial intelligence models. Each of the categories has its advantages and disadvantages, and hybrid models usually provide higher performance than models established with a single algorithm [5]. Currently, combinations of intelligent methods and hybrid methods have been used extensively in wind speed forecasting as a result of the rapid development of artificial intelligence.

As a typical artificial intelligence model, ANNs (artificial neural networks) are used in the study of wind speed forecasting because of their ability to determine arbitrary functions to approximate mechanisms from observed data. In, BPANN (back-propagation artificial neural network) Gnana Sheela and Deepa [6] introduced an intelligent back-propagation algorithm for wind speed prediction that produced more accurate simulation results than conventional back-propagation algorithms [6]. To test the algorithm's performance in one-hour-ahead wind speed forecasting, three types of typical neural networks (adaptive linear element, back-propagation, and radial basis function) were investigated by Li and Shi [7], who showed that no single neural network model consistently out-performed the other models. In Ref. [8], a recurrent neural network was used for wind speed forecasting and out-performed the ARIMA models. Cadenas and Rivera [9] developed an ANN model with two input neurons and one output neuron for short-term wind speed forecasting with good performance in accuracy.

In hybrid models featuring ANNs, the time-frequency transformations EMD (empirical mode decomposition) and WTT (wavelet transform technique) are employed to extract signals from original wind speed data and thereby achieve accurate forecasting. The hybrid model EMD–ANN, which is based on EMD and ANNs, was proposed by Liu et al. [10] for wind speed predictions, producing highly accurate predictions compared to those of the ANN and ARIMA models. Two hybrid forecasting frameworks, the wavelet-genetic algorithm (GA) multilayer perceptron (MLP) and

* Corresponding author.
*E-mail addresses:* xiongshh12@lzu.edu.cn, wjz@lzu.edu.cn (S. Xiong).

http://dx.doi.org/10.1016/j.energy.2014.08.064
0360-5442/

the wavelet-particle swarm optimization (PSO) multilayer perceptron (MLP), were used to predict non-stationary wind speeds and showed suitable performance under the diverse accuracy requirements of wind speed predictions in Ref. [5]. Liu et al. [11] presented three hybrid models, the Wavelet Packet–BFGS, Wavelet Packet–ARIMA–BFGS, and Wavelet–BFGS, in their wind speed predictions and achieved satisfactory performance. Zhang et al. [12] used a novel approach named WTT–SAM–RBFNN that consisted of the seasonal adjustment method (SAM), RBFNN, and wavelet transform technique (WTT) for short-term wind speed forecasting, which was shown to be an effective method for improving prediction accuracy .

As types of hybrid models featuring ANNs, numerous forecasting models have been combined to achieve high forecasting performance in wind speed data series. Pourmousavi Kani and Ardehali [13] proposed a new ANN-MC model that consisted of an ANN (artificial neural network) and MC (Markov chain); the model was demonstrated to be effective in short-term wind speed forecasting [13]. Two new ARIMA–ANN and ARIMA–Kalman hybrid methods were compared by Liu et al. [14], who obtained high accuracy in wind speed forecasting. In [15], two hybrid models (ARIMA–ANN and ARIMA–SVM) were compared individually to the ARIMA, ANN, and SVM forecasting models and were shown to be viable options for wind speed forecasting. A hybrid model consisting of ARIMA (autoregressive integrated moving average) models and ANN (artificial neural network) models was developed for wind speed forecasting and demonstrated higher accuracy than the ARIMA and ANN models for three examined sites [16]. In Ref. [17], a hybrid MM5 (fifth generation mesoscale model)-neural network approach was presented for short-term wind speed predictions with good performance at specific points.

In another type of hybrid model featuring an ANN, different methods were utilized to improve the performance of the ANN. Nan et al. [18] presented a synthesis of correction forecasting models that consisted of a combination forecasting and prediction bias correction model that resulted in higher accuracy in short-time wind speed forecasting than the combination of BP neural network prediction and an ARMA model. A MAS (multiple architecture system) approach involving predictions obtained from different regression algorithms (MLR, MLP, RBF, and SVM) combined in an ensemble was proposed by Bouzgou and Benoudjit [19]; the model was able to improve the precision of wind speed predictions yielded by traditional prediction methods. In Ref. [20], a hybrid computing model integrating self-organizing feature maps and a multilayer perceptron network was developed and provided improved performance in terms of error minimization. Haque et al. [21] proposed a soft computing model (SCM) formulated on a BPNN (back-propagation neural network), RBFNN (radial basis function neural network), and ANFIS (adaptive neuro-fuzzy inference system), which demonstrated enhanced forecasting accuracy for short-term wind speed predictions [21]. A robust two-step methodology consisting of a Bayesian combination algorithm and three neural network models [adaptive linear element network (ADALINE), back-propagation (BP) network, and radial basis function (RBF)] was used by Li et al. [22] for 1-h-ahead wind speed forecasting, providing a unified approach for tackling the challenging model selection issues associated with wind speed forecasting [22]. A new method using a multilayer feed-forward neural network (MFNN) trained by the SPSA was developed for wind speed forecasting at time scales varying from a few minutes to an hour; the model was observed to be superior to the back-propagation algorithm, time-interval averaging approach, ARMA, and persistence method reported in Ref. [23]. In Ref. [4], a hybrid method that consisted of a back-propagation (BP) neural network and seasonal exponential adjustment was proposed, and it was able to forecast the wind speed on a daily scale with smaller mean absolute errors compared to those of results achieved without adjustments. A method called IS-PSO-BP that combines PSO-BP with comprehensive parameter selection was presented; the method afforded higher forecast performance than the basic back-propagation neural network and ARIMA model considered in the case study [24]. In Refs. [5,7,8,10,11,13–16,18–23], hybrid models using different ANNs with different models were proposed or studied to forecast wind speed.

The SVM (support vector machine) is another common type of artificial intelligence model that has attracted the attention of researchers who study wind speed forecasting. Mohandes et al. [25] employed SVM to wind speed prediction with better performance than MLP. A hybrid forecasting approach combining the ensemble empirical mode decomposition (EEMD) and support vector machine (SVM) models was proposed by Hu et al. [26], and exhibited improved wind speed forecasting capability compared to three other methods (ARIMA, SARIMA, and EMD–SVM) [26]. Chen and Yu [27] presented an SVR–UKF approach that integrated an unscented Kalman filter (UKF) with a support vector regression (SVR)-based state-space model; the method showed improved performance in both one-step-ahead and multi-step-ahead wind speed predictions compared with artificial neural networks (ANNs), SVR, autoregressive (AR), and autoregressive integrated with Kalman filter (AR-Kalman) approaches. In Ref. [28], a hybrid model combined with an input selected by deep quantitative analysis, wavelet transform (WT), genetic algorithms (GAs), and support vector machines (SVMs) was used for short-term wind speed forecasting, outperforming the model to which it was compared (persistent model and SVM-GA model). A hybrid model called SARIMA–LSSVM (seasonal auto-regression integrated moving average and least square support vector machine) was developed to predict the mean monthly wind speed in the Hexi Corridor and was shown to be simple and efficient [29]. A new approach called GMCM–GPR that integrates GMCM (Gaussian mixture copula model) and localizes GPR (Gaussian process regression) was developed by Yu et al. [30]; the model was demonstrated to outperform the GMCM–ARIMA and GMCM–SVR methods in long-term wind speed forecasting.

Numerous models have been employed to improve the accuracy of wind speed forecasting. In Ref. [31], four modified methods were proposed for wind speed forecasting. The first two approaches are based on particle swarm optimization (PSO) and the first-order and second-order adaptive coefficient (FAC and SAC) methods, the remaining two methods decompose wind speed data into seasonal and trend components to achieve improved forecasting that is more efficient than that of the original FAC and SAC models. Carapellucci and Giordano [32] constructed a model that adopts a physical–statistical approach for synthetically generating hourly wind speed data [32]. To forecast wind speed, the Mycielski algorithm [33], k-nearest neighbor (k-NN) classification [34], evolutionary support vector regression algorithms [35], modified TK (Taylor Kriging) model [36], kernel ridge regression with active learning [37], fine tuning support vector machines [38], methods of wavelet and classical time-series analysis [39], Bayesian theory and structural break-based modeling [40], and single exponential smoothing methods [41] have been proposed to improve forecasting accuracy.

Although many improvements in forecasting methods having been, wind speed forecasting is still a difficult and challenging task. In 1989, Moghram and Rahman [42] reviewed five forecasting methods: (1) time series; (2) multiple linear regressions; (3) knowledge-based approaches; (4) general exponential smoothing; and (5) state-space and Kalman filtering. No method was determined to be superior, and the research revealed that a single model could not provide the best performance under all situation.