Energy 112 (2016) 208–220

Contents lists available at ScienceDirect

Energy

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

On practical challenges of decomposition-based hybrid forecasting algorithms for wind speed and solar irradiation

Yamin Wang, Lei Wu*

Electrical and Computer Engineering Department, Clarkson University, Potsdam, NY 13699, USA

ARTICLE INFO

Article history:
Received 30 March 2016
Received in revised form
2 June 2016
Accepted 14 June 2016
Available online 5 August 2016

Keywords:
Empirical mode decomposition
Wind and solar forecast

ABSTRACT

This paper presents a comprehensive analysis on practical challenges of empirical mode decomposition (EMD) based algorithms on wind speed and solar irradiation forecasts that have been largely neglected in literature, and proposes an alternative approach to mitigate such challenges. Specifically, the challenges are: (1) Decomposed sub-series are very sensitive to the original time series data. That is, sub-series of the new time series, consisting of the original one plus a limit number of new data samples, may significantly differ from those used in training forecasting models. In turn, forecasting models established by original sub-series may not be suitable for newly decomposed sub-series and have to be trained more frequently; and (2) Key environmental factors usually play a critical role in non-decomposition based methods for forecasting wind speed and solar irradiation. However, it is difficult to incorporate such critical environmental factors into forecasting models of individual decomposed sub-series, because the correlation between the original data and environmental factors is lost after decomposition. Numerical case studies on wind speed and solar irradiation forecasting show that the performance of existing EMD-based forecasting methods could be worse than the non-decomposition based forecasting model, and are not effective in practical cases. Finally, the approximated forecasting model based on EMD is proposed to mitigate the challenges and achieve better forecasting results than existing EMD-based forecasting algorithms and the non-decomposition based forecasting models on practical wind speed and solar irradiation forecasting cases.

1. Introduction

Forecasting of nonlinear and stochastic wind speed and solar irradiation time series (TS) is facing with a significant challenge because of their inherent intermittent and uncertain characteristics [1–3]. From the methodology point of view, forecasting algorithms for wind speed and solar irradiation TS studied in literature can be classified into three main categories. (i) The first one is physical methods, which adopt numerical equations to describe the relationship between the TS to be forecasted and the related geographical and/or environmental factors. For instance, the global solar radiation arriving at the surface of the earth can be calculated by numerical equations with given geographical and numerical weather prediction (NWP) information [4]. In addition, environmental factors like terrain, obstacle, pressure, and temperature can be considered in physical approaches for the wind speed forecasting [5–7]. The Prediktor in Ref. [6] is developed by the Risoe National Laboratory in Denmark, which considers local weather conditions by using the NWP data from the High Resolution Limited Area Model (HIRLAM). The eWind in Ref. [7] developed by AWS TrueWind Inc. in the U.S. adopts a similar physical approach as Prediktor, but includes a high-resolution boundary layer approach as the numerical weather model to take local conditions into account. (ii) The second one is statistic regression methods. Widely used linear regression models include autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA) models [8,9]. (iii) The third one is artificial intelligence (AI) based methods, including artificial neural network (ANN), support vector machine (SVM), extreme learning machine (ELM), and fuzzy logic model [1]. Physical methods usually perform well in the long-term forecasting (i.e., one day to one week ahead), while regression and AI based models usually present good performance in very-short-term forecasting (i.e., a few seconds to half an hour ahead) and short-term forecasting (i.e., half an hour to 6-h ahead) [2]. Indeed, AI based

* Corresponding author.
E-mail address: lwu@clarkson.edu (L. Wu).

http://dx.doi.org/10.1016/j.energy.2016.06.075
0360-5442/

methods may provide more accurate forecasting results than regression models such as ARMA [2]. However, they usually require more extensive adjustments on various parameters and may encounter inefficient or over-fitted training, especially with a large data set, while regression methods are usually easy to implement and can provide timely prediction.

Recently, hybrid forecasting approaches, such as combining physical and statistical approaches, have been demonstrated to be able to effectively leverage advantages of different approaches and, in turn, perform better than individual methods. In some hybrid algorithms, physical methods are used as the first step to forecast wind and solar TS, which supply auxiliary inputs to other statistical/AI based models in the following steps. For instance, in Ref. [10] three indices for the solar irradiation forecasting are first derived from total sky imager, infrared radiation, and pyranometer measurement, respectively, and ANN is trained to establish the correlations between solar irradiation and these three physical indices. Ref. [11] investigated a hybrid method that combines information from processed satellite images with ANN for predicting global horizontal irradiance (GHI) at time horizons of 30, 60, 90, and 120 min.

Indeed, the decomposition based forecasting method is a kind of hybrid forecasting approach that combines advanced decomposition algorithms with different forecasting models for individual decomposed sub-series. The decomposition based forecasting algorithm works as follows [12–15]:

Step 1. The original TS is decomposed into a set of sub-series via decomposition algorithms;

Step 2. Each sub-series is forecasted via a specific forecasting model;

Step 3. Forecasting results of individual sub-series are combined to obtain the final forecasting results of the original TS.

In Step 2, various forecasting models have been applied to individual sub-series according to their distinct characteristics [12–27]. On the other hand, the earliest method to achieve the decomposition is the Fourier transform developed by Joseph Fourier in 1807, while the currently widely used decomposition algorithms in Step 1 include wavelet transform [14,24], and empirical mode decomposition (EMD) [12–13,15–23,25–27]. The Fourier transform is useful for stationary and pseudo-stationary signals, but it may not provide satisfactory results for signals that are highly non-stationary, noisy, or periodic. Using short wavelets instead of long waves as the analysis function is a major advantage of wavelet transform over Fourier transform, and this benefit makes wavelet analysis an interesting alternative for many applications [28]. Indeed, by the wavelet transform, the signal can be decomposed into a sum of more flexible functions, called wavelets, which are localized in both time and frequency [28]. Step forward, compared to the wavelet transform, EMD method is easy to implement because it is an adaptive and fully data-driven technique and does not use any predetermined transforms that depend on the choice of a particular theoretical structure [28,29].

The application of decomposition techniques on forecasting could date back to 1990s. Ref. [30] was among the first of using decomposition techniques on forecasting applications, which adopted wavelet decomposition and ARIMA for the one step ahead forecasting of monthly car sales in Spanish market. Recently, the EMD-based forecasting method has been widely applied in many areas such as crude oil price [12], passenger flow [13], wind speed [15–22], electricity load [25], and solar radiation [26], and the forecasting time horizon ranges from short-term (i.e., half an hour to 6-h ahead) [13,15–20,22], and [26] to long-term (i.e., one day to one week ahead) [18,21], and [25]. In Ref. [12], characteristics of crude oil price TS in West Texas Intermediate were analyzed by EMD, and the decomposed trend series were forecasted by ANN and SVM. Ref. [13] predicted the short-term passenger flow forecasting based on EMD and BP ANN. The short-term hourly wind speed forecasting based on EMD was illustrated in Refs. [15,17]. Ref. [16] predicted 1-h, 3-h and 5-h ahead hourly wind speed based on three improved versions of EMD. One step ahead wind speed forecasting models for TS sampled every 20 min and every month were illustrated in Ref. [18]. Multi-step ahead short-term wind speed forecasting based on EMD and ANN was proposed in Refs. [19] and [20]. EMD-based daily mean wind speed forecast was discussed in Ref. [21]. Ref. [22] predicted multi-step ahead wind speed TS sampled every 1 min via the EMD-recursive ARIMA model. In Ref. [25], EMD was applied on the long-term forecasting of electricity load and Ref. [26] described solar radiation forecasting based on EMD and ANN.

This paper focuses on analyzing the challenges of EMD-based forecasting methods. Existing researches on EMD-based forecasting algorithms mainly focus on two aspects. (i) The first one is to improve the performance of classical EMD. For instance, the ensemble EMD (EEMD) and Genetic Algorithm-Back Propagation (GA-BP) ANN based wind speed forecasting method was proposed in Ref. [15] to solve the mode mixing issue of EMD. Ref. [16] discussed three improved versions of EMD, including EEMD, complementary EEMD, and complete EEMD with adaptive white noises. Furthermore, different settings on parameters of the added white noise in EEMD were compared in Ref. [17]. Ref. [18] investigated the application of fast EEMD which could improve the computational performance of EEMD on wind speed forecast. (ii) The second aspect is to build different forecasting models for individual sub-series after EMD according to their distinct characteristics. For instance, ANN was used in Refs. [19,20] to forecast individual sub-series of wind speed TS data. In Ref. [20], the first sub-series with the highest frequency is excluded from the whole forecasting procedure, and input variables for ANN are determined by the partial autocorrelation function which is inspired by ARMA models. SVM based forecasting models were discussed in Ref. [25] to forecast individual sub-series of electricity load TS data, and ELM was applied in Ref. [26] to forecast individual sub-series of solar radiation TS data. Improved versions of traditional forecasting methodologies have also been adopted on wind speed forecast. For instance, Ref. [21] applied a multiple kernel-based relevance vector regression (RVR) learning algorithm to forecast sub-series of the original wind speed TS. As the number of relevance vectors in RVR is much smaller than that of SVM, and there is no need to set the penalty parameter, RVR can be applied more conveniently than SVM [21]. Recursive ARIMA, which can update parameters of ARIMA models in real time by using previous forecasted data, is applied to forecast individual sub-series of wind speed TS data in Ref. [22]. In Ref. [23], instead of establishing specific forecasting models for individual subseries, only one forecasting model is established, which includes historical data of all sub-series as input variables and wind speed forecasts as outputs.

However, existing EMD-based forecasting methods present two significant challenges that need to be fully addressed before they can be effectively applied in practical cases. The first challenge is that existing EMD-based forecasting methods usually execute Step 1 only once to decompose the original TS into sub-series, which are further divided into training and forecasting data sets. The training data set is used to build forecasting models, and the forecasting data set is used to forecast future values and quantify corresponding forecasting errors. That is, the EMD method would assume that certain future data are known when performing the decomposition procedure in Step 1. However, in the forecasting