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

procedure of practical cases, the TS need to be re-decomposed more frequently with newly obtained data, in order to provide more accurate input data to the next step forecasting of each sub-series. In fact, our case study shows that sub-series of TS with newly obtained data could be significantly different from the one used in training the forecasting models, and in turn the established models may not be suitable for new sub-series and have to be updated more frequently (see detailed analysis in Section 3).

The second challenge is related to the effective utilization of key environmental factors and similar days. For instance, solar irradiation arriving at the surface of the earth is dramatically influenced by the cloud index in the sky. Another example is the electric energy consumption, which presents strong correlation with outdoor temperature. These key environmental factors are usually considered as inputs for improving the accuracy of forecasting models. However, after applying EMD, such correlation information is lost and environmental factors do not present strong correlations with any sub-series (see detailed analysis in Section 4). In addition, in decomposition based forecasting methods, the training data set must be constructed as chronological TS, and in turn the correlation between hourly data and the data of the same hour in previous day may not be properly addressed. Indeed, EMD-based forecasting methods may not be able to effectively utilize correlated environmental factors and similar day analysis, although they have significant potentials for improving the forecasting accuracy. Thus, in Step 2 of existing EMD-based forecasting models, historical TS data are exclusively used and environmental factors have been largely ignored.

This paper presents an in-depth analysis on the challenges of the existing EMD-based forecasting algorithm and proposes potential alternatives. The EMD-based model is compared with non-decomposition based forecasting models including ANN, SVM, and the persistence method, via practical applications on wind speed and solar irradiation forecasting. Furthermore, the correlation between environmental factors and individual sub-series is also analyzed. That is, environmental factors are used for training each sub-series, and the corresponding forecasting results are compared with those from the model without environmental factors. The major contributions of the paper are:

(1) Two challenges for applying existing EMD-based forecasting algorithms in practical cases are analyzed via detailed studies on wind speed and solar irradiation forecasts. The reasons that make the existing EMD-based forecasting algorithms fail to perform well are analyzed in details.
(2) Possible alternatives that may improve the performance of existing EMD-based algorithms are proposed. Practical case studies on wind speed and solar irradiation forecasts show that the performance of the proposed approximated forecasting model based on EMD outperforms the existing EMD-based forecasting methods and the non-decomposition based method.

2. Existing EMD-based forecasting method

This section focuses on the traditional EMD-based forecasting method in literature, including the procedure of EMD, the traditional EMD-based forecasting method (i.e., Algorithm 1), and the performance analysis via a practical cast study on the wind speed forecast.

2.1. EMD

Hilbert-Huang transform is an adaptive and efficient method for analyzing nonlinear and non-stationary TS. Its key component is EMD [31], which can be used to decompose the original TS data into a set of intrinsic mode functions (IMFs). An IMF is a function that satisfies the following two conditions: (i) In the entire data set, the number of extreme values and the number of zero crossings must be equal or differ at most by one; and (ii) At any point, the mean value of the envelopes defined by local maxima and local minima must be zero.

With the above definition of IMF, an TS can be decomposed into a set of IMFs via the following steps [31]:

a) For a TS $x(t)$, identify all local maxima and minima. Connect all local maxima by a cubic spline curve to produce the upper envelope, and connect all local minima by another cubic spline curve to derive the lower envelope. The mean value of the upper and the lower envelopes is defined as $m(t)$, and $h(t)$ is the difference between $x(t)$ and $m(t)$ which can be calculated via (1).

$$h(t) = x(t) - m(t) \tag{1}$$

b) Replace $x(t)$ in (1) by $h(t)$ and repeat the procedure in Step (a) multiple times, until in a certain iteration k $h_k(t)$ becomes an IMF. That is, if D_k defined in (2) is smaller than the predetermined threshold, $h_k(t)$ can be viewed as a sub-series satisfying the two conditions defined for IMF [31]. Designate the first IMF $h_k(t)$ as $\mathrm{IMF}_1(t)$.

$$D_k = \sum_{t=0}^{T} \left| h_{(k-1)}(t) - h_k(t) \right|^2 \Big/ \sum_{t=0}^{T} \left| h_{(k-1)}(t) \right|^2 \tag{2}$$

c) Once $\mathrm{IMF}_1(t)$ is determined, the residue $r_1(t)$ is obtained by subtracting $\mathrm{IMF}_1(t)$ from the original TS data $x(t)$ via (3). Take $r_1(t)$ as the new original TS, and repeat Steps (a)-(b) until the second IMF $\mathrm{IMF}_2(t)$ is obtained. In order to derive all IMFs, the above procedure is repeated n times until $r_n(t)$ is smaller than the predetermined threshold or $r_n(t)$ becomes a monotone function. Finally, a set of IMFs (i.e., $\mathrm{IMF}_1(t)$, $\mathrm{IMF}_2(t)$, …, $\mathrm{IMF}_n(t)$) and the final residue $r(t)=r_n(t)$ are obtained.

$$r_n(t) = x(t) - IMF_n(t) \tag{3}$$

2.2. Existing EMD-Based forecasting procedure

Algorithm 1 shown in Table 1 and Fig. 1 presents the detailed procedure of the existing EMD-based forecasting method in literature [12,13,15–23], and [25–27].

One-step ahead and multiple-step ahead rolling procedures are usually used to forecast individual sub-series. For instance, Fig. 2 presents a GA-BP ANN model based one-step ahead rolling forecasting procedure, in which weights and thresholds among different layers are optimized by GA. The one-step ahead rolling forecasting procedure in Fig. 2 includes three inputs and one output. That is, the first three actual data are used to forecast the fourth data. Once the actual value of the fourth data is available, the second to the fourth actual data are used to forecast the fifth data and so on.

2.3. Performance analysis of the existing EMD-Based forecasting method

In this section, the performance of the existing EMD-based

Table 1
Algorithm 1-existing EMD-based approach.

Algorithm 1: Existing EMD-based forecasting procedure
1. Decompose the original TS and get individual sub-series through Steps (*a*)–(*c*) in Section 2.1.
2. Divide each sub-series into training and forecasting data sets. Build a specific forecasting model for each sub-series based on the training data set.
3. Obtain forecasting results of individual sub-series.
4. Aggregate results of individual sub-series to obtain the final forecasting results of the original TS. Evaluate the forecasting error via the forecasting data set.

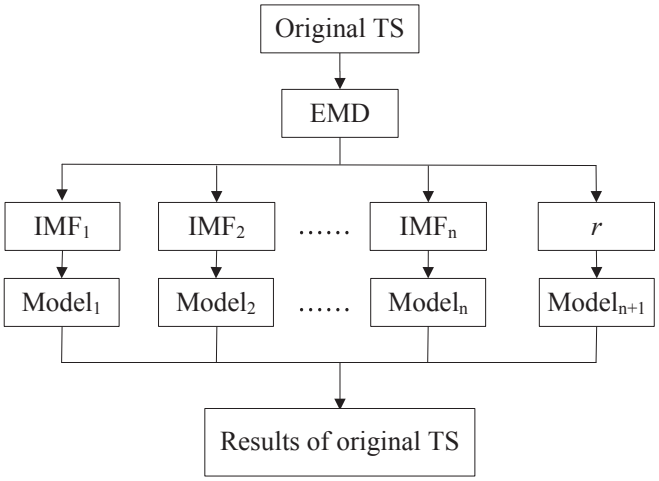

Fig. 1. Flowchart of the existing EMD-based forecasting method.

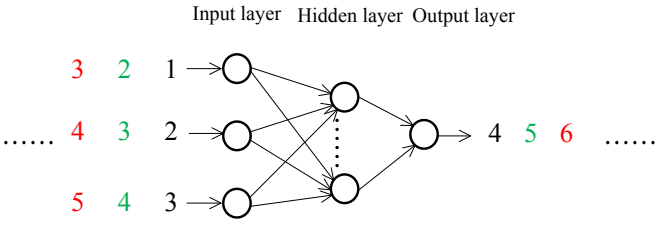

Fig. 2. Illustration of the one-step ahead rolling forecasting procedure with ANN.

approach is analyzed via a practical case study on the wind speed forecast. Wind speed data samples for every 10-min in a wind farm of Inner Mongolia in China are used. The rolling forecasting procedure is applied to forecast each sub-series, while weather factors and similar day analysis are not included in this case study.

Fig. 3 shows the original wind speed TS and sub-series after decomposition. As shown in Fig. 3, the original wind speed TS is decomposed into 8 sub-series by EMD, including seven IMFs of IMF1-IMF7 and the final residue r. IMF1 and IMF2 present high frequency components with small magnitudes. In comparison, frequencies of IMF 3-IMF7 are low, while the residue r is monotonous with a large magnitude.

Both GA-BP ANN and SVM based forecasting models are established for individual sub-series. 274 data which is about 75% of the entire data series is used for training the model (data to the left of red curves in Fig. 3), and the rest 25% data (data to the right of red curves in Fig. 3) are used for testing the forecasting performance. For each sub-series, the one-step ahead rolling forecasting procedure shown in Fig. 2 is applied, in which the first three actual data are used to forecast the fourth data. For the sake of comparison, the persistence forecasting approach is also applied as the benchmark method [1] to forecast each sub-series. The persistence forecasting method is the simplest approach, which assumes that the forecasting value at time $(t + \Delta t)$ is the same as the actual data at time t when Δt is small enough. Root mean square error (RMSE) (4) and normalized RMSE (nRMSE) (5) are used as metrics to evaluate the performance of different forecasting models. In (4) and (5), y_t and f_t denote the actual value and the forecast value, respectively. T is the total number of data used for the performance evaluation, and in (5) y_{max} and y_{min} are the maximum and minimum values of y_t, respectively. The forecasting results are shown in Table 2. In Table 2, the bold numbers are final forecasting errors of the entire TS with and without EMD.

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - f_t)^2} \tag{4}$$

$$nRMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - f_t)^2} \Big/ (y_{max} - y_{min}) \tag{5}$$

Results in Table 2 indicate that forecasting errors of IMF3-IMF7 and r are much smaller than those of IMF1 and IMF2. The reason is that sub-series with lower frequencies present less sharp ramps, have relatively clearer trend, and in turn could be forecasted more accurately. Indeed, Fig. 3 shows that IMF1 and IMF2 present an insignificant proportion in the original wind speed TS, as the magnitudes of IMF1 and IMF2 are much smaller as compared to the original wind speed TS. In turn, although forecasting errors of IMF1 and IMF2 are relatively high, the final forecasting error of the entire TS is smaller than that of the non-decomposition based forecasting model. Results in Table 2 also indicate that performances of GA-BP ANN and SVM on forecasting individual sub-series are not significantly different. In addition, the persistence method performs very well in forecasting IMF6, IMF7, and *r*. Thus, when the computational burden of forecasting models is a concern, advanced forecasting methods may not be necessary to forecast sub-series with small deviations such as IMF6, IMF7, and *r*.

3. Significant changes in sub-series with newly obtained data

This section analyzes the first practical challenge of the existing EMD-based forecasting method (i.e., Algorithm 1 in Section 2). That is, Algorithm 1 may not be suitable in practical cases because it assumes that certain future data are known when performing the decomposition procedure in Step 1. Indeed, in practical situations while certain future data are unknown, sub-series of the original TS with newly obtained data may be significantly different from the one used in training the forecasting models, and in turn the established models may not be suitable for newly decomposed sub-series. This section presents two possible alternatives (i.e., Algorithms 2 and 3) which do not assume any known future data, and analyzes their performance via a practical case study on the wind speed forecast.

3.1. Changes in sub-series

In existing EMD-based forecasting applications, each sub-series is divided into training and forecasting data sets. The training set is used to build proper forecasting models for individual sub-series,

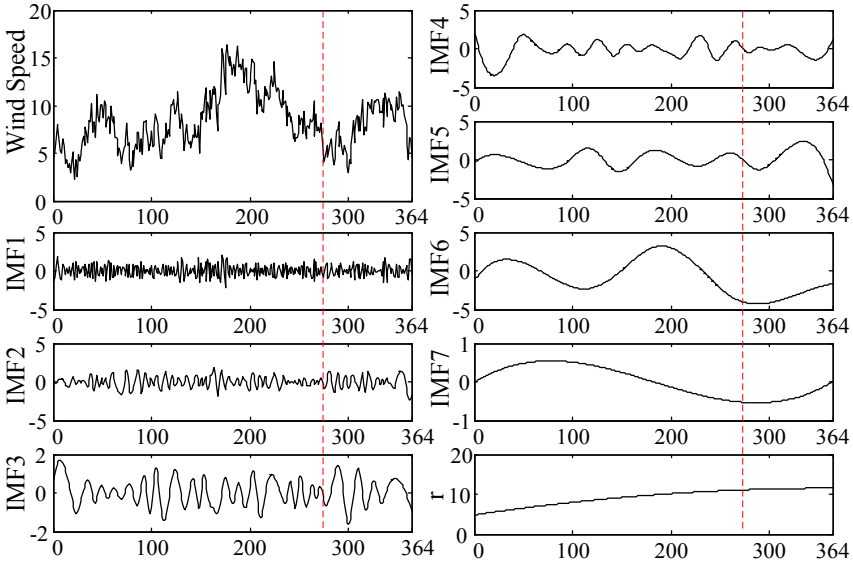

Fig. 3. Sub-series obtained by EMD.

Table 2
RMSE (m/s) and nRMSE of individual IMFs and *r* obtained by the existing EMD-based forecasting procedure.

		GA-BP ANN	SVM	Persistence
RMSE/nRMSE	IMF1	0.7146/0.1709	0.7829/0.1872	1.0911/0.2609
	IMF2	0.2612/0.0624	0.2558/0.0611	0.5326/0.1273
	IMF3	0.0247/0.0073	0.0428/0.0127	0.2240/0.0665
	IMF4	0.0176/0.0032	0.0243/0.0044	0.0716/0.0129
	IMF5	0.0056/0.0010	0.0397/0.0071	0.1166/0.0208
	IMF6	0.0027/0.0004	0.0161/0.0021	0.0350/0.0047
	IMF7	0.0023/0.0021	0.0079/0.0073	0.0070/0.0064
	r	$0.0003/4.5e^{-5}$	$0.0039/5.8e^{-4}$	$0.0055/8.2e^{-4}$
	Entire TS with EMD	**0.7340/0.0518**	**0.7646/0.0539**	**1.2568/0.08861**
	Entire TS without EMD	**1.1856/0.0836**	**1.1873/0.0837**	**1.2568/0.0886**

while the forecasting set is used to analyze forecasting errors and seek for proper model improvement. That is, inputs in each step of the rolling based forecast procedure are assumed to be known before the entire forecasting procedure starts.

In the non-decomposed method, the assumption that the data set is given and can be divided into training and forecasting data sets is reasonable, as they would not change significantly when newly data is available. However, this assumption may not be applicable for the EMD-based method in practical applications. The reason is that in order to obtain input data in the next step of the rolling based forecast procedure for each sub-series, the original TS needs to be re-decomposed when a new actual data is obtained, and individual sub-series could be significantly different from the ones that are decomposed before. For instance, Fig. 4 shows two sets of decomposed sub-series including IMF4-IMF7 and r that are used in the case study of Section 2.3. Red lines are sub-series obtained by decomposing 274 training wind speed data points, and blue lines are sub-series obtained by decomposing the total of 364 wind speed data points, which contain 274 training wind speed data points and 90 new wind speed data points. It shows that with newly available data, values of individual sub-series are significantly different. Furthermore, Fig. 5 shows the residues r obtained from five wind data sets, including 275 up to 279 wind data samples. It indicates that even though the length of TS differs by one data sample, the decomposed sub-series could change significantly. This means that individual sub-series input data could change significantly along the rolling forecasting procedure. In turn, it is highly possible that the forecasting model built based on an old sub-series (even with one step lag) may no longer be suitable for the newly decomposed sub-series.

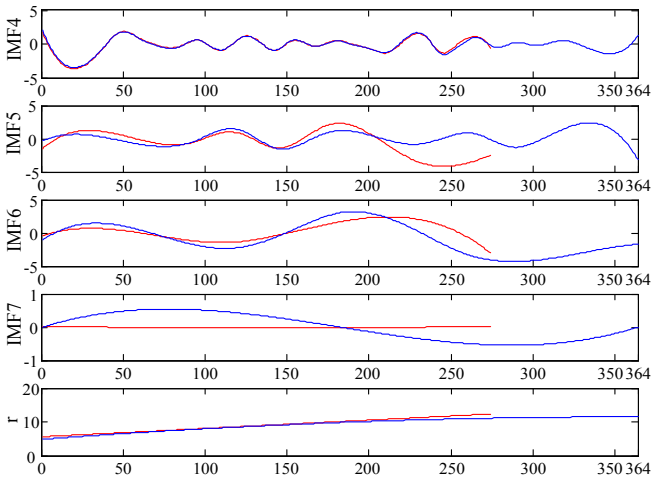

Fig. 4. Sub-series decomposed by 274 (red) and 364 (blue) samples of wind speed data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Another issue caused by the changes is that in the rolling procedure, different values of the same data point are used in different

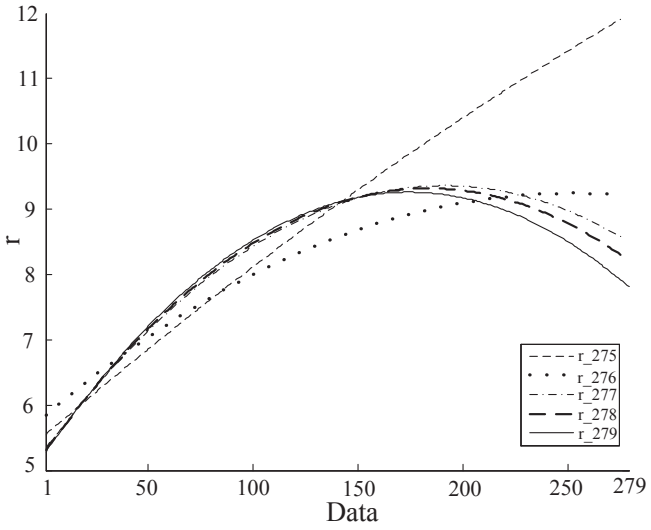

Fig. 5. Residues *r* obtained from 275 up to 279 samples of wind speed TS.

forecasting steps. For instance, in the rolling forecasting steps of case study in Section 2.3, the 273th–275th data are used to forecast the 276th data (i.e., curve r_275), and similarly the 274th–276th data are used to forecast the 277th data in the next rolling forecasting step (i.e., curve r_276) and so on. The issue is that 274th data used in the two steps are different. Thus, it is important to define the consistent data that will be used in each forecasting step. In the following case studies, we define that the input data used in each step are the most up-to-date data obtained in this step (i.e., the 274th data in curve r_276 is used to forecast the 277th data).

3.2. Practical EMD-based forecasting procedure

From the above analysis, the existing EMD forecasting procedure in Algorithm1 may not be suitable for practical applications. In EMD-based practical forecasting cases, only the training data can be treated as known, while the forecasting data should be viewed as unknown data that are dynamically changing over times. Once a new actual data is available, the original TS should be decomposed again to include the new input data into sub-series. This revised procedure for practical applications is shown in Table 3 as Algorithm2. The main difference between Algorithms 1 and 2 is that in Algorithm 1, the original TS is decomposed only once and in turn the inputs to each step of the rolling forecasting procedure are determined in advance, while in Algorithm 2 inputs for each step of the rolling forecasting procedure are dynamically obtained by re-decomposing the original TS with newly obtained data. That is, in Algorithm 2 the most up-to-date 274 data TS is decomposed to forecast the next data point in each step.

Fig. 6 shows the training data set and the forecasting data set for each sub-series in Algorithm 2. For instance, the first data on the right-hand-side of the red curve (i.e., the 275th data in the TS) in each sub-series of Fig. 6 are obtained by decomposing the data set of the 2nd–274th training data plus the forecasted value of the 275th data, the second data in each sub-series are obtained by decomposing the data set of the 3rd–274th training data plus two forecasting data, and so on. It is shown that the right-hand-side profiles of the red curve in Fig. 6 are significantly different from

Table 3
Algorithm 2-practical EMD approach.

Algorithm 2: Practical EMD-based forecasting procedure
1. Divide the data set into training and forecasting data sets. The forecasting data set is assumed to be unknown.
2. Decompose the most up-to-date 274 data TS (which is a combination of actual historical data and forecast values of the most recent time slots) and get individual sub-series. Build a specific forecasting model for each sub-series based on the training data set.
3. Do one-step ahead forecasting. If all forecasting tasks are done, go to Step 4; Otherwise, go back to Step 2.
4. Aggregate results of individual sub-series to obtain the final forecasting results to the original TS. Evaluate the forecasting error via the forecasting data set.

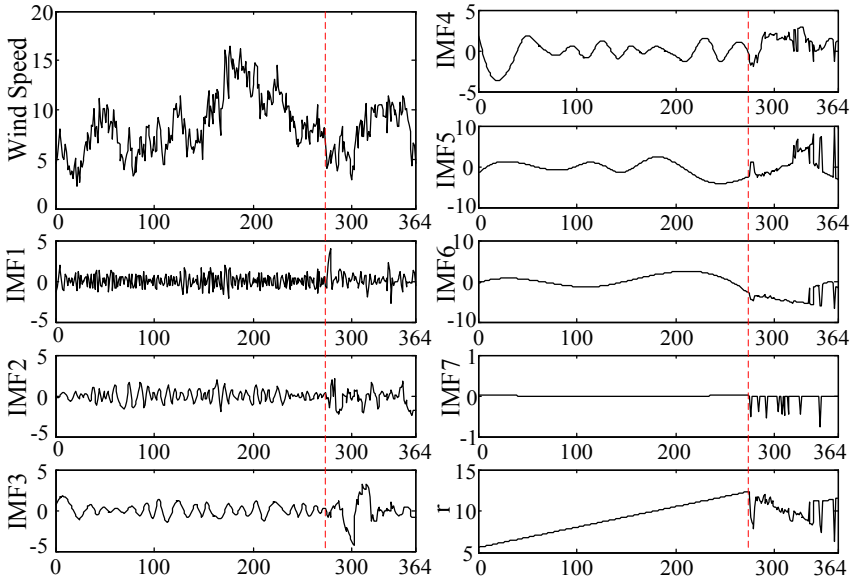

Fig. 6. Sub-series obtained in practical cases.

Table 4
RMSE (m/s) and nRMSE of each IMF and *r* obtained by Algorithm 2.

		GA-BP ANN	SVM	Persistence
RMSE/nRMSE	IMF1	1.0599/0.1582	1.9115/0.2854	1.0320/0.1541
	IMF2	0.9099/0.2056	0.8397/0.1897	0.8501/0.1921
	IMF3	0.9399/0.1242	0.7377/0.0975	0.6937/0.0916
	IMF4	0.8937/0.1882	0.8247/0.1737	0.7756/0.1633
	IMF5	2.3404/0.1914	2.4451//0.2000	1.4678/0.1200
	IMF6	1.2805/0.1874	1.3774/0.2015	1.5990/0.2339
	IMF7	0.2114/0.2631	0.2940/0.3659	0.4591/0.5714
	r	1.0402/0.0864	1.2481/0.1038	1.9281/0.1603
	Entire TS with EMD	**1.9480/0.1374**	**2.0806/0.1467**	**1.2655/0.0892**
	Entire TS without EMD	**1.1856/0.0836**	**1.1873/0.0837**	**1.2655/0.0892**

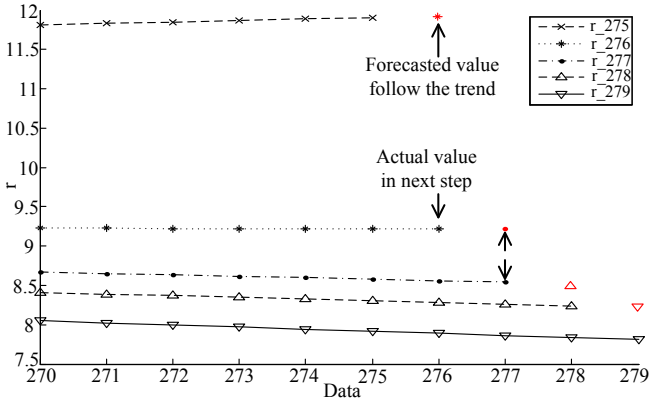

Fig. 7. Residue decomposed by 275, 276, 277, 278, and 279 data TS respectively.

those in Fig. 3. Indeed, because of the changes analyzed in Section 3.1, profiles of each sub-series on the right-hand-side are different from those training data sets on the left-hand-side, which may induce significant errors if the same forecasting models are used. An extreme case is the residue r, in which the left-hand-side presents a monotonous trend while the right-hand-side profile fluctuates intensively.

The wind speed case study in Section 2 is revisited by applying Algorithm 2. GA-BP ANN, SVM, and persistence based forecasting methods are used in this EMD-based practical forecasting case, and the results are presented in Table 4. In Table 4, the bold numbers are final forecasting errors of the entire TS with and without EMD. It is shown that all forecasting errors of sub-series in Table 4 are larger than the corresponding ones in Table 3 from Algorithm 1.

To further analyze this issue, the last data segments of the five curves in Fig. 5, including data 270–279, are illustrated in Fig. 7. It is clearly shown that forecasted values based on the training data set (i.e., marked as red) present significant deviations as compared to the actual values obtained in the next time step of the rolling forecasting procedure. This large deviation is caused by the changes in sub-series when new actual data are available.

It is worth mentioning that the forecasting error of IMF5 In Table 4 is even larger than the error of entire TS. A simple example is shown in Fig. 8 to explain this phenomenon. Assuming the original TS is decomposed into one IMF and r, as shown in Fig. 8. The left-hand side of the red curve is the training data set, and the right-hand side is the forecasting result. Following the trend in the training data set, forecasting results for IMF1 and r are 6 and 5 (marked as red) respectively, while the real values in the next step are 4 and 7 respectively. In turn, the forecasted TS in this step is $6 + 5 = 11$, which is the same as the actual value. Thus, the forecasting result of the original TS could be the same as the real value, even though large forecasting errors may exist in certain sub-series.

In addition, Table 4 shows that forecasting errors of the EMD-

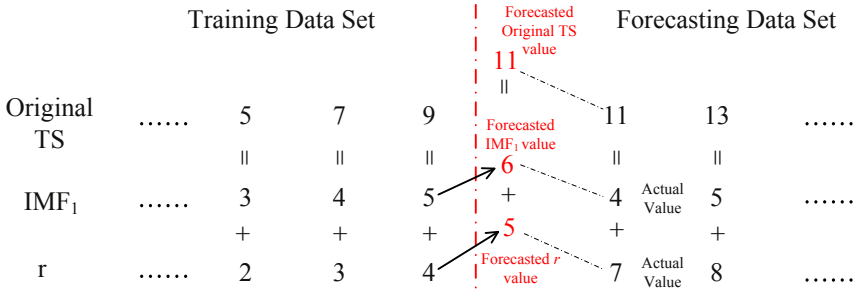

Fig. 8. Illustration of larger RMSE of IMF than Original TS.

Table 5
Algorithm 3-the revised practical EMD approach.

Algorithm 3: revised practical EMD-based forecasting procedure
1. Divide the data set into training and forecasting data sets.
2. Decompose a predefined TS plus one new data in the training data set and get individual sub-series.
3. Repeat Step (2) *N* times and obtain all inputs and outputs for the training data set. Build a specific forecasting model for each sub-series based on the training data set.
4. Decompose the previous 274 data TS and get individual sub-series.
5. Perform forecasting tasks with newly obtained data. If all forecasting tasks are done, go to Step 6; Otherwise, go back to Step 4.
6. Evaluate the forecasting error via the forecasting data set.

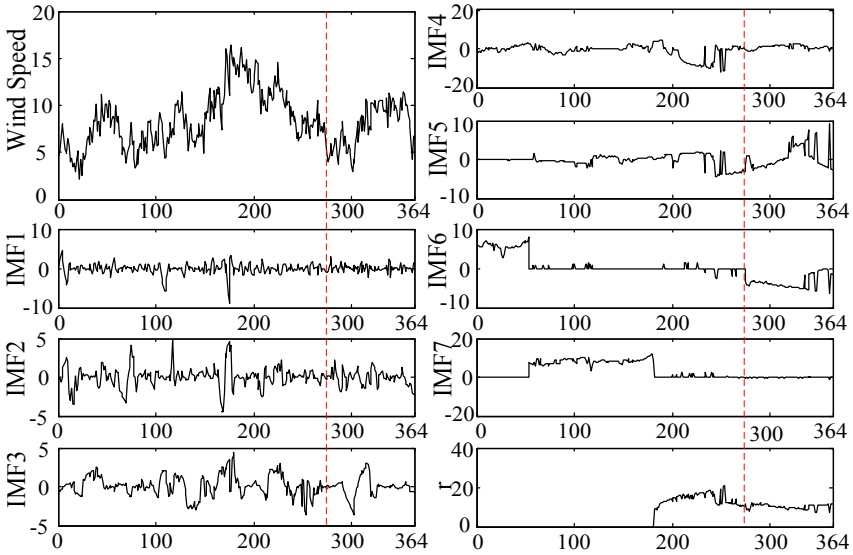

Fig. 9. Sub-series obtained by Algorithm 3.

Table 6
RMSE (m/s) and nRMSE of each IMF and *r* obtained by Algorithm 3.

		GA-BP ANN	SVM	Persistence
RMSE/nRMSE	IMF1	0.9480/0.1415	0.9824/0.1467	1.0503/0.1570
	IMF2	0.4748/0.1073	0.4847/0.1095	0.8635/0.1951
	IMF3	0.1963/0.0258	0.2314/0.0305	0.7004/0.0922
	IMF4	0.9972/0.2100	0.9875/0.2080	0.7516/0.1583
	IMF5	3.9999/0.3271	3.8795/0.3173	1.9030/0.1556
	IMF6	3.8051/0.5567	3.7451/0.5479	1.1341/0.1659
	IMF7	0.2122/0.2641	0.2213/0.2754	0.3061/0.3810
	r	2.1411/0.1780	2.2113/0.1838	0.9053/0.0753
	Entire TS with EMD	**3.8455/0.2711**	**3.8542/0.2717**	**1.2568/0.0886**
	Entire TS without EMD	**1.1856/0.0836**	**1.1873/0.0837**	**1.2568/0.0886**

Table 7
RMSE (m/s) and nRMSE of individual IMFs and the residue r obtained by EEMD-based procedure in Algorithm 2.

		GA-BP ANN	SVM	Persistence
RMSE/nRMSE	IMF1	0.7140/0.1904	0.7624/0.2036	0.6860/0.1832
	IMF2	0.5551/0.1819	0.4933/0.1616	0.4356/0.1427
	IMF3	0.5096/0.1727	0.4105/0.1391	0.2796/0.0947
	IMF4	0.3645/0.1774	0.3242/0.1578	0.2056/0.1001
	IMF5	0.8323/0.1021	0.8845/0.1085	0.5080/0.0623
	IMF6	0.6525/0.1595	0.6752/0.1650	0.4008/0.0979
	IMF7	0.1456/0.5407	0.1622/0.6023	0.1240/0.4605
	r	0.4350/0.0389	0.5031/0.0451	0.3111/0.028
	Whole TS	**1.7019/0.1200**	**1.7321/0.1221**	**1.3635/0.0961**
	Without EMD	**1.1856/0.0836**	**1.1873/0.0837**	**1.3635/0.0961**

based method for the original TS could be larger than those of the non-decomposition based forecasting method, which indicates that the combination of EMD and forecasting model may not be always effective. The results in Table 4 also present that errors of the EMD-based method are larger than that of the benchmark persistence method. In sum, directly applying the EMD-based forecasting algorithm in practical cases may not be effective.

Indeed, the poor performance is mainly caused by the inconsistency between the training data TS and the forecasting data TS. Since re-decomposing TS is inevitable in practical forecasting procedure, in order to make training data set to be consistent with the forecasting data set, the training procedure in Algorithm 2 is revised to Algorithm 3 as shown in Table 5. In Algorithm 3, the training data set is also structured by decomposing the TS recursively. Sub-series obtained by Algorithm 3 are shown in Fig. 9. Forecasting results are shown in Table 6. In Table 6, the bold

Table 8
Correlation coefficients between different TS data and the environmental factors.

	C	*H*	*Ws*
Original TS	−0.337	−0.549	0.030
IMF 1	−0.045	−0.099	−0.010
IMF 2	0.052	−0.248	0.017
IMF 3	0.027	−0.116	0.035
IMF 4	0.154	−0.130	−0.162
r	−0.174	0.0164	−0.1437

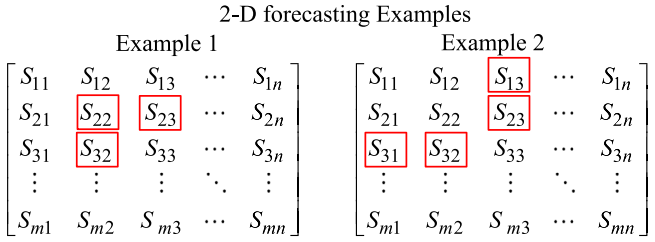

Fig. 10. Illustration of 2-D solar irradiation forecasting.

numbers are final forecasting errors of the entire TS with and without EMD. Results in Table 6 show that all sub-series still present large forecasting errors, and the final forecasting error of the original TS from Algorithm 3 is larger than that of Algorithm 2. Thus, although Algorithm 3 makes the training data set to be more consistent with the forecasting data set, the sub-series become more difficult to forecast and the performance gets even worse. In turn, the EMD-based forecasting procedure in Algorithm 3 may still be inefficient in practical cases.

3.3. Improved EMD-Based forecasting procedure

Existing literature have shown that a better performance could be obtained by applying the improved EMD in Algorithm 1 [15,16]. This section further tests the performance of improved EMD when applied to practical case studies.

As an improved version of EMD, EEMD is a noise-resistant data analysis method for solving the problem of mode mixing in EMD. EEMD, instead of EMD, is applied to Algorithm 2 in the following case study. The standard deviation of the white noise is set as 0.2, and the ensemble number is 50 in EEMD. The forecasting procedure is based on Algorithm 2. The results are shown in Table 7. Compared to Table 4, results in Table 7 show that the forecasting performance is improved by applying EEMD in Algorithm 2. However, forecasting errors of the entire TS in Table 7 are still larger than those of the non-decomposition based method.

In conclusion, the existing EMD-based forecasting procedure in Algorithm1 cannot be applied to practical cases directly. In addition, forecasting accuracy of EMD and EEMD-based forecasting methods in Algorithm 2 and Algorithm 3 does not show advantage over non-decomposition based forecasting algorithms in practical forecasting cases. As different forecasting models such as GA-BP ANN and SVM do not show significant difference in forecasting individual sub-series, the research focus in this direction should be more on solving the challenges of applying the decomposition based forecasting methods in practical cases, instead of exploring various forecasting models and improved versions of EMD.

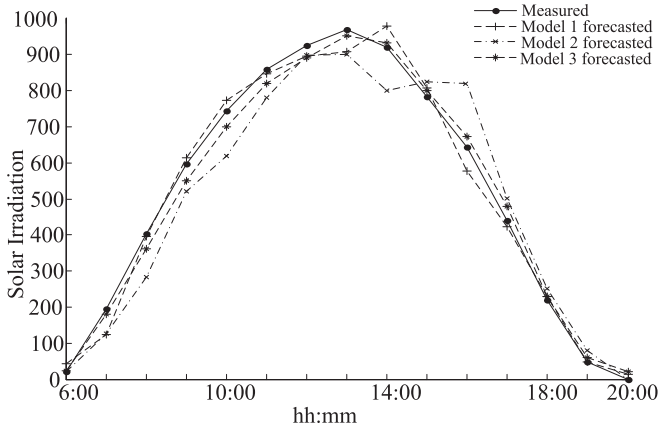

Fig. 11. Forecasting results of Models 1–3 in 5/5/2010.

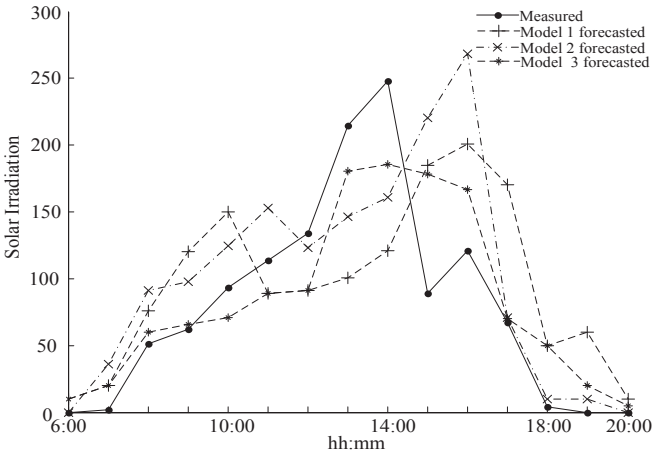

Fig. 12. Forecasting result of Models 1–3 in 10/17/2010.

4. Environmental factors and similar day analysis

This section analyzes the second challenge of the existing EMD-based forecasting method (i.e., Algorithm 1 in Section 2 and Algorithms 2–3 in Section 3). That is, strong correlations between key environmental factors and the original TS are not sustained in sub-series and in turn fail to provide useful information for improving forecasting results of individual sub-series. In addition, similar day analysis cannot be effectively applied in the existing EMD-based methods.

In existing EMD-based forecasting models, only historical TS data are used to train forecasting models for individual sub-series. This procedure may be acceptable for wind speed forecasting in which environmental factors may not be a significant driven. However, environmental factors could play vital roles in electricity load and solar irradiation forecasting. For instance, cloud index dramatically influences the solar irradiation received on the ground. However, different from non-decomposition based forecasting methods, EMD-based forecasting algorithm cannot effectively deal with environmental factors, because strong correlation between environmental factors and the original TS is lost after decomposition. That is, decomposed sub-series do not present strong correlations with environmental factors. In addition, similar day analysis has been widely applied on electricity load and solar irradiation forecasting. Since similar days are constructed by environmental factors, the similar day analysis cannot be efficiently applied in EMD-based forecasting algorithm either.

A practical solar irradiation forecasting case study is used to illustrate this issue with the EMD-based forecasting method. Solar irradiation and related key environmental data including cloud index C, humidity H, and wind speed Ws are obtained from the National Solar Radiation Database [29]. The first 1380 hourly data in San Francisco in 1991 is used to analyze the correlation between global solar radiation TS and the environmental factors. After the original TS is decomposed by EMD, the correlation between each IMF and the environmental factors is also analyzed. The Pearson correlation coefficient (6) is used to measure the correlation effect. $\boldsymbol{p}$ and $\boldsymbol{q}$ are $1 \times n$ TS vectors. p_i and q_i are the ith elements in $\boldsymbol{p}$ and $\boldsymbol{q}$, respectively. $\overline{p}$ and $\overline{q}$ are average values of all elements in $\boldsymbol{p}$ and $\boldsymbol{q}$, respectively.

Table 9
RMSE (W/m^2) and nRMSE of the solar irradiation forecast.

	5/5/2010				10/17/2010			
	Model 1	Model 2	Model 3	CP Model	Model 1	Model 2	Model 3	CP Model
RMSE/nRMSE	83.41/0.09	37.45/0.04	28.34/0.03	40.59/0.04	69.02/0.27	62.26/0.25	37.56/0.15	49.08/0.20

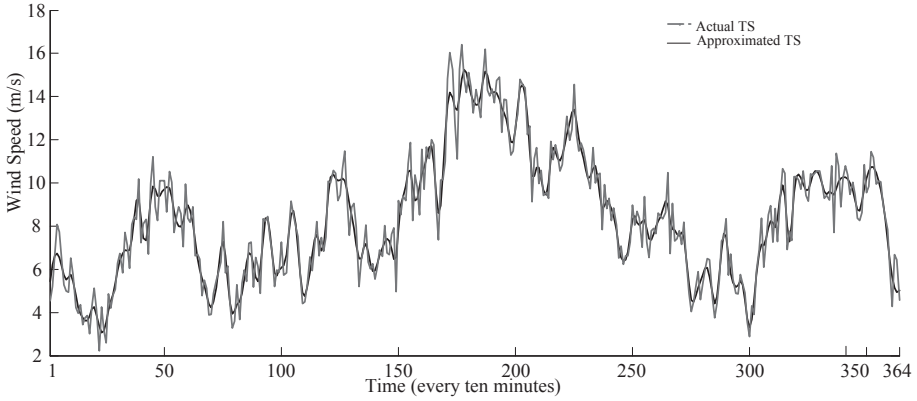

Fig. 13. Approximated wind speed TS.

$$r_{pq} = \frac{\sum_{i=1}^{n}(p_i - \overline{p})(q_i - \overline{q})}{\sqrt{\sum_{i=1}^{n}(p_i - \overline{p})^2}\sqrt{\sum_{i=1}^{n}(q_i - \overline{q})^2}} \quad (6)$$

Correlation coefficients between different TS data (including original TS, IMF1-IMF4, and r) and the environmental factors (including *C*, *H*, and *Ws*) are shown in Table 8. Table 8 shows that the original solar irradiation TS presents strong correlations with cloud index and humidity. However, after decomposition, this correlation with individual IMFs is not obvious. In turn, including environmental factors into the training set of each sub-series may even deteriorate the training effect.

Days which have similar cloud index and humidity conditions with the forecasting day can be viewed as similar days. For non-decomposition based solar irradiation forecasting method, hourly irradiation in previous similar days are also taken as inputs to forecast the irradiation of same hour in the forecasting day, which is called a 2-D forecasting method [33]. The 2-D model for solar irradiation forecasting is illustrated in Fig. 10, where S_{mn} means the solar irradiation in the nth hour of the mth day. In the 2-D model, the inputs for forecasting S_{mn} include hourly irradiation in previous hours of the same day and in the same hour of previous days. For instance, S_{33} could be forecasted by S_{22}, S_{23}, and S_{32} as illustrated in Example 1 of Fig. 10, or by S_{13}, S_{23}, S_{31}, and S_{22} as illustrated in Example 2 of Fig. 10 [33].

However, the 2-D forecasting technique cannot be effectively applied in the EMD-based forecasting method. In the decomposition based method, data of all similar days should be gathered together into a single TS, and after decomposition correlations between solar irradiation in the same hour of different days are lost. In other words, the decomposition based forecasting method takes any kinds of data as a TS without recognizing its physical characteristics, and forecasts each sub-series using the simple rolling procedure.

To further analyze the impact of environmental factors and similar day analysis, three models are built to forecast the hourly solar irradiation. Model 1 is the EMD-based approach, and its inputs include three previous historical TS data for each sub-series as well as related cloud index and humidity information. In this model, data of historical days which have similar environmental factors with the forecasting day are used to construct the original solar irradiation TS, which is then decomposed by EMD. Model 2 and Model 3 are non-decomposition based forecasting method. Inputs of Model 2 only include four previous historical TS data in the constructed similar day series. Model 3 is a 2-D forecasting model, and its inputs include data in previous 2 h and data of the same hour in the previous two similar days. SVM is used to forecast each sub-series in Model 1 and TSs in Model 2 and Model 3. The rolling forecasting procedure in Model 1 is based on Algorithm 2. The hourly solar irradiation of two days in San Francisco including 5/5/2010 and 10/17/2010 are forecasted. Forecasting results are shown in Fig. 11 and Fig. 12, respectively. RMSE (W/m^2) and nRMSE results are shown in Table 9.

The clearness persistence (CP) model proposed in Ref. [34] is used as a benchmark to evaluate the performance of the above three forecasting models. In the CP model, the ratio of the measured irradiance to the extraterrestrial solar radiation is assumed to be same in two successive time-steps. In Ref. [34], the extraterrestrial solar radiation used in the CP model is computed by (7), where I_0 is the extraterrestrial solar radiation, $G_0 = 1367$ W/m^2 is the solar constant, and θ_z is the solar zenith angle. The extraterrestrial data from Ref. [32] are used to set the value of I_0 in this paper. Forecasting results via the CP model are also shown in Table 9.

Table 10
RMSE (m/s) and nRMSE of the approximate forecasting model on wind speed forecasting.

	Algorithm 2	Algorithm 3	Without EMD	Approximated model
RMSE/nRMSE	1.9480/0.1374	3.8455/0.2711	1.1856/0.0836	1.0024/0.0707

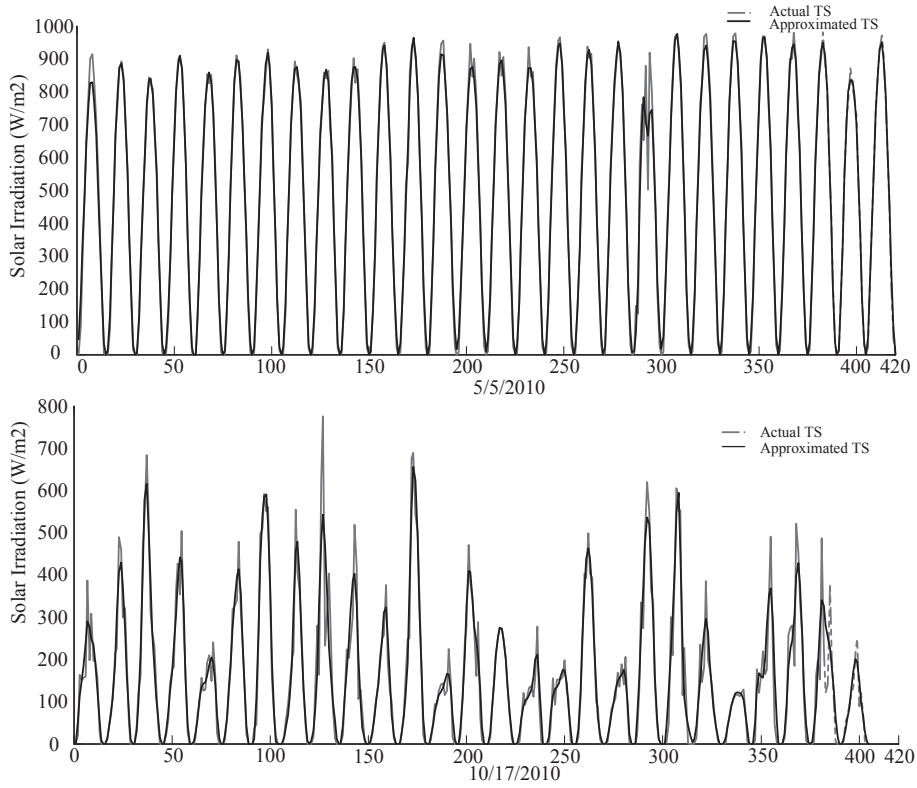

Fig. 14. Approximated solar irradiation TS.

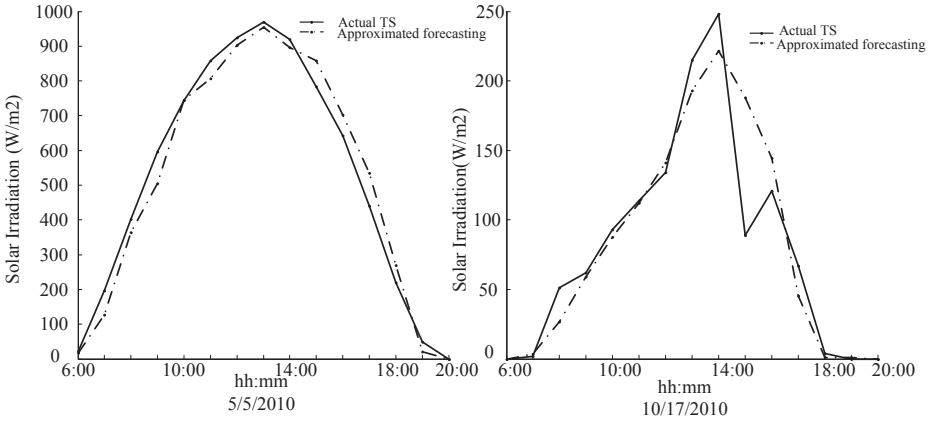

Fig. 15. Forecasting results of solar irradiation on 5/5/2010 and 10/17/2010.

Table 11
RMSE (W/m^2) and nRMSE of the approximated forecasting model on solar irradiation forecasting.

	5/5/2010	10/17/2010
RMSE/nRMSE	33.35/0.04	30.08/0.12

$$I_0 = G_0 \cos(\theta_z) \tag{7}$$

5/5/2010 is a blue-sky day. Fig. 11 and the results in Table 9 show that Model 3 presents the best performance among the three models. The lag phenomenon in Model 2 is common in the rolling forecasting procedure, which is effectively solved in Model 3 by considering solar data of previous days. In addition, the performance of Model 2 is better than Model 1, which means that the EMD-based forecasting model cannot be efficiently applied on the solar irradiation forecasting. On the other hand, 10/17/2010 is a cloudy day, which means that the solar irradiation fluctuates significantly with the change of cloud index. Fig. 12 also consistently shows that forecasting results of Model 3 is the best, and the performance of Model 2 is better than that of Model 1. In both cases, the forecasting error of Model 1 are larger than the CP model, which indicates that Model 1 is not an effective forecasting method in practical solar forecasting cases.

In conclusion, after decomposition, environmental factors, which have strong correlations to the original TS data, do not present strong correlations to individual sub-series. Thus, the training effect may be deteriorated when environmental factors are

included as additional inputs. The 2-D forecasting method cannot be efficiently applied in the EMD-based forecasting procedure. In turn, the EMD-based forecasting method fails to show advantages over non-decomposition based forecasting models when applied to solar irradiation forecasts.

5. Approximated forecasting model based on EMD to mitigate the challenges

The previous two sections analyze the two major challenges that would impede the application of existing EMD-based forecasting algorithms on practical wind speed and solar radiation forecasts. In this section, an approximated forecasting model based on EMD is proposed as an alternative way to mitigate such challenges when integrating EMD into practical wind speed and solar radiation forecasting procedures.

The basic idea is to obtain the approximated TS by removing IMF1 with the highest frequency from the original TS, and establish a single forecasting model for the approximated TS directly. As shown in Fig. 3, sub-series with higher frequencies such as IMF1 present more sharp ramps, relatively unclearer and more volatile trends, and smaller magnitudes. In comparison, as shown in Fig. 13, the approximated wind speed TS by excluding IMF1 presents less sharp ramps and relatively clearer trend as compared to original wind speed TS studied in Section 3. In turn, it is expected that the forecasting performance of the approximated TS could be improved via traditional forecasting technique (e.g., ANN and SVM), because it presents much clearer trend compared to the original TS. The approximated forecasting idea is inspired by the work in Ref. [20], in which IMF1 is removed from the whole forecasting procedure and forecasting models are established for each subseries except IMF1. However, the forecasting procedures for all other subseries in Ref. [20] still face with the first challenge discussed in Section 3. Different from Ref. [20], the proposed approximated forecasting model is established for the approximated TS directly in this paper. It is noted that as illustrated in Section 3, IMF1 changes when the original TS is decomposed with newly obtained data, and such changes in IMF1 would also result in the change of the approximated TS. However, the change of approximated TS is not significant because IMF1 presents a negligibly small proportion in the original wind speed TS.

The same wind speed forecast case study discussed in Section 3 is used to illustrate the effectiveness of the proposed approximated forecasting model based on EEMD. GA-BP ANN is established to forecast the approximated wind speed TS. The results in Table 10 show that both RMSE and nRMSE of the proposed approximated model are smaller than those of Algorithm 2, Algorithm 3, and the algorithm without EMD. This study indicates that the proposed approximated forecasting model could potentially mitigate the challenge and improve the performance of EMD-based forecasting algorithms.

The proposed approximated forecasting model is also applied on the solar irradiation forecasting case study to mitigate the second challenge discussed in Section 4. The approximated similar day TS of the two days studies in Section 4 are shown in Fig. 14. As shown in Fig. 14, the main trend of the original TS is captured in the approximated TS profile, which can help avoid the second challenge proposed in Section 4. In this case study, the inputs of the approximated forecasting method include two previous approximated TS data as well as related cloud index and humidity information. The forecasting results are shown in Fig. 15.

Forecasting errors of the proposed approximated forecasting model on 5/5/2010 and 10/17/2010 are shown in Table 11, which are all smaller than those of Model 1, Model 2, and CP Model in Table 9. This indicates that the proposed approximated forecasting model could potentially mitigate the second challenge discussed in Section 4. In addition, the results also show that the proposed approximated forecasting model outperforms Model 3 on the cloudy day 10/17/2010, but presents a slightly worse performance on the blue-sky day 5/5/2010. This is because on a blue-sky day, the similar day solar irradiation TS has relatively clear trend and Model 3 could derive a forecasting model that better matches the original TS, while the proposed approximated forecasting model performs better on a cloudy day because the sharp ramps in the similar day TS are reduced and the forecasting model could be trained more efficiently.

6. Conclusion

This paper discusses two major challenges for applying the existing EMD-based forecasting methods in practical wind speed and solar radiation forecasts, which have been largely neglected in literature. Numerical case studies on practical wind speed and solar irradiation forecasts highlight that:

1. In existing EMD-based forecasting methods (i.e., Algorithms 2 and 3), sub-series of the original TS with newly obtained data are significantly different from the ones used in training the forecasting models. Our case studies in wind speed forecasting show that Algorithms 2 and 3 do not present advantages over non-decomposition based forecasting models. The degraded performance of Algorithms 2 and 3 over the persistence method indicates that existing EMD-based forecasting method is not effective in practical cases.
2. The existing EMD-based forecasting methods (Algorithms 1–3) are not able to effectively incorporate key environmental factors, and in turn may result in poor forecasting performance for TS that are usually sensitive and strongly correlated to environmental changes. In fact, the performance of the existing EMD-based forecasting models while considering environmental factors is even worse than the models without environmental factors and the non-decomposition based models, which indicates that the EMD-based forecasting model cannot effectively handle environmental factors. Our case study also shows that similar day analysis may not be effectively applied to the EMD-based method either.
3. An approximated forecasting model based on EMD is proposed as an alternative way to mitigate the above two challenges in existing EMD-based forecasting methods. In this method, approximated TS by excluding IMF1 from the original TS is directly predicted by forecasting algorithms. Our case studies show that its performance is better than the existing EMD-based forecasting methods and the non-decomposition based method.

Acknowledgment

This work was supported in part by the U.S. National Science Foundation grant ECCS-1254310.

References

[1] Soman SS, Zareipour H, Malik H, Mandal P. A review of wind power and wind speed forecasting methods with different time horizons. In: Proc. North American power symposium, Arlington, TX, USA; 2010. p. 1–8.
[2] Ma L, Luan S, Jiang C, Liu H, Zhang Y. A review on the forecasting of wind speed and generated power. Renew Sustain Energy Rev 2009;13:915–20.
[3] Mellit A, Kalogirous SA. Artificial intelligence techniques for photovoltaic application: a review. Prog Energy Combust Sci 2008;34:574–632.
[4] Hammer A, Heinemann D, Lorenz E, Luckehe B. Short-term forecasting of solar radiation: a statistic approach using satellite data. Sol Energy 1999;67:139–50.
[5] Jaesung J, Robert PB. Current status and future advances for wind speed and

power forecasting. Renew Sustain Energy Rev 2014;31:762–77.
[6] Landberg L. Short-term prediction of local wind conditions. J Wind Eng Ind Aerodynamics 2001;89:235–45.
[7] McGowin C. California regional wind energy forecasting system development. Exclusive summary, volume 1. Palo Alto, California: Electrical Power Research Institute (EPRI); Sep., 2006.
[8] Wu L, Shahidehpour M. A hybrid model for integrated day-ahead electricity price and load forecasting in smart grid. IET Generation Transm Distribution 2014;8(12):1937–50.
[9] Wu L, Shahidehpour M. A hybrid model for day-ahead price forecasting. IEEE Trans Power Syst 2010;25(3):1519–30.
[10] Marquez R, Gueorguiev VG, Coimbra CFM. Forecasting of global horizontal irradiance using sky cover indices. J Sol Energy Eng 2013;135(01–1017):1–5.
[11] Marquez R, Pedro HTC, Coimbra CFM. Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. Sol Energy 2014;92:176–88.
[12] Zhang X, Lai KK, Wang SY. A new approach for crude oil price analysis based on Empirical Mode Decomposition. Energy Econ 2008;30:905–18.
[13] Wei Y, Chen M. Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transp Res Part C 2012;21:148–62.
[14] Lei C, Ran L. Short-term wind speed forecasting model for wind farm based on Wavelet decomposition. In: Proc. Electric utility deregulation and restructuring and power technologies, Nanjing, China; 2008. p. 2525–9.
[15] Wang Y, Wang S, Zhang N. A novel wind speed forecasting method based on EEMD and GA-BP neural network. In: Proc. IEEE PES general meeting, Vancouver, Canada; 2013. p. 1–5. 2013.
[16] Ren Y, Suganthan PN, Srikanth N. A comparative study of empirical mode decomposition-based short-term wind speed forecasting method. IEEE Transaction Sustain Energy 2015;6(1):236–44.
[17] 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;94:629–36.
[18] Sun W, Liu M. Wind speed forecasting using FEEMD echo state networks with RELM in Hebei, China. Energy Convers Manag 2016;114:197–208.
[19] Liu H, Chen C, Tian HQ, Li YF. A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks. Renew Energy 2012;48:545–56.
[20] Guo Z, Zhao W, Lu H, Wang J. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. Renew Energy 2012;37:241–9.
[21] Fei SW. A hybrid model of EMD and multiple-kernel RVR algorithm for wind speed prediction. Int J Electr Power & Energy Syst 2016;78:910–5.
[22] Liu H, Tian H, Li Y. An EMD-recursive ARIMA method to predict wind speed for railway strong wind warning system. J Wind Eng Ind Aerodynamics 2015;141:27–38.
[23] Zhang C, Wei H, Zhao J, Liu T, Zhu T. Short-term wind speed forecasting using empirical mode decomposition and feature selection. Renew Energy 2016;96:727–37.
[24] Cao JC, Cao SH. Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis. Energy 2006;31:3435–45.
[25] Ghelardoni L, Ghio A, Anguita D. Energy Load forecasting using empirical mode decomposition and support vector regression. IEEE Trans Smart Grid 2013;4(1):549–56.
[26] Wang S, Wang Y, Liu Y, Zhang N. Hourly solar radiation forecasting based on EMD and ELM neural network. Electr Power Autom Equip 2014;8:7–12.
[27] Bai W, Liu ZG, Zhou DD, Wang Q. Research of the load forecasting model base on HHT and combination of ANN. In: Proc. Power and energy engineering conference, Wuhan, China; 2009. p. 1–4.
[28] Amar A, Guennoun ZEA. Contribution of wavelet transformation and empirical mode decomposition to measurement of U.S. core inflation. Appl Math Sci 2012;6(135):6739–52.
[29] Labate D, Foresta F, Occhiuto G. Empirical mode decomposition vs. wavelet decomposition for extraction of respiratory signal from single-channel ECG: a comparison. IEEE Sensors J 2013;13(7):2666–74.
[30] Arino MA. Time series forecasts via wavelets: an application to car sales in the Spanish market. Institute of Statistics & Decision Sciences, Duke University; 1995.
[31] Huang NE, Shen Z, Long SR. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc R Soc Lond 1998;454:903–95.
[32] National solar radiation data base, http://rredc.nrel.gov/solar/old_data/nsrdb/, accessed 21.03.16.
[33] Fatih OH, Omer NG, Mehmet K. Hourly solar radiation forecasting using optimal coefficient 2-D linear filters and feed-forward neural networks. Sol Energy 2008;82:714–26.
[34] Ricardo M, Carlos FM Coimbra. Proposed metric for evaluation of solar forecasting models. J Sol Energy Eng 2012;135(01–1016):1–9.