



# Exploiting sparsity of interconnections in spatio-temporal wind speed forecasting using Wavelet Transform



Akin Tascikaraoglu<sup>a,b,\*</sup>, Borhan M. Sanandaji<sup>a</sup>, Kameshwar Poola<sup>a</sup>, Pravin Varaiya<sup>a</sup>

<sup>a</sup> Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA

<sup>b</sup> Department of Electrical Engineering, Yildiz Technical University, Istanbul, Turkey

## HIGHLIGHTS

- We propose a spatio-temporal approach for wind speed forecasting.
- The method is based on a combination of Wavelet decomposition and structured-sparse recovery.
- Our analyses confirm that low-dimensional structures govern the interactions between stations.
- Our method particularly shows improvements for profiles with high ramps.
- We examine our approach on real data and illustrate its superiority over a set of benchmark models.

## ARTICLE INFO

### Article history:

Received 14 September 2015

Received in revised form 19 November 2015

Accepted 17 December 2015

Available online 8 January 2016

### Keywords:

Wind forecasting  
Compressive sensing  
Spatial correlation  
Wavelet Transform

## ABSTRACT

Integration of renewable energy resources into the power grid is essential in achieving the envisioned sustainable energy future. Stochasticity and intermittency characteristics of renewable energies, however, present challenges for integrating these resources into the existing grid in a large scale. Reliable renewable energy integration is facilitated by accurate wind forecasts. In this paper, we propose a novel wind speed forecasting method which first utilizes Wavelet Transform (WT) for decomposition of the wind speed data into more stationary components and then uses a *spatio-temporal* model on each sub-series for incorporating both temporal and spatial information. The proposed spatio-temporal forecasting approach on each sub-series is based on the assumption that there usually exists an intrinsic low-dimensional structure between time series data in a collection of meteorological stations. Our approach is inspired by Compressive Sensing (CS) and structured-sparse recovery algorithms. Based on detailed case studies, we show that the proposed approach based on exploiting the sparsity of correlations between a large set of meteorological stations and decomposing time series for higher-accuracy forecasts considerably improve the short-term forecasts compared to the temporal and spatio-temporal benchmark methods.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

### 1.1. Variable energy resources

Environmental and geopolitical concerns associated with the current power grid have motivated many countries and many states in United States to setup aggressive Renewable Portfolio Standard (RPS). Volatility, stochasticity, and intermittency characteristics of renewable energies, however, present challenges for integrating these resources into the existing grid in a large scale

as the proper functioning of an electric grid requires a continuous power balance between supply and demand [1,2]. While ancillary services such as load following and frequency regulation have been proposed to compensate for such mismatches [3,4], obtaining more precise and reliable forecasts remains as a crucial step towards renewable energy integration [5].

### 1.2. Wind speed forecasting methods

Among different renewable energy resources, wind energy is one of the most attractive energies and its development has been radically increased around the world. According to a report, total wind power capacity in the world has doubled every three years since 2000 and reached an installed capacity of 197 GW in 2010

\* Corresponding author at: Department of Electrical Engineering, Yildiz Technical University, Istanbul, Turkey. Tel.: +90 212 3835866; fax: +90 212 3835858.

E-mail addresses: [atasci@berkeley.edu](mailto:atasci@berkeley.edu), [atasci@yildiz.edu.tr](mailto:atasci@yildiz.edu.tr) (A. Tascikaraoglu).

and 369 GW in 2014 [6]. Wind power can be forecasted directly; however, forecasting wind speed and then converting the speed forecasts to power forecasts using commercially available wind turbine power curves is generally considered as a better approach [7]. This approach is more effective as, for example, wind turbines with different ratings in a wind farm might experience the same wind speed but have different wind power [8]. Therefore, we focus on wind speed forecasting in this paper.

Wind speed forecasting approaches can be classified to various categories: (i) shorter-term forecasting vs. longer-term forecasting, (ii) data-driven methods vs. model-based methods, and (iii) point forecasting vs. probabilistic forecasting. There are advantages and disadvantages associated with each approach. For instance, data-driven models are widely considered to be the most competitive methods for higher temporal resolutions due to the slightly changing wind speeds during these periods [9]. Recently, the main interest has turned from individual models to advanced combination models, which leverage the unique advantages of combining several single models in order to mitigate the prediction errors [10,11]. Also much effort has been devoted at the last few years to the wind speed forecasting methods accounting for the data collected from both the site of interest and other points (wind farms and meteorological stations) on its surrounding region over a reasonable period of time. In order to utilize both two state-of-art approaches, we aim to perform short-term point forecasting based on a two-stage model that combines a data preprocessing method, namely Wavelet Transform (WT), and a *spatio-temporal* model in this paper.

### 1.3. WT-based spatio-temporal wind speed forecasting

The main objective of data preprocessing models is to realize a preliminary process on data sets by decomposing the nonlinear time series into more stationary sub-series which are generally easier to analyze. Among various techniques, WT-based methods can be pointed out as the prevailing approach in wind forecasting due to its easy implementation and adaptive ability of time-frequency analysis [10]. The benefits of WT on the prediction performance have been proven for forecasting approaches using both linear and nonlinear models. Catalão et al. [12] present an approach which combines WT and Neural Network (NN) for short-term wind power forecasting. Several approaches are studied for wind speed forecasting which utilize WT, wavelet packet decomposition and Artificial Neural Network (ANN) [13]. Support Vector Machines (SVM), which is closely related to ANN in terms of its structure, is also used in combination with WT [14]. Liu et al. [15] also propose a model based on WT and SVM, in which the parameters of SVM are optimized using Genetic Algorithm (GA). A hybrid model consists of two-layer WT and Artificial Bee Colony (ABC) algorithm-based Relevance Vector Machine (RVM), which has a sparser representation compared to SVM, is also presented for speed forecasts [16]. In [17], WT is utilized in combination with a Feed Forward Neural Network (FFNN) model for wind power predictions with the objective of increasing the efficiency in a stand-alone system.

In this study, we first apply WT for decomposing the wind speed data into some sub-series. We then apply our proposed *spatio-temporal* wind speed forecasting method on each sub-series. We show that the overall procedure improves the prediction performance. Spatio-temporal forecasting approaches combine spatial information from different locations in a geographical area to get higher-accuracy predictions compared with predictions generated only using local data. The underlying assumption is that a considerable correlation exists between data of target meteorological station and data from the stations in its vicinity. It is expected that these correlations also appear in the wind forecasts, possibly

with some time lags depending primarily on the wind speed and distance between the sites. The potential benefits of these models and the analysis of the contribution of each input candidate on the system performance have been investigated in a number of studies. Alexiadis et al. [18] present an ANN-based spatio-temporal technique for wind speed and power forecasts up to several hours ahead. A fuzzy model using data from various stations in and around a wind farm is put forward by Damousis et al. [19] in order to benefit from autocorrelation and cross-correlation between a set of wind speed data. Gneiting et al. [20] introduce a wind speed prediction method, namely the Regime Switching Space–Time Diurnal (RSTD) model, which is based on spatial and temporal information. This approach is enhanced by Hering and Genton [21] in the Trigonometric Direction Diurnal (TDD) model, incorporating wind direction in the model. Tastu et al. [22] present different data-driven models (such as Autoregressive (AR)-based models) in order to investigate the relation between various variables and the corresponding forecast error. Similarly, multivariate AR models are used to exploit geographically dispersed wind data by Hill et al. [23]. Xie et al. [24] propose a probabilistic forecast model based on TDD model including spatial as well as the temporal correlations and geostrophic wind information. Dowell et al. [25] present an adaptive filter for wind speed and direction forecasting on the basis of spatial correlations at neighbor areas. A method with probabilistic wind generation power forecasts is developed in [26] taking the spatial information into account. Wind farm generation forecast is studied in another work [27] by applying spatio-temporal analysis to data from multiple classes of wind turbines, in which finite-state Markov chain models are obtained for short-term forecasts.

### 1.4. Our contribution

We propose a novel spatio-temporal approach for wind speed forecasting. Our contributions are as follows. We first show that there exist low-dimensional structures governing the interactions among a set of meteorological stations, and exploiting such structures significantly enhances forecasting performance. We formulate the forecasting task as a linear inverse problem. We then use tools from structured-sparse recovery to find a *block-sparse* solution. The proposed approach dynamically determines the wind speed data set to be used in the forecast of the target station for the following prediction horizon. In other words, the proposed algorithm determines the contribution of each meteorological station on the next forecast by assigning them coefficients proportional to their similarities with the recent data of the target station. Contrary to the widely-used conventional time series methods considering predefined structures in which the order of each meteorological station is determined once and then kept constant during all of the prediction periods, the proposed adaptive algorithm provides an up-to-date input set for each prediction horizon without realizing the time-consuming order selection process. Thus, the model structure of the proposed algorithm helps improve the prediction accuracy, especially for longer prediction times including daily, seasonal, and annual cycles in weather conditions.

Second, we show that combining the WT with the proposed spatio-temporal method further improves the forecasting accuracy, particularly for wind profiles with high up and down ramps. Furthermore, the indicated improvements are accomplished using a smaller training data set for the proposed algorithm while usually larger data sets are included in other benchmark methods considered in this paper for obtaining reasonable results. Lastly, our analysis shows that the proposed method results in highly concentrated prediction error values which further facilitate the applicability of the forecasts in energy market commitment and dispatch.

### 1.5. Paper organization

In Section 2, forecasting problem is formulated and the proposed algorithms are presented in Section 3 with their related concepts. In Section 4, the proposed method is applied on real wind speed data and the results of a comparison against a set of benchmark methods are provided. We provide some additional information about the methods in Section 5. Section 6 summarizes our results and propose future directions.

## 2. Modeling framework

One can represent the output of a system as a weighted linear combination of its past observations. This representation is often referred to an AR model. This framework can be extended to a more general setting in which the output variable is a  $P$ -dimensional vector. The so-called Multivariate Autoregressive (M-AR) model of order  $n$  can be given as

$$\underbrace{\begin{bmatrix} y_{n+1}^i \\ y_{n+2}^i \\ \vdots \\ y_{n+M}^i \end{bmatrix}}_{\mathbf{b} \in \mathbb{R}^M} = \underbrace{\begin{bmatrix} y_n^1 & \cdots & y_1^1 & \cdots & y_n^P & \cdots & y_1^P \\ y_{n+1}^1 & \ddots & \vdots & \cdots & y_{n+1}^P & \ddots & \vdots \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ y_{n+M-1}^1 & \cdots & y_M^1 & \cdots & y_{n+M-1}^P & \cdots & y_M^P \end{bmatrix}}_{A \in \mathbb{R}^{M \times N}} \underbrace{\begin{bmatrix} X_1^{\text{tr}}(1, i) \\ \vdots \\ X_n^{\text{tr}}(1, i) \\ \vdots \\ X_1^{\text{tr}}(P, i) \\ \vdots \\ X_n^{\text{tr}}(P, i) \end{bmatrix}}_{\mathbf{x} \in \mathbb{R}^N} \quad (2)$$

Block 1  
Block P

$$\begin{aligned} \mathbf{y}(t) &= X_1 \mathbf{y}(t-1) + \cdots + X_n \mathbf{y}(t-n) + \mathbf{e}(t) \\ &= \sum_{j=1}^n X_j \mathbf{y}(t-j) + \mathbf{e}(t), \end{aligned} \quad (1)$$

where  $X_j \in \mathbb{R}^{P \times P}$  is a coefficient matrix for the  $j$ -th time lag and  $\mathbf{e}(t)$  is the noise component. Let  $y_t^i$  be the wind speed of the  $i$ -th meteorological station at time  $t$  ( $t = 1, 2, \dots, M+n$ ). The M-AR model (1) for each station can be recast in a format as given in (2), where  $N := nP$ .

In the training phase, our task is to recover a vector  $\mathbf{x} \in \mathbb{R}^N$  that best explains the measurements  $\mathbf{b} \in \mathbb{R}^M$  and  $A \in \mathbb{R}^{M \times N}$ . As can be seen from (2),  $\mathbf{x}$  has a block structure in which the coefficients belonging to each station appear in one vector-block.

## 3. Compressive spatio-temporal wind speed forecasting based on Wavelet Transform

The proposed prediction model consists of three stages, as shown in Fig. 1. In the first stage, the original wind speed time series of the target location is decomposed by WT into a certain number of more regular components which have different frequency bands. In the study, a second-level Wavelet decomposition is

employed, therefore three subsets are extracted from the original series: second level approximation component ( $A_2$ ), and first and second level detail components ( $D_1$  and  $D_2$ , respectively). Then the next values of each sub-series are separately forecasted using Compressive Spatio-Temporal Wind Speed Forecasting (CST-WSF) model in which one of the components from WT and the data from the other stations excepting the target station are used as inputs. It is also to be noted that a forecasted sub-series is obtained for each extracted component using a CST-WSF model having different model parameters (i.e., order values), which are determined regarding the prediction horizon considered and the profile of component type applied to the model as input. For instance, lower order levels are generally selected for the relatively short prediction horizons and relatively stable component profiles. Finally, the single component forecasts are aggregated to build the original wind speed time series forecast.

In order to forecast each sub-series, we claim that only a few meteorological stations have a high correlation with the target station among a large

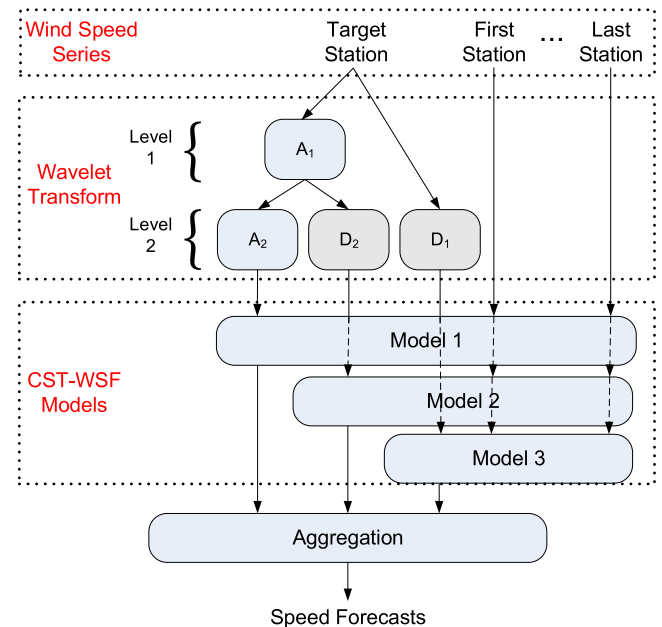


Fig. 1. General scheme of the proposed forecasting model.

number of stations. Under the assumption of *sparsity* of the inter-connections, we show that  $\mathbf{x}$  contains a particular format. In fact, the non-zero entries of the recovered solution appear in a limited number of block locations, chosen based on their contribution to the output of the target station. These vectors are called *block-sparse*. For a particular meteorological station, we then solve the augmented minimization as:

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 \quad \text{subject to} \quad (\mathbf{x} \text{ is block-sparse}). \quad (3)$$

We namely call this algorithm CST-WSF. As mentioned before, the final forecasts of the original series is obtained by aggregating the predictions of each sub-series. Different approaches employed in our study are briefly explained in the following subsections.

### 3.1. Wavelet Transform

WT decomposes a time series into different data sub-series at different frequency levels: one approximation subset which represents the general trend of the signal and multiple detail subsets which include more volatile fluctuations on different levels, as shown in Fig. 1. These subsets exhibit relatively more stationary behavior than the original time series, which facilitates the prediction task and thereby lowers the overall prediction errors.

WT can be divided into two classes, namely, Continuous Wavelet Decomposition (CWT) and Discrete Wavelet Decomposition (DWT) in which the wavelets are discretely sampled. In the literature of wind forecasting, DWT is generally preferred due to its efficiency in providing the appropriate information with a more reasonable computational time compared to CWT. Besides, the number of decomposition level, which is another essential parameter in WT, can be determined by comparing the similarity between the approximation level and the original wind speed time series. Considering more decomposition levels results in more stationary subsets; however, high number of levels might cause decomposition information loss and thus large prediction errors in forecasting. Therefore, the number of decomposition level is mostly limited to a small number ranging from 2 to 4 in wind speed and power forecasting studies [16,17,28,29]. Selection of the mother wavelet has also a considerable effect on the profile of the decomposed levels, hence on the prediction performance. The Daubechies wavelet of order 4 (Db4) is proposed in the wind forecasting literature as the most suitable type of mother wavelet, due to its ability of providing a balance between wavelength and smoothness [30].

Considering the above-mentioned characteristics of WT, in our study, a second-level DWT, which generates one approximation (A2), and two detail components (D1 and D2), is used with the mother wavelet of Db4. We particularly analyzed the effects of these parameters on the forecasting performance of the proposed models for the wind speed data considered in the case studies. For detailed explanations about the method and parameter selection, please see [12,14].

### 3.2. Background on Compressive Sensing (CS)

CS is a recent paradigm in signal processing which allows recovery of a signal from its underdetermined measurement set under given conditions on the matrix  $\mathbf{A}$  [31] and under the assumption of signal sparsity. CS recovery algorithms provide tools to recover a  $K$ -sparse signal  $\mathbf{x} \in \mathbb{R}^N$  from its measurements  $\mathbf{b} = \mathbf{A}\mathbf{x} \in \mathbb{R}^M$  where  $\mathbf{A} \in \mathbb{R}^{M \times N}$  is the measurement matrix with  $M < N$ . A  $K$ -sparse signal  $\mathbf{x} \in \mathbb{R}^N$  is a signal of length  $N$  with  $K$  non-zero entries where  $K < N$ . From basic linear algebra, one knows that there are infinitely many possible solutions to the inverse problem  $\mathbf{b} = \mathbf{A}\mathbf{x}$ . If  $\mathbf{A}$  satisfies some

conditions, then CS is guaranteed to find the sparsest solution among others.

Various recovery guarantees have been presented in the literature of CS and these methods can be categorized into two classes, namely convex optimization algorithms such as Basis Pursuit (BP) [32] and greedy algorithms such as Orthogonal Matching Pursuit (OMP) [33] and Compressive Sampling Matching Pursuit (CoSaMP) [34].

### 3.3. Uniform CST-WSF

As mentioned earlier, we use CS tools for a block-sparse recovery:

**Definition 1** (*Block  $K$ -Sparse Signal*). Let  $\mathbf{x} \in \mathbb{R}^N$  be a concatenation of  $P$  vector-blocks  $\mathbf{x}_i \in \mathbb{R}^n$ , i.e.,

$$\mathbf{x} = [\mathbf{x}_1^{\text{tr}} \cdots \mathbf{x}_i^{\text{tr}} \cdots \mathbf{x}_P^{\text{tr}}]^{\text{tr}},$$

where  $N = nP$ . A signal  $\mathbf{x} \in \mathbb{R}^N$  is called *block  $K$ -sparse* if it has  $K < P$  non-zero blocks.  $\square$

Over the past years, several CS algorithms have been proposed for dealing with additional structures in the sparse signal to be recovered. In our work, we use Block Orthogonal Matching Pursuit (BOMP) algorithm [35,36] for block sparse recovery [37]. In order to find a block-sparse solution to the equation  $\mathbf{b} = \mathbf{A}\mathbf{x}$ , the BOMP steps are given in Algorithm 1, where  $\mathbf{A} \in \mathbb{R}^{M \times N}$  is a concatenation of  $P$  matrix-blocks  $\mathbf{A}_i \in \mathbb{R}^{M \times n}$  ( $\forall n_i = n$ ) as

$$\mathbf{A} = [\mathbf{A}_1 \cdots \mathbf{A}_i \cdots \mathbf{A}_P].$$

**Algorithm 1.** The BOMP algorithm for block-sparse recovery.

---

**Require:**  $\mathbf{A}, \mathbf{b}, \{n_i\}_{i=1}^P$ , stopping criteria  
**Ensure:**  $\mathbf{r}^0 = \mathbf{b}$ ,  $\mathbf{x}^0 = \mathbf{0}$ ,  $\mathcal{A}^0 = \emptyset$ ,  $l = 0$   
**repeat**  
     1. **match:**  $\mathbf{h}_i = \mathbf{A}_i^{\text{tr}} \mathbf{r}^l$ ,  $i = 1, 2, \dots, P$   
     2. **identify support:**  $\lambda = \arg \max_i \|\mathbf{h}_i\|_2$   
     3. **update the support:**  $\mathcal{A}^{l+1} = \mathcal{A}^l \cup \lambda$   
     4. **update signal estimate:**  
          $\mathbf{x}^{l+1} = \arg \min_{\mathbf{s}: \text{supp}(\mathbf{s}) \subseteq \mathcal{A}^{l+1}} \|\mathbf{b} - \mathbf{A}\mathbf{s}\|_2$ ,  
         where  $\text{supp}(\mathbf{s})$  indicates the blocks  
         on which  $\mathbf{s}$  may be non-zero  
     5. **update residual estimate:**  $\mathbf{r}^{l+1} = \mathbf{b} - \mathbf{A}\mathbf{x}^{l+1}$   
     6. **increase index  $l$  by 1**  
**until** stopping criteria true  
**output:**  $\hat{\mathbf{x}} = \mathbf{x}^l = \arg \min_{\mathbf{s}: \text{supp}(\mathbf{s}) \subseteq \mathcal{A}^l} \|\mathbf{b} - \mathbf{A}\mathbf{s}\|_2$

---

Due to the block sparsity of  $\mathbf{x}$ , the observation vector  $\mathbf{b}$  can be given as a combination of the columns of  $\mathbf{A}$ , with the selections of columns occurring in clusters due to the block structure of the sparsity pattern in  $\mathbf{x}$ . BOMP attempts to identify the participating indices by correlating the measurements  $\mathbf{b}$  against the columns of  $\mathbf{A}$  and comparing the correlations between different blocks. Once a considerable block has been found, its influence is removed from the measurements  $\mathbf{b}$  via an orthogonal projection, and the correlation values are recalculated for the rest of the blocks. This process repeats until the residual equals zero [37]. BOMP and its recovery conditions have been investigated in topology identification and causal inference of interconnected networks [37].

### 3.4. Nonuniform CST-WSF

One can generalize the uniform CST-WSF model (2) to the case where the interconnection links between any two stations are AR models with different orders. This model structure distinguishes between the stations with low and high correlations with the target station. Let  $n_i$  be the order for the  $i$ -th station ( $i = 1, 2, \dots, P$ ). An Nonuniform Multivariate Autoregressive (NM-AR) version of (2) can be considered as given in (4), where  $n_{\max} \geq \max_i n_i$  and  $N := \sum_{i=1}^P n_i$ . This model structure results in a nonuniform block-sparse vector  $\mathbf{x}$  with blocks having different length.

$$\underbrace{\begin{bmatrix} y_{n_{\max}+1}^i \\ y_{n_{\max}+2}^i \\ \vdots \\ y_{n_{\max}+M}^i \end{bmatrix}}_{\mathbf{b} \in \mathbb{R}^M} = \underbrace{\begin{bmatrix} y_{n_{\max}}^1 & \cdots & \cdots & \cdots & y_{n_{\max}-n_1+1}^1 & \cdots & y_{n_{\max}}^P & \cdots & y_{n_{\max}-n_P+1}^P \\ y_{n_{\max}+1}^1 & \ddots & \ddots & \ddots & \vdots & \cdots & y_{n_{\max}+1}^P & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ y_{n_{\max}+M-1}^1 & \cdots & \cdots & \cdots & y_{n_{\max}-n_1+M}^1 & \cdots & y_{n_{\max}+M-1}^P & \cdots & y_{n_{\max}-n_P+M}^P \end{bmatrix}}_{\mathbf{A} \in \mathbb{R}^{M \times N}} \times \underbrace{\begin{bmatrix} X_1^{\text{tr}}(1, i) \\ \vdots \\ \vdots \\ X_{n_1}^{\text{tr}}(1, i) \\ \vdots \\ X_1^{\text{tr}}(P, i) \\ \vdots \\ X_{n_P}^{\text{tr}}(P, i) \end{bmatrix}}_{\mathbf{x} \in \mathbb{R}^N} \quad (4)$$

} Block 1  
} Block P

**Definition 2** (Nonuniform Block  $K$ -Sparse Signal). Let  $\mathbf{x} \in \mathbb{R}^N$  as a concatenation of  $P$  vector-blocks  $\mathbf{x}_i \in \mathbb{R}^{n_i}$  where  $N = \sum_{i=1}^P n_i$ . A signal  $\mathbf{x} \in \mathbb{R}^N$  is called *nonuniform block  $K$ -sparse* if it has  $K < P$  non-zero blocks.  $\square$

**Remark 1.** A nonuniform block  $K$ -sparse signal is a generalization of a block  $K$ -sparse signal (Definition 1) whose blocks have different length, i.e.,  $n_i = n$ ,  $\forall i$ .  $\square$

Given  $\{n_i\}_{i=1}^P$ , the BOMP Algorithm 1 can be used for recovery of  $\mathbf{x}$  with  $\mathbf{A}_i \in \mathbb{R}^{M \times n_i}$ . So as to find the order set  $\{n_i\}_{i=1}^P$ , we apply auto- and cross-correlation analyses. The orders are then adjusted for the best forecasting performance.

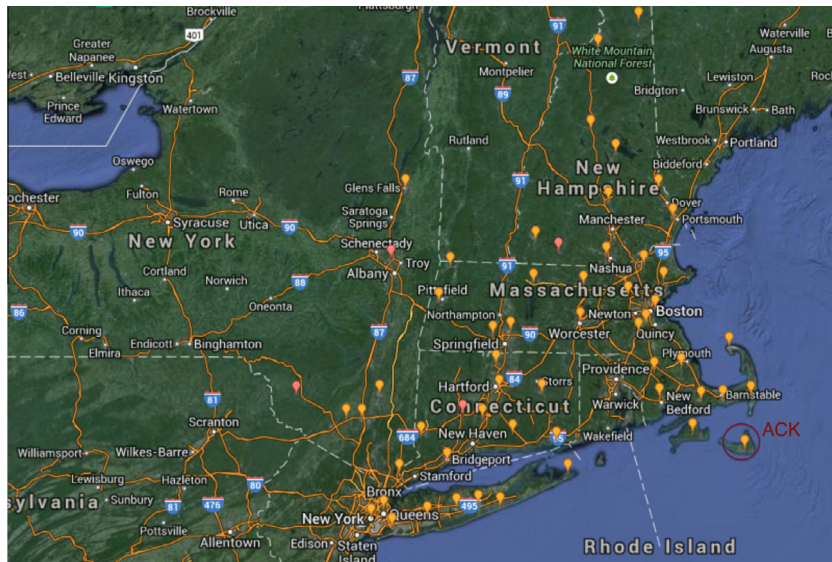
## 4. Case study of 57 meteorological stations in United States

In this section, we apply our CST-WSF approaches together with WT to real wind speed data of an area in the east coast of the United States. These states generally face wind profiles with high speed all over the year and also, it is possible to find regions in these states containing a large number of stations close to each other.

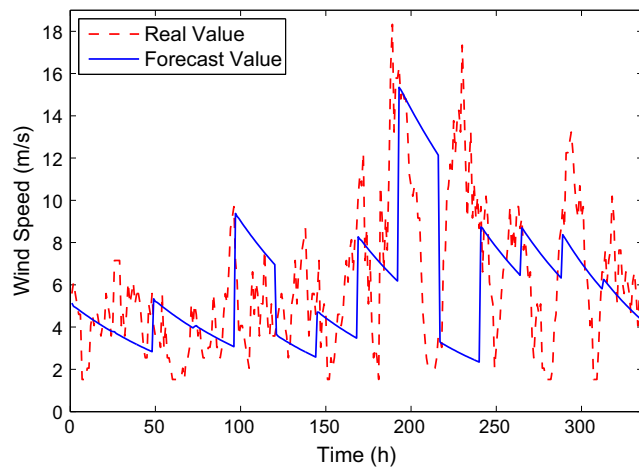
### 4.1. Data description

Hourly wind speed data from a combination of weather stations and airports (Meteorological Terminal Aviation Routine (METAR) data) in four states of United States is used in the case study. Fig. 2 illustrates the area including the states Massachusetts, Connecticut, New York and New Hampshire [38], and the location of 57 stations from which the measurements are taken. A station (circled in red) located on an island, namely Nantucket Memorial Airport (ACK), is chosen as target station due to its relatively higher wind speed profile and its good cross-correlations with other stations (probably caused by the prevailing wind directions of north-west and southeast).

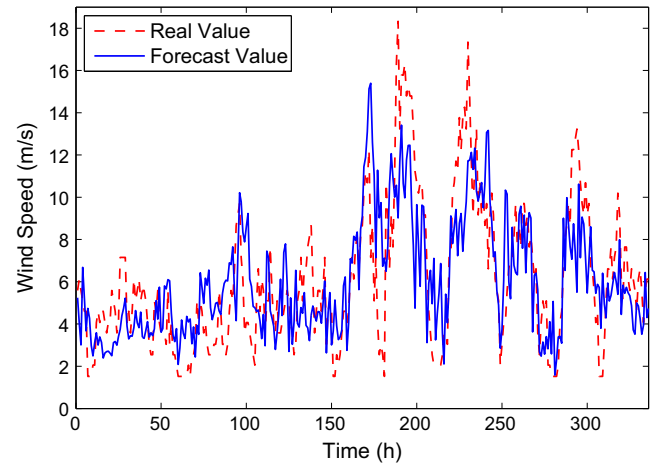




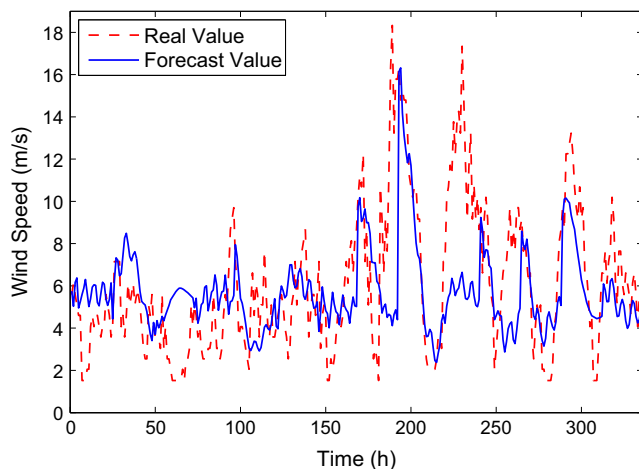
**Fig. 2.** Map of the area under study. The 57 measuring locations in east coast are shown with yellow points. Circled in red is the target station ACK. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



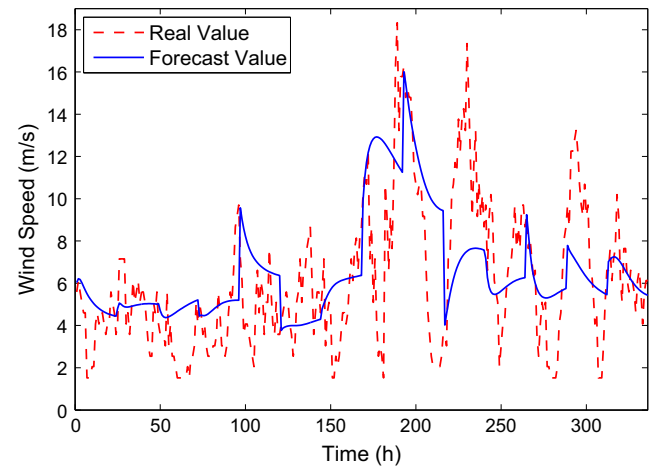
(a) AR model of order 3



(a) Spatio-temporal ANN model



(b) WT-ANN model



(b) Spatio-temporal LS M-AR model

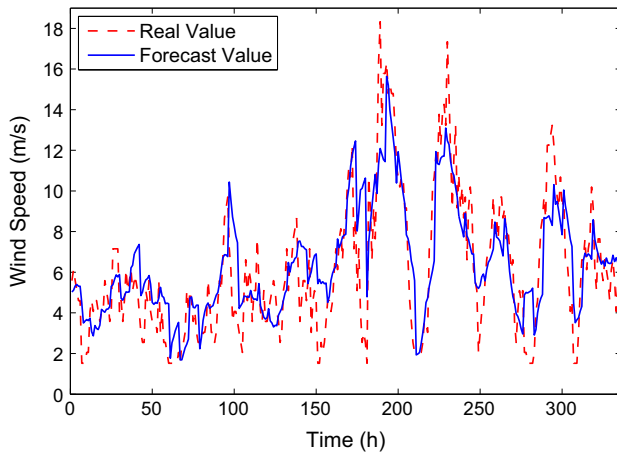
**Fig. 3.** Comparison of different temporal forecasting algorithms for 24 h.

**Fig. 4.** Comparison of different spatio-temporal forecasting algorithms for 24 h.

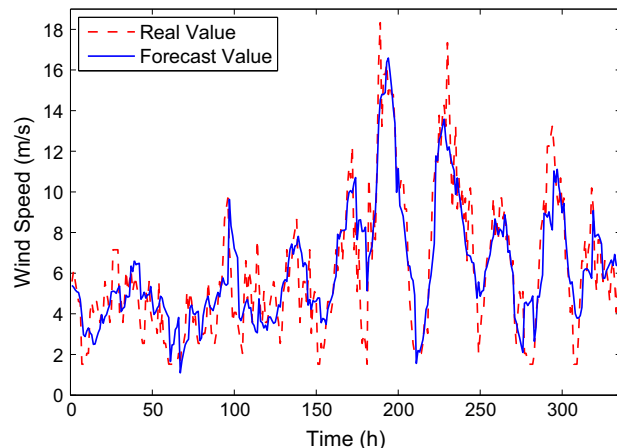
The simulations are performed for an approximately three-month period in 2014 during which data is available for all stations. Various performance metrics are calculated for the proposed model as well as several benchmark methods are provided in related tables for this three-month period; however, only the results associated with a two-week period from February 6, 2014 to February 20, 2014 are presented in the figures in order to avoid data redundancy. This two-week period has the most irregular wind profiles in the data set used in the study and thus shows the contribution of the proposed model more clearly. Note that a period of 14 days (for instance, the data from January 23, 2014 to February 6, 2014 for the two-week test period selected for the illustration of the model performances) is used for the proposed models as the training data set while a period of 30 days is used for the AR model and ANN-based benchmark models (WT-ANN and ANN-based spatio-temporal) due to their model structure dependency on relatively large data sets for having effective forecasts.

#### 4.2. Comparison with other benchmark algorithms

We compare CST-WSF based on WT with temporal and spatio-temporal wind speed forecasting algorithms for two different prediction horizons, 6-h and 24-h. For temporal forecasting, we first

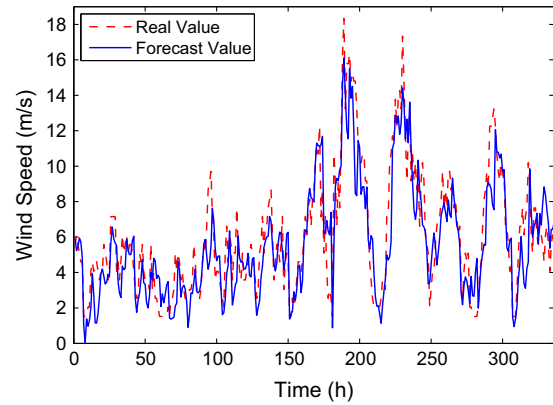


(a) Uniform CST-WSF with equal orders of 3 for 6 hours.

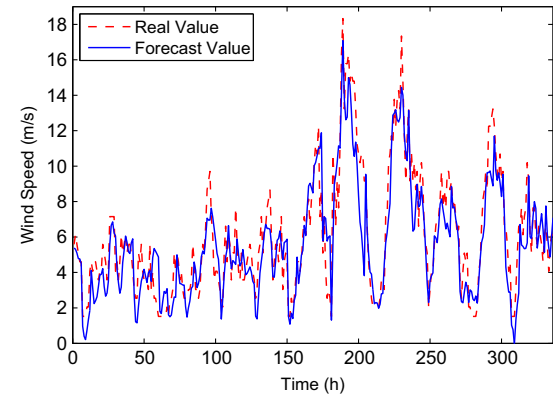


(b) Nonuniform CST-WSF with nonuniform orders for 6 hours.

Fig. 5. Comparison of the uniform and nonuniform CST-WSF.



(a) Uniform WT-CST-WSF with equal orders of 4,5 and 5, respectively for each sub-series for 6 hours.



(b) Nonuniform WT-CST-WSF with nonuniform orders for 6 hours.

Fig. 6. Comparison of the uniform and nonuniform WT-CST-WSF for 6 h.

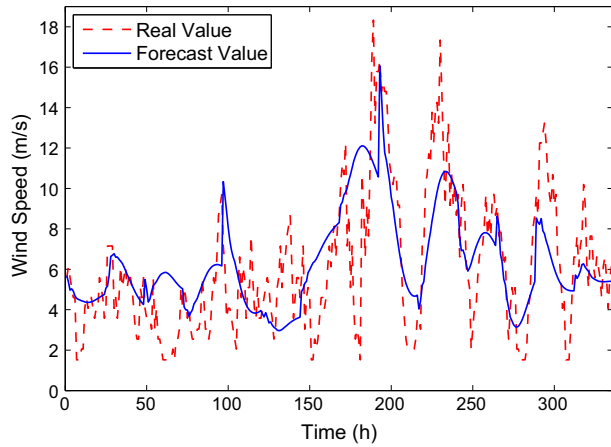
Table 1

Statistical error measure comparison of different methods for 6-h predictions.

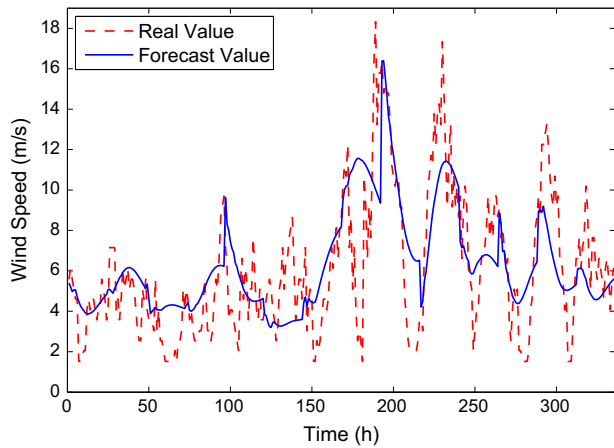
Forecasting approach	MAE (m/s)	RMSE (m/s)	NRMSE (%)
AR of order 3	2.0711	2.7602	16.43
WT-ANN	1.8287	2.4713	14.71
ANN-based ST	1.7996	2.3083	13.74
LS-based ST	1.7209	2.1991	13.09
CST-WSF of order 3	1.5705	2.0647	12.29
Nonuniform CST-WSF	1.3414	1.7522	10.43
WT-CST-WSF	1.3186	1.7674	10.52
Nonuniform WT-CST-WSF	1.1867	1.5809	9.41

consider AR model, which is the fundamental model in traditional time series models [39]. Then, we consider an advanced prediction method, which integrates WT with ANN to capture the existing nonlinearity in time series of wind speed. In this method, wind speed series is decomposed to a set of sub-series by WT and each sub-series is estimated separately employing an ANN model (see [17] for more details). Fig. 3 shows the results of temporal methods in which a multi-step recursive approach is adopted for a time period of 14 days with 24-h updates.

As can be seen, the considered temporal ANN-based method provides reasonable forecasting compared to the AR model. In addition to temporal methods, we employ two spatio-temporal methods in our comparison. We consider a spatio-temporal model based on NN [40] and also an M-AR approach using Least Squares (LS) [22]. Fig. 4 shows the superiority of incorporation of spatial information in forecasting as compared to the methods using only temporal data.



(a) Uniform CST-WSF with equal orders of 4.



(b) Nonuniform CST-WSF with nonuniform orders.

**Fig. 7.** Comparison of the uniform and nonuniform CST-WSF for 24 h.

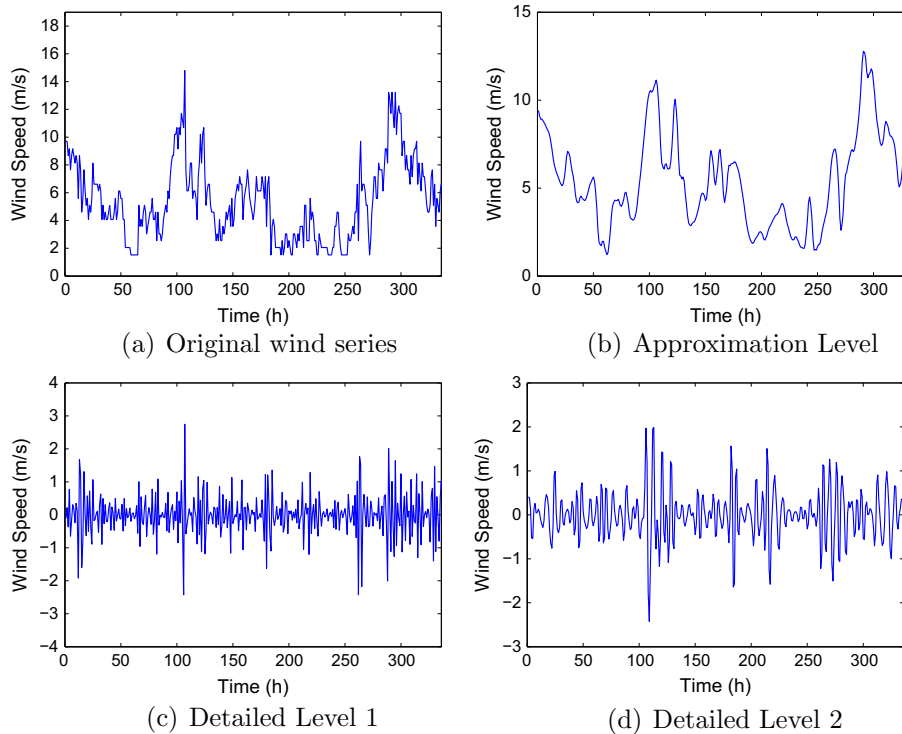
We only report the 24-h ahead predictions of the benchmark models for the sake of saving space. In fact, 24-h predictions are of more importance for power system operations and more appropriate for showing the efficiency of a wind forecasting model [5]. For detailed information about the 6-h predictions see [41].

#### 4.3. Uniform and nonuniform CST-WSF based on WT

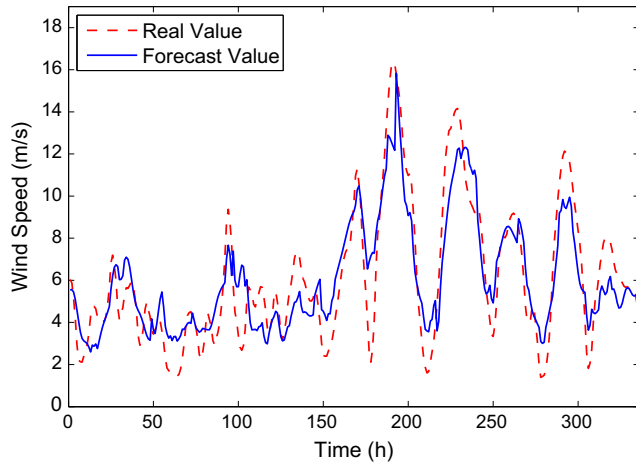
The proposed CST-WSF approaches are applied in this section. A coefficient vector  $\mathbf{x}$  is calculated every 6 or 24 h in the simulations. In other words,  $\mathbf{x}$  is updated every 6 or 24 time steps (each of which is 1 h) based on new measurements. Also, a recursive process, in which the forecast of all stations at time  $n + M + 1$  ( $\hat{y}_{n+M+1}^i, \forall i$ ) is incorporated to the matrix  $A$  for the subsequent forecasts at time  $n + M + 2$  ( $\hat{y}_{n+M+2}^i, \forall i$ ), is employed in wind speed predictions and this process continues through the prediction horizon. The matrix  $A$  is then updated for another 6 or 24 steps with the real data measured during the previous prediction horizon.

The forecasting results for the uniform CST-WSF model and for the nonuniform CST-WSF model for the 6-h case are shown in Fig. 5a and b, respectively. The results are superior to the benchmark methods explained in the previous section. Lastly, the uniform and nonuniform CST-WSF algorithms together with WT method are applied, as shown in Fig. 6.

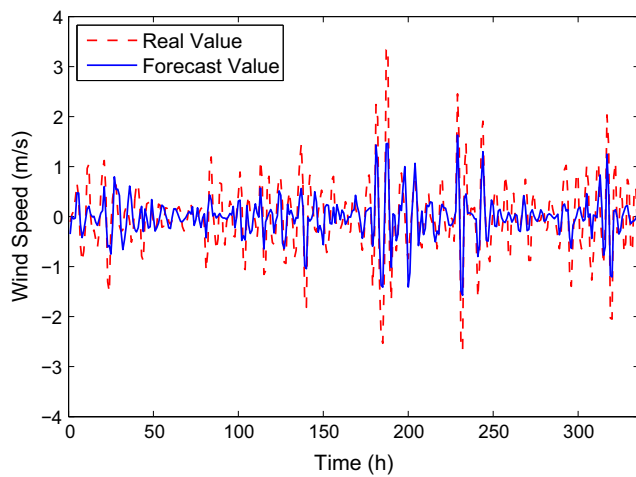
With the objective of better observing the performance of the proposed methods for 6-h predictions, the associated Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Normalized Root Mean Squared Error (NRMSE) are compiled in Table 1 for models explained in this paper for the whole simulation period, i.e., nearly three months. As can be seen, the proposed algorithms outperforms the temporal and spatio-temporal models. Besides, incorporation of the WT leads to an important decrease in error metrics. Considering the NRMSE, nonuniform CST-WSF based on WT achieves a reduction of 36.03% and 28.11% compared to the WT-ANN and LS-based spatio-temporal models, respectively.

**Fig. 8.** Wind speed signal and its approximation and detail components in the training data set.

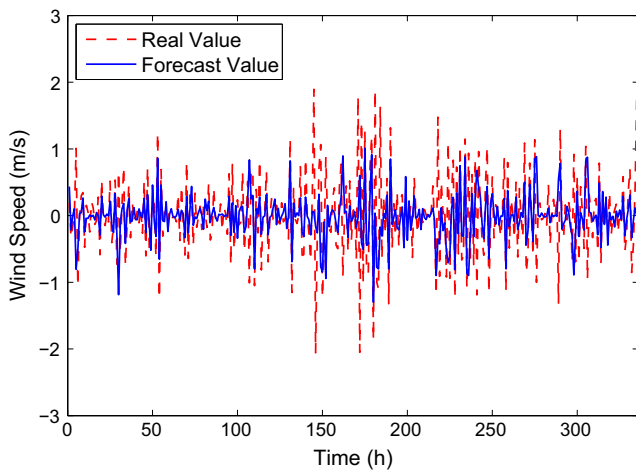




(a) Approximation Level



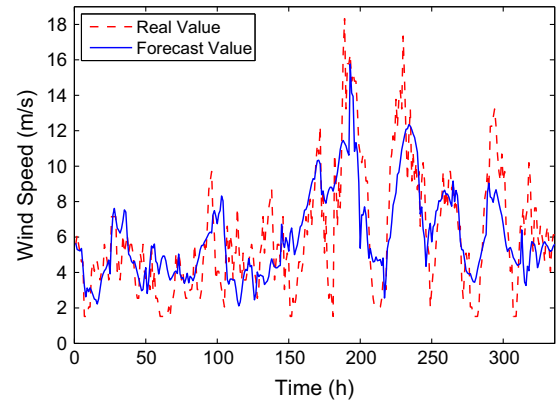
(b) Detailed Level 1



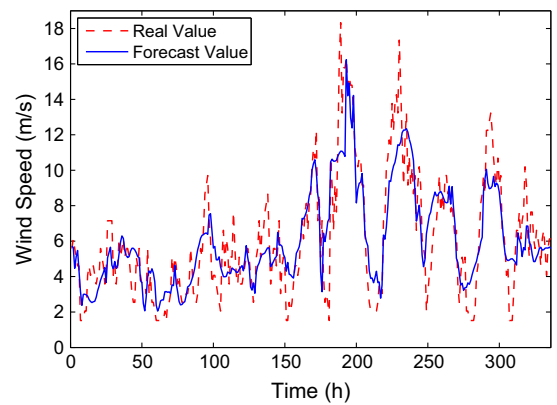
(c) Detailed Level 2

**Fig. 9.** Wind speed signal and its approximation and detail components in the test data set.

The simulation studies are repeated for examining the performance of the proposed models in longer prediction horizons. The results of both CST-WSF methods are given in Fig. 7, respectively. As mentioned in the previous section, the uniform and nonuniform CST-WSF approaches are applied to the decomposed wind speed series of the target station to increase the prediction performance.

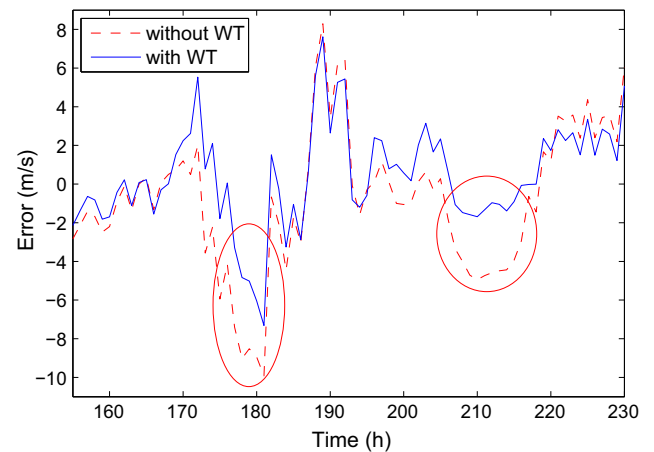


(a) Uniform WT-CST-WSF with equal orders of 6, 7, and 7, respectively for each sub-series.



(b) Nonuniform WT-CST-WSF with nonuniform orders.

**Fig. 10.** Comparison of the uniform and nonuniform WT-CST-WSF for 24 h.

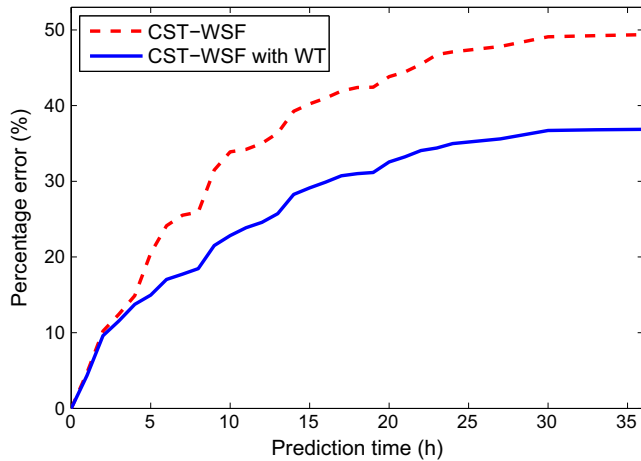


**Fig. 11.** Predictions errors for nonuniform WT-CST-WSF for 24 h. The areas circled in red show the high ramps. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

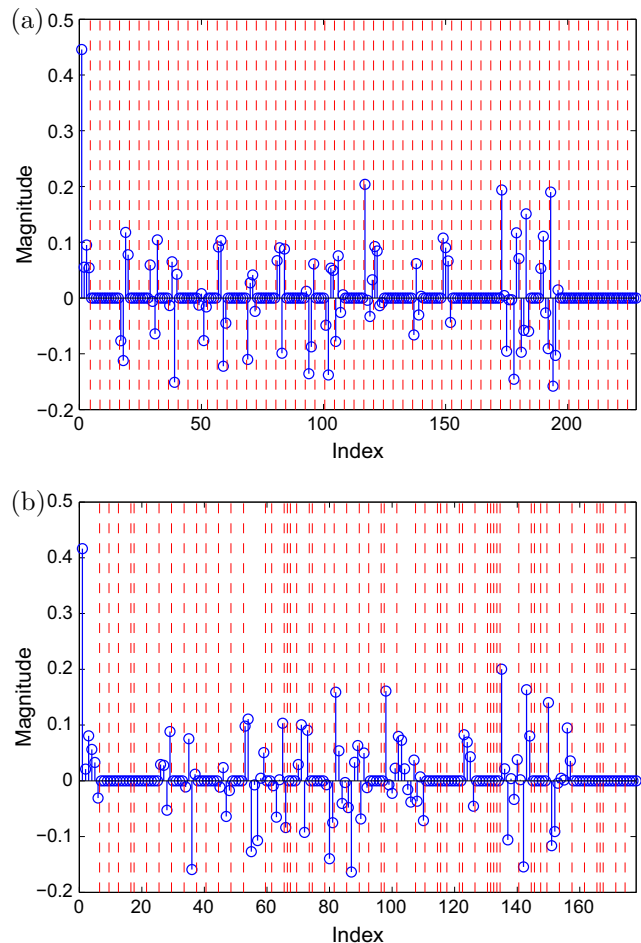
For this purpose, first, the decomposed levels of the wind speed data used in the training stage are derived, as shown in Fig. 8. The future values of the Wavelet components are then forecasted separately using modified CST-WSF algorithms. Fig. 9 shows the forecasted values for each Wavelet components ( $A_2$ ,  $D_1$  and  $D_2$ ) during the two-week test period. As can be seen from Fig. 9a, the forecast accuracy is significantly high due to the relatively stable

**Table 2**  
Statistical error measure comparison of different methods for 24-h predictions.

Forecasting approach	MAE (m/s)	RMSE (m/s)	NRMSE (%)
AR of order 3	2.9591	4.0522	24.12
WT-ANN	2.2417	3.0307	18.04
ANN-based ST	2.0517	2.6191	15.59
LS-based ST	2.2285	3.0979	18.44
CST-WSF of order 4	2.0101	2.6191	15.59
Nonuniform CST-WSF	1.9106	2.5267	15.04
WT-CST-WSF	1.8393	2.3218	13.82
Nonuniform WT-CST-WSF	1.5384	1.9841	11.81



**Fig. 12.** Percentage error vs. prediction time up to 36 h.



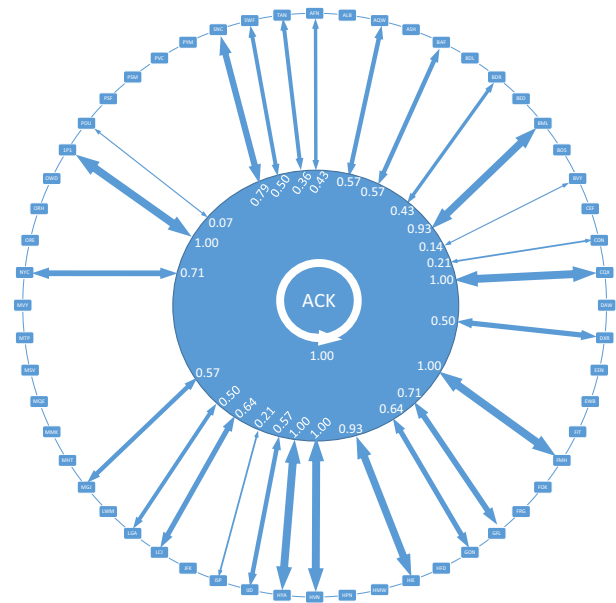
**Fig. 13.** Block-sparse coefficient vector. (a) Uniform CST-WSF of order 4. (b) Nonuniform CST-WSF with different orders.

profile of  $A_2$  component, as explained before. On the other hand, the sub-series shown in Fig. 9b and c have a volatile characteristic since these high-frequency time series represent the sudden changes in wind speed, which makes it difficult to achieve reasonable results. However, the erroneous wind speed forecasts obtained for these components do not cause a considerable damage to the total prediction accuracy because of their generally low wind speed magnitude compared to that of  $A_2$  component. Finally the predictions of the uniform and nonuniform CST-WSF approaches based on WT are given in Fig. 10.

As seen from Fig. 10, sub-series with lower fluctuations are predicted with a lower prediction error. WT method improves the

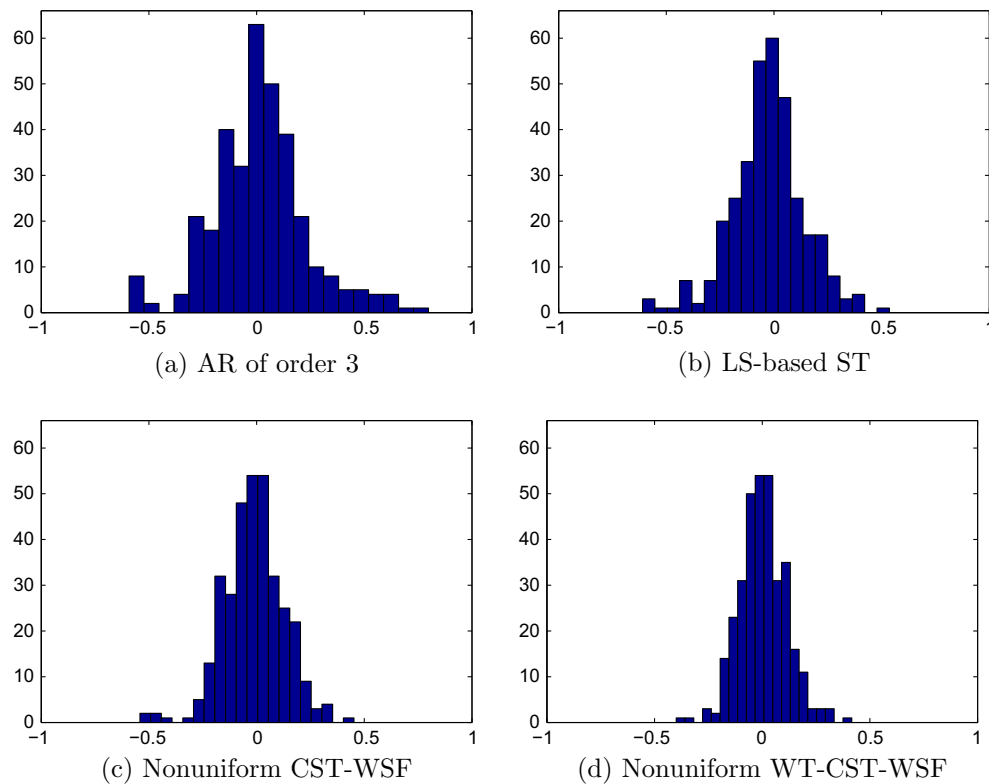


(a) Correlations for the period from February 6, 2014 to February 20, 2014.



(b) Correlations for the period from February 20, 2014 to March 6, 2014.

**Fig. 14.** The correlation levels between the target site and other sites for two consecutive 14-day periods.



**Fig. 15.** Comparison of prediction error histograms of different forecasting algorithms.

error particularly in areas with high ramp. This is illustrated in Fig. 11 by zooming in the periods in which the improvements are more distinctive. Decomposition of the wind speed data enables us to predict each of the sub-series with different model orders depending on the characteristics of the data. As a result, an improved performance is achieved in forecasting the original time series. The associated performance metrics are listed in Table 2 for the whole simulation period.

The proposed models are even more effective for longer prediction horizons, as seen from Table 2. In terms of NRMSE values, nonuniform CST-WSF approach based on WT provides an improvement of 28.11% for 6-h and 35.95% for 24-h predictions compared to the LS-based spatio-temporal model.

Apart from the capability of capturing high ramps, the other key benefit of WT method is that it helps to decrease prediction errors for longer prediction horizons [10]. Fig. 12 shows the variation of the percentage errors as a function of forecast horizon up to 36 h. As can be seen, there is a relatively linear relation between the percentage error and time for shorter prediction time horizons for both of the models. The error percentages become more steady for the longer prediction horizons, particularly with the method using WT. In other words, WT allows CST-WSF approach to extend its prediction horizon effectively by revealing the long-term trend in wind speed series.

For further analysis, Fig. 13 shows the block-sparse coefficient vectors for both CST-WSF models. In Fig. 13, only a few of the blocks in both CST-WSF approaches are non-zero. These blocks represent the most correlated stations with the target station and it can be pointed out that these stations are commonly the same for both cases. The correlated stations are substantially associated with the wind speed, wind direction, distance to the target station, and geographical layout of the region. Therefore, the correlation level of each station is changed continuously over time.

Fig. 14 shows the normalized correlation values for each station during two consecutive 14-day periods. A higher score indicates

that there is a relatively high correlation between the target site and the corresponding site for most of the prediction times (i.e. 24 h) in the given period. Fig. 14 indicates two important points: (i) only some stations are dominant in each 14 days and (ii) the dominant structure is changing over time, even for the consecutive periods.

Lastly, Fig. 15 shows the distribution of the predictions errors over the test period. The highly concentrated error values obtained with the proposed models indicate less requirement for regulating reserve capacity, which can be pointed out as another benefit of the proposed models.

## 5. Discussion

In this paper we propose a CST-WSF method in order to accomplish high-accuracy forecasts and combined this approach with WT to improve the forecasting performance, particularly for longer prediction horizons due to their importance in power system operations such as unit commitment and scheduling. Combining various models, in fact, might cause the wind speed forecasting to be more complicated. However, this potential drawback can be compromised and justified considering the high performance of the combined models over the individual models (please see [10] for further information and a large number of examples).

Besides, we choose a deterministic approach in our forecasting framework. One might take a probabilistic approach towards forecasting where statistical testing should be performed to evaluate the forecasting algorithm. As our approach is deterministic and we are not assuming any probability distribution on the wind speed time series data, we can only compare our approach to other approaches that follow a deterministic framework. The metrics based on which we decide our algorithm is superior to other considered algorithms are the ones listed in Tables 1 and 2.

It is worth noting that our spatio-temporal approach can be used for wind power forecasting, with only minor modifications.

In our paper, one of the reasons of choosing forecasts of wind speed is the data availability. In order to emphasize the contribution of spatial data on forecast performance, a larger data set measured within the same area for a reasonable period is required; however, finding such a data set including power information of many wind power plants located in the same area is generally not possible. Furthermore, in the case of power data, the forecasts might be sometimes complex if wind farm includes wind turbines with different capacities. In this case, in order to forecast the power output of different turbines, the model must be performed for several times, with different input data and possibly different model parameters, which results in a considerably increased computational load. However, the power forecast of each turbine can be easily calculated by using the power curve of the related turbine once the common speed forecasts of the area are obtained. Therefore, one can claim that wind speed forecasting approach is more effective in terms of calculation time compared to the power forecasting approaches for the wind plants having various types of turbines.

Lastly, the most representative features apart from the wind speed, such as wind direction, temperature and pressure, can be included easily in the proposed model structure for the purpose of increasing accuracy. In this case, the proposed algorithm can also remove the requirement of such a feature selection process in forecasting. This is because the algorithm already assigns a higher value to the most informative variables in a continuously updated coefficient determination process. Therefore, potentially and with some considerations, all of the available features can be applied to the proposed forecasting approach without realizing a time-consuming data selection treatment.

## 6. Conclusion

In this paper, we present two approaches, called uniform and nonuniform CST-WSF, for short-term wind speed forecasting. These approaches are based on CS and structured-sparse recovery algorithms and exploit the sparsity of the correlations between a large number of weather stations. We showed that the proposed method provides significantly improved the forecasts compared to various benchmark models and advanced spatio-temporal approaches. Moreover, we enhanced the prediction performance considerably utilizing WT as part of the proposed forecasting scheme.

There are several future directions for this work. Our first objective is to employ the CST-WSF approach on a larger data set. Our conjecture is that the proposed forecasting approach have a better performance when applied to a larger collection of stations. Including different weather variables such as wind direction, pressure and temperature in the proposed algorithm is another future work. Investigating the impact of more accurate short-term forecasts on the energy markets is yet another direction. Lastly, the performance evaluation of the proposed approach for different Wavelet parameters (decomposition level, type of mother Wavelet, etc.), particularly for some special cases such as very short- and very long-term forecasts, ramps of wind speed, and very small and very large training data sets, is considered as a future study.

## Acknowledgments

This work is supported in part by EPRI and CERTS under sub-award 09-206; PSERC S-52; NSF under Grants 1135872, EECs-1129061, CPS-1239178, CNS-1239274; the Republic of Singapore National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore for the SinBerBEST

Program; Robert Bosch LLC through its Bosch Energy Research Network funding program; TUBITAK-2219 Program.

## References

- [1] Rajagopal R, Bitar E, Varaiya P, Wu F. Risk-limiting dispatch for integrating renewable power. *Int J Electr Power Energy Syst* 2013;44(1):615–28.
- [2] Quan H, Srinivasan D, Khambadkone AM, Khosravi A. A computational framework for uncertainty integration in stochastic unit commitment with intermittent renewable energy sources. *Appl Energy* 2015;152:71–82.
- [3] Callaway DS. Tapping the energy storage potential in electric loads to deliver load following and regulation, with application to wind energy. *Energy Convers Manage* 2009;50(5):1389–400.
- [4] Sanandaji BM, Hao H, Poola K. Fast regulation service provision via aggregation of thermostatically controlled loads. In: *Proceedings of the 47th Hawaii International Conference on System Sciences – HICSS47*; 2014. p. 2388–97.
- [5] Tascikaraoglu A, Erdinc O, Uzunoglu M, Karakas A. An adaptive load dispatching and forecasting strategy for a virtual power plant including renewable energy conversion units. *Appl Energy* 2014;119:445–53.
- [6] International Energy Agency (IEA). Technology roadmap: wind energy; 2013 <[http://www.iea.org/publications/freepublications/publication/Wind\\_2013\\_Roadmap.pdf](http://www.iea.org/publications/freepublications/publication/Wind_2013_Roadmap.pdf)>.
- [7] Lei M, Shiyan L, Chuanwen J, Hongling L, Yan Z. A review on the forecasting of wind speed and generated power. *Renew Sustain Energy Rev* 2009;13(4):915–20.
- [8] Zhu X, Genton MG. Short-term wind speed forecasting for power system operations. *Int Statist Rev* 2012;80(1):2–23.
- [9] Li G, Shi J. On comparing three artificial neural networks for wind speed forecasting. *Appl Energy* 2010;87(7):2313–20.
- [10] Tascikaraoglu A, Uzunoglu M. A review of combined approaches for prediction of short-term wind speed and power. *Renew Sustain Energy Rev* 2014;34:243–54.
- [11] Zhang W, Wu J, Wang J, Zhao W, Shen L. Performance analysis of four modified approaches for wind speed forecasting. *Appl Energy* 2012;99:324–33.
- [12] Catalão J, Pousinho H, Mendes V. Short-term wind power forecasting in Portugal by neural networks and wavelet transform. *Renew Energy* 2011;36(4):1245–51.
- [13] Liu H, Tian H-Q, Pan D-F, Li Y-F. Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks. *Appl Energy* 2013;107:191–208.
- [14] Liu Y, Shi J, Yang Y, Lee W-J. Short-term wind-power prediction based on wavelet transform-support vector machine and statistic-characteristics analysis. *IEEE Trans Indust Appl* 2012;48(4):1136–41.
- [15] Liu D, Niu D, Wang H, Fan L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renew Energy* 2014;62:592–7.
- [16] Fei S-W, He Y. Wind speed prediction using the hybrid model of wavelet decomposition and artificial bee colony algorithm-based relevance vector machine. *Int J Electr Power Energy Syst* 2015;73:625–31.
- [17] Tascikaraoglu A, Uzunoglu M, Vural B. The assessment of the contribution of short-term wind power predictions to the efficiency of stand-alone hybrid systems. *Appl Energy* 2012;94:156–65.
- [18] Alexiadis M, Dokopoulos P, Sahsamanoglou H. Wind speed and power forecasting based on spatial correlation models. *IEEE Trans Energy Convers* 1999;14(3):836–42.
- [19] Damousis IG, Alexiadis MC, Theocharis JB, Dokopoulos PS. A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Trans Energy Convers* 2004;19(2):352–61.
- [20] Gneiting T, Larson K, Westrick K, Genton MG, Aldrich E. Calibrated probabilistic forecasting at the stateline wind energy center: the regime-switching space-time method. *J Am Statist Assoc* 2006;101(475):968–79.
- [21] Hering AS, Genton MG. Powering up with space-time wind forecasting. *J Am Statist Assoc* 2010;105(489):92–104.
- [22] Tastu J, Pinson P, Kotwa E, Madsen H, Nielsen HA. Spatio-temporal analysis and modeling of short-term wind power forecast errors. *Wind Energy* 2011;14(1):43–60.
- [23] Hill DC, McMillan D, Bell KR, Infield D. Application of auto-regressive models to uk wind speed data for power system impact studies. *IEEE Trans Sustain Energy* 2012;3(1):134–41.
- [24] Xie L, Gu Y, Zhu X, Genton MG. Short-term spatio-temporal wind power forecast in robust look-ahead power system dispatch. *IEEE Trans Smart Grid* 2014;5(1):511–20.
- [25] Dowell J, Weiss S, Hill D, Infield D. Short-term spatio-temporal prediction of wind speed and direction. *Wind Energy* 2014;17(12):1945–55.
- [26] Tastu J, Pinson P, Trombe P-J, Madsen H. Probabilistic forecasts of wind power generation accounting for geographically dispersed information. *IEEE Trans Smart Grid* 2014;5(1):480–9.
- [27] He M, Yang L, Zhang J, Vittal V. A spatio-temporal analysis approach for short-term forecast of wind farm generation. *IEEE Trans Power Syst* 2014;29(4):1611–22.
- [28] Liu H, Tian H-Q, Chen C, Li Y-F. A hybrid statistical method to predict wind speed and wind power. *Renew Energy* 2010;35(8):1857–61.
- [29] Wang J-Z, Wang Y, Jiang P. The study and application of a novel hybrid forecasting model—a case study of wind speed forecasting in china. *Appl Energy* 2015;143:472–88.

- [30] Amjady N, Keynia F. Short-term load forecasting of power systems by combination of wavelet transform and neuro-evolutionary algorithm. *Energy* 2009;34(1):46–57.
- [31] Candès E, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inform Theory* 2006;52(2):489–509.
- [32] Chen SS, Donoho DL, Saunders MA. Atomic decomposition by basis pursuit. *SIAM J Scient Comput* 1999;20(1):33–61.
- [33] Tropp J. Greed is good: algorithmic results for sparse approximation. *IEEE Trans Inform Theory* 2004;50(10):2231–42.
- [34] Needell D, Tropp J. CoSaMP: iterative signal recovery from incomplete and inaccurate samples. *Appl Comput Harmonic Anal* 2009;26(3):301–21.
- [35] Eldar Y, Mishali M. Robust recovery of signals from a structured union of subspaces. *IEEE Trans Inform Theory* 2009;55(11):5302–16.
- [36] Eldar YC, Kuppinger P, Bölcskei H. Block-sparse signals: uncertainty relations and efficient recovery. *IEEE Trans Signal Process* 2010;58(6):3042–54.
- [37] Sanandaji BM, Vincent TL, Wakin MB. A review on sufficient conditions for structure identification of interconnected systems. In: Proceedings of the 16th IFAC Symposium on System Identification; 2012. p. 1623–8.
- [38] Iowa Environmental Mesonet. ASOS historical data; 2014 <<http://mesonet.agron.iastate.edu/ASOS/>>.
- [39] Song Z, Jiang Y, Zhang Z. Short-term wind speed forecasting with markov-switching model. *Appl Energy* 2014;130:103–12.
- [40] Bilgili M, Sahin B, Yasar A. Application of artificial neural networks for the wind speed prediction of target station using reference stations data. *Renew Energy* 2007;32(14):2350–60.
- [41] Sanandaji BM, Tascikaraoglu A, Poolla K, Varaiya P. Low-dimensional models in spatio-temporal wind speed forecasting. In: Proceedings of the 2015 American Control Conference – ACC; 2015. p. 4485–90.