Renewable Energy 133 (2019) 1352–1365

Contents lists available at ScienceDirect

Renewable Energy

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

Characterizing forecastability of wind sites in the United States

Cong Feng [a], Mucun Sun [a], Mingjian Cui [a], Erol Kevin Chartan [b], Bri-Mathias Hodge [b], Jie Zhang [a, *]

[a] The University of Texas at Dallas, Richardson, TX 75080, USA
[b] National Renewable Energy Laboratory, Golden, CO 80401, USA

ARTICLE INFO

Article history:
Received 17 January 2018
Received in revised form
15 August 2018
Accepted 26 August 2018
Available online 29 August 2018

Keywords:
Forecastability
Time series analysis
Wind power forecasting

ABSTRACT

With the rapid growth of wind power, managing its uncertainty and variability becomes critical in power system operations. Wind forecasting is one of the enablers to partially tackle challenges associated with wind power uncertainty. To improve the 'forecasting ability', defined as forecastability, different forecasting methods have been developed to assist grid integration of wind energy. However, forecasting performance not only relies on the power of forecasting models, but is also related to local weather conditions and (known as wind characteristics) wind farm properties. In this study, geospatial and instance spatial distributions of six wind characteristics and two forecasting error metrics are first analyzed based on 126,000 + wind sites in the United States. Forecasts in different look-ahead times are generated by using a machine learning based multi-model forecasting framework and the Weather Research and Forecasting model. A forecastability quantification method is developed by characterizing the relationship between forecastability and wind series entropy using three regression methods, i.e., linear approximation, locally weighted scatterplot smoother nonlinear nonparametric regression, and quantile regression. It is found that the forecastability of a wind site can be successfully characterized by wind series characteristics, thereby providing valuable information at different stages of wind energy projects.

1. Introduction

Wind power has captured considerable attention from researchers, utilities, and energy policymakers in recent years due to its rapid growth. Wind energy had 54.6 GW global capacity additions in 2016, reaching a cumulative capacity of more than 486.8 GW and 7% of the total power capacity in the world [1]. In the U.S., there are over 53,000 wind turbines operating, generating more than 84.1 GW of power across 41 states. Currently, a combined 20.9 GW of wind capacity is under construction in the United States according to the American Wind Energy Association [1]. However, the uncertain and variable characteristics of wind pose challenges to further increases in wind penetration. These challenges can be partially addressed by improving the accuracy of wind speed and power forecasting.

The 'forecasting ability' of a time series is defined as the ***forecastability*** (or predictability in some literature) [2,3]. Forecastability is of great importance to power system individuals, such as independent power producers (IPPs). This is because IPPs have limited experience with wind power forecasting but are required to provide wind forecasting (also known as decentralized wind power forecasts) in different look-ahead times to power purchasers and grid operators [4]. Ideally, IPPs are paid at the market clearing price for the amount of energy they supply. However, low-forecastability wind farms can have large deviations between the wind forecasts and actual outputs, which will be subjected to penalties for these forecast errors [5]. In addition to the imbalance costs, higher forecastability can help IPPs to better schedule maintenance [6]. The forecastability also concerns regional transmission organizations (RTOs)/independent system operators (ISOs) by (i) assisting to reduce reserve requirements, therefore achieving economic savings [7,8]; and (ii) optimizing the commitment and dispatch of thermal units to minimize production costs [9]. From the wind farm developers' perspective, the forecastability of a potential wind farm is a critical factor during the wind farm design [10].

The forecastability of a wind time series is influenced by different factors, such as learning power of the forecasting models and the time series characteristics (e.g., local weather conditions and wind farm properties). A number of studies have analyzed the forecastability by error distributions and spatio-temporal

* Corresponding author.
E-mail address: jiezhang@utdallas.edu (J. Zhang).

https://doi.org/10.1016/j.renene.2018.08.085
0960-1481/

Table 1
Case study summary.

State	ISO/RTO	No. of wind sites in WIND Toolkit	Wind penetration (by 2016)
Texas	ERCOT	7,869	12.6%
Oklahoma	SPP RTO	4,280	25.1%
Wyoming	N/A	8,196	9.4%
Kansas	SPP RTO	4,154	29.6%
New York	NYISO	2,859	2.9%

Note: ERCOT - Electric Reliability Council of Texas, SPP RTO - Southwest Power Pool Regional Transmission Organization, NYISO - New York ISO.

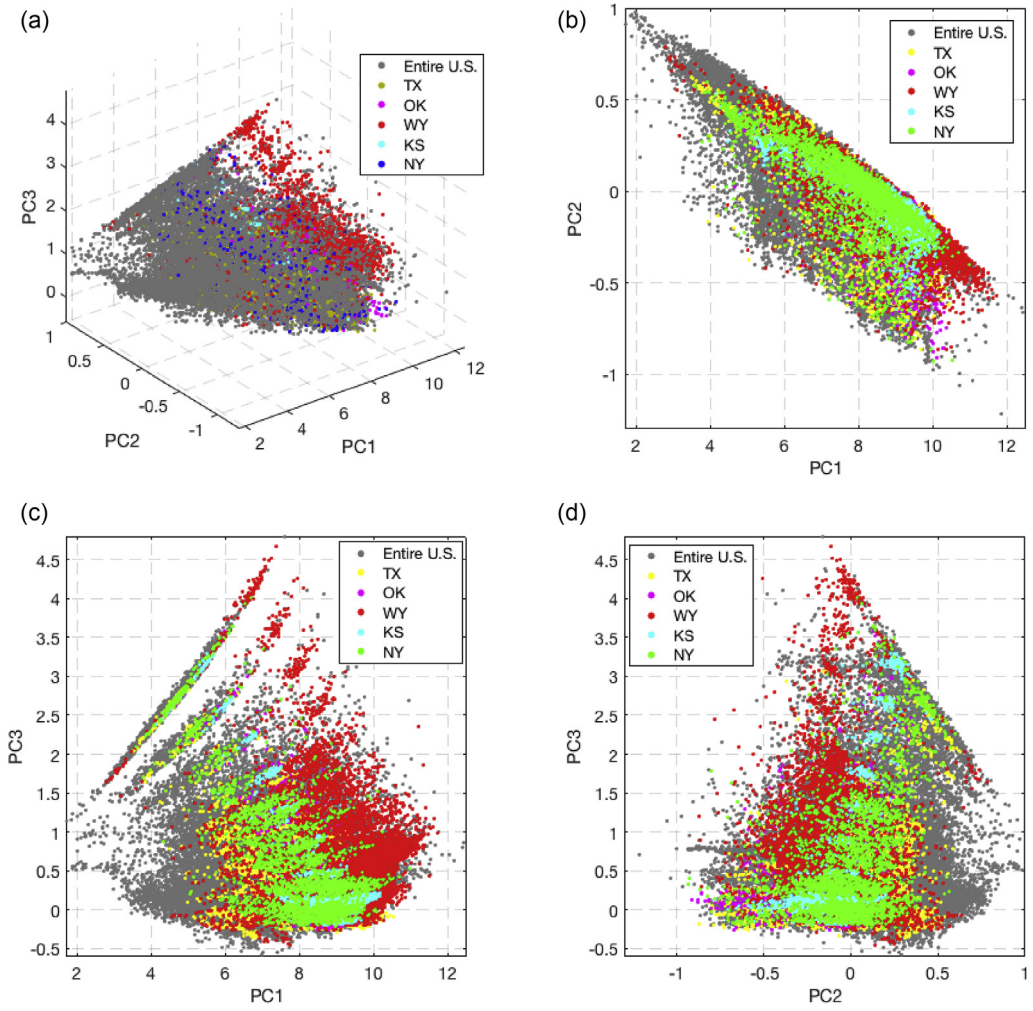

Fig. 1. Instance space of the wind sites of the entire WIND Toolkit and case studies. (a) Scatterplot in a 3D instance space; (b) PC1 vs: PC2 in a 2D instance space; (c) PC1 vs: PC3 in a 2D instance space; (d) PC2 vs: PC3 in a 2D instance space.

correlations. For example, spatio-temporal patterns of wind power forecasting errors in Denmark were proved to exist in Ref. [11]. Wind power forecasting error distributions were examined and compared in Refs. [12] and [13]. However, the aforementioned research didn't explore the forecasting errors and forecastability in conjunction with the forecasting process. Many studies tried to improve the forecastability by developing forecasting models (e.g., artificial neural networks [14], random forests [15], and Takagi-Sugeno fuzzy models [16,17]) with better learning power, which can be referred in review papers [18–21]. These forecasting methods are categorized based on different criteria (e.g., look-ahead times [22], algorithm principles, and input data [23]). With the development of big data analytics and machine learning, wind forecasting (especially short-term wind forecasting) has been significantly improved. Among different statistical models for short-term wind forecasting, ensemble models and deep learning models have been found to perform better than single-algorithm machine learning models [24–26]. Though these powerful machine learning methods are helpful for wind forecasting, the forecastability is largely affected by other factors, such as local weather conditions and wind farm properties. In this paper, we seek to quantify the forecastability of a wind farm and also investigate the impacts of local weather conditions (i.e., nonlinearity, spectral entropy, and mean wind speed) and wind farm properties (i.e., mean absolute gradient of 1-h wind power series, mean wind power, and capacity factor) on the forecatability, using the data of over 126,000 wind sites in the U.S. The relationships between the forecastability in different look-ahead times and the local weather conditions/

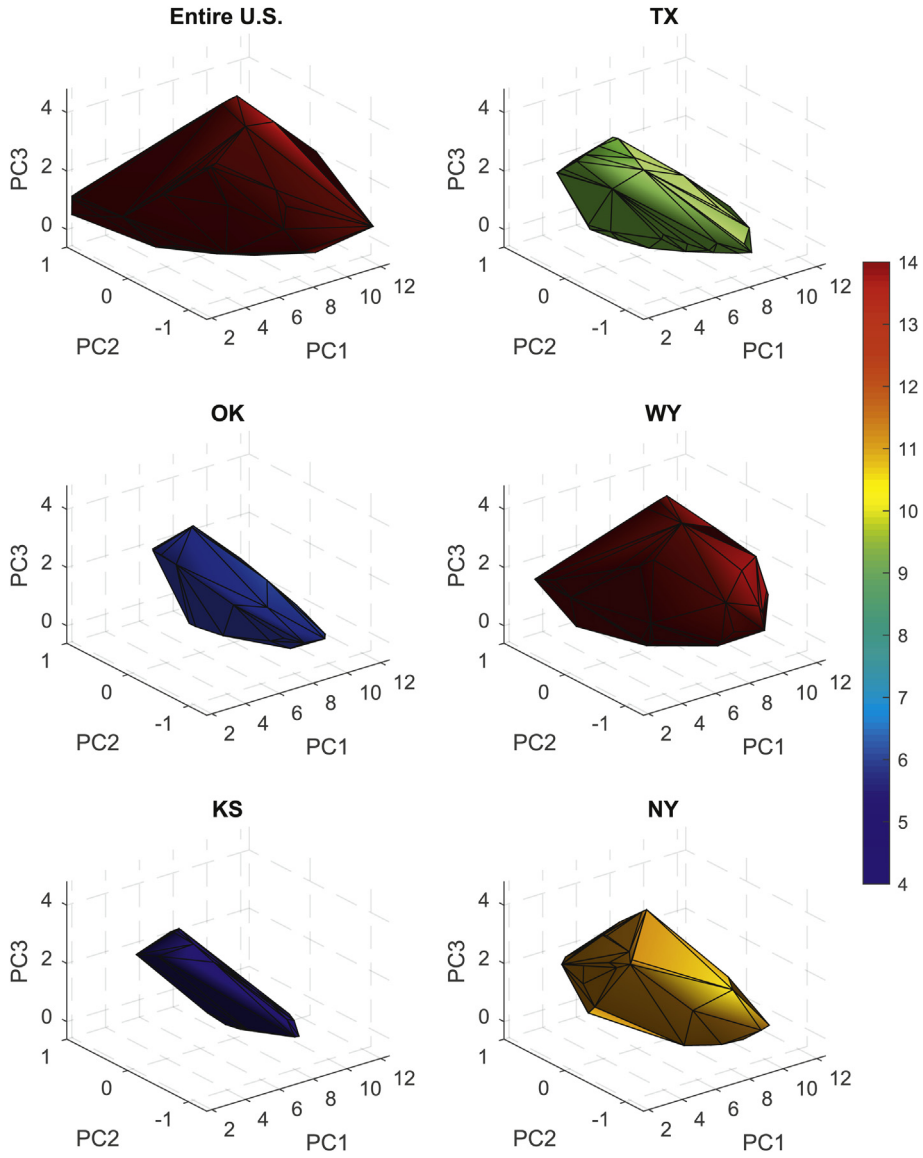

Fig. 2. The minimum convex polytope of the entire WIND Toolkit dataset and five study cases. The color bar indicates the value of *Div*. The *Div* values are 23.51, 9.30, 5.89, 16.79, 3.61, and 11.15 with respect to datasets of the WIND Toolkit dataset (the entire U.S.), TX, OK, WY, KS, and NY, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

wind farm properties are characterized. The main innovations and contributions of this paper include:

(i) Analyzing wind power time series characteristics of over 126,000 wind sites in the United States;
(ii) Performing wind power forecasting in different look-ahead times by the machine learning based multi-model forecasting framework and the Weather Research and Forecasting model;
(iii) Characterizing and quantifying the forecastability of wind sites based on wind power time series characteristics.

The remainder of the paper is organized as follows. Section 2 briefly describes the analyzed dataset and quantifies the data diversity. Geospatial and instance spatial distributions of local weather conditions, wind farm properties, and forecasting errors are analyzed in Section 3. Section 4 quantifies the forecastability of the wind sites based on different wind characteristics. Section 5 provides the concluding remarks.

2. Data description and pre-analysis

A large and diverse dataset is required to ensure the generality of this study. Thus, the Wind Integration National Dataset (WIND) Toolkit, the largest publically available wind dataset developed for the next generation grid integration studies, is adopted in this paper. The WIND Toolkit includes a meteorological dataset, wind power time series, and wind power forecasts generated by the Weather Research and Forecasting (WRF) Model. The WIND Toolkit power data was simulated at 100 m hub height. It covers 126,000+ locations with a 5-min resolution, spanning 2007 through 2013 [27]. Some critical wind power features, such as ramping characteristics, spatial and temporal correlations, wind plants' capacity factors, and time synchronized with load profiles, are represented in the WIND Toolkit.

The wind power time series and NWP forecasts at different

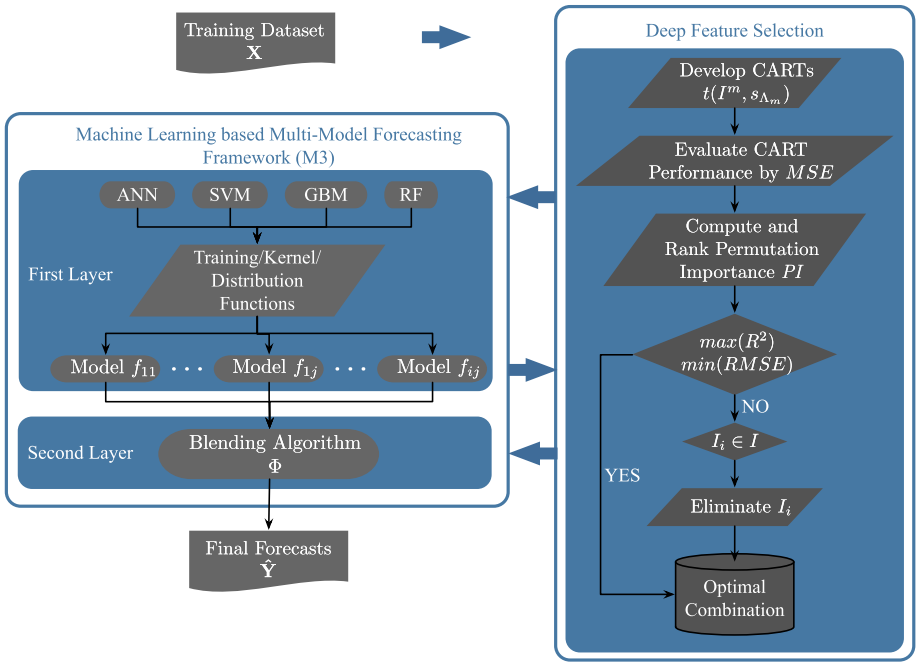

Fig. 3. Flowchart of the M3 method. The recursive feature elimination (RFE) is adopted in the deep feature selection (DS) module to carry out feature selection in the first layer and model selection in the second layer. $\boldsymbol{I^m}$ is the input to the mth single classification and regression tree (CART) in the RFS. s_{Λ_m} is the random vector to extract the bootstrap sample. R^2 and *RMSE* are coefficient determination and root mean square error, which are used to evaluate the RF model performance. $\boldsymbol{I_i}$ is a feature vector.

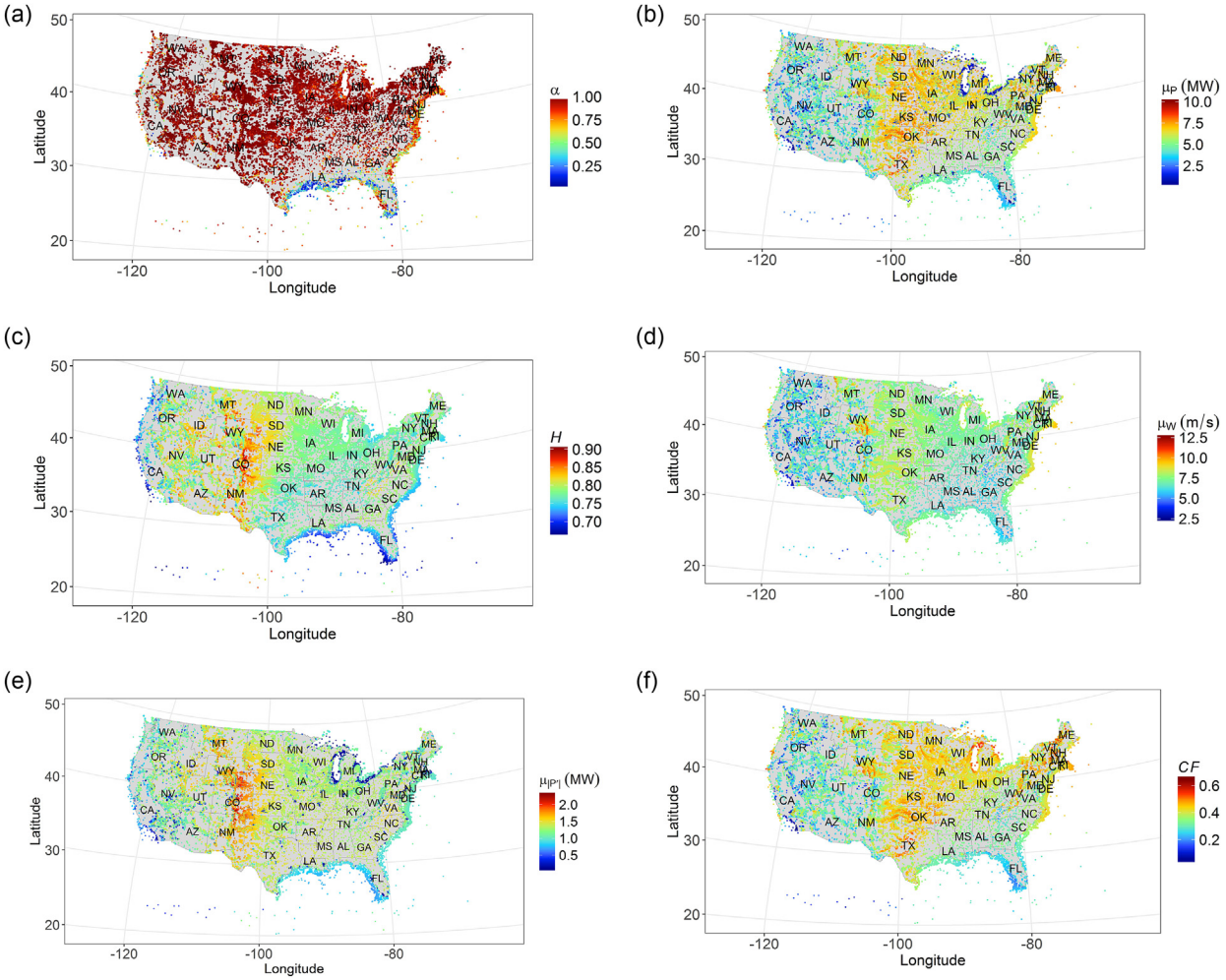

Fig. 4. Geospatial distributions of the wind site characteristics. (a) Geospatial distribution of power series nonlinearity; (b) Geospatial distribution of power series entropy; (c) Geospatial distribution of wind power variability; (d) Geospatial distribution of mean wind power; (e) Geospatial distribution of mean wind speed; (f) Geospatial distribution of capacity factor.

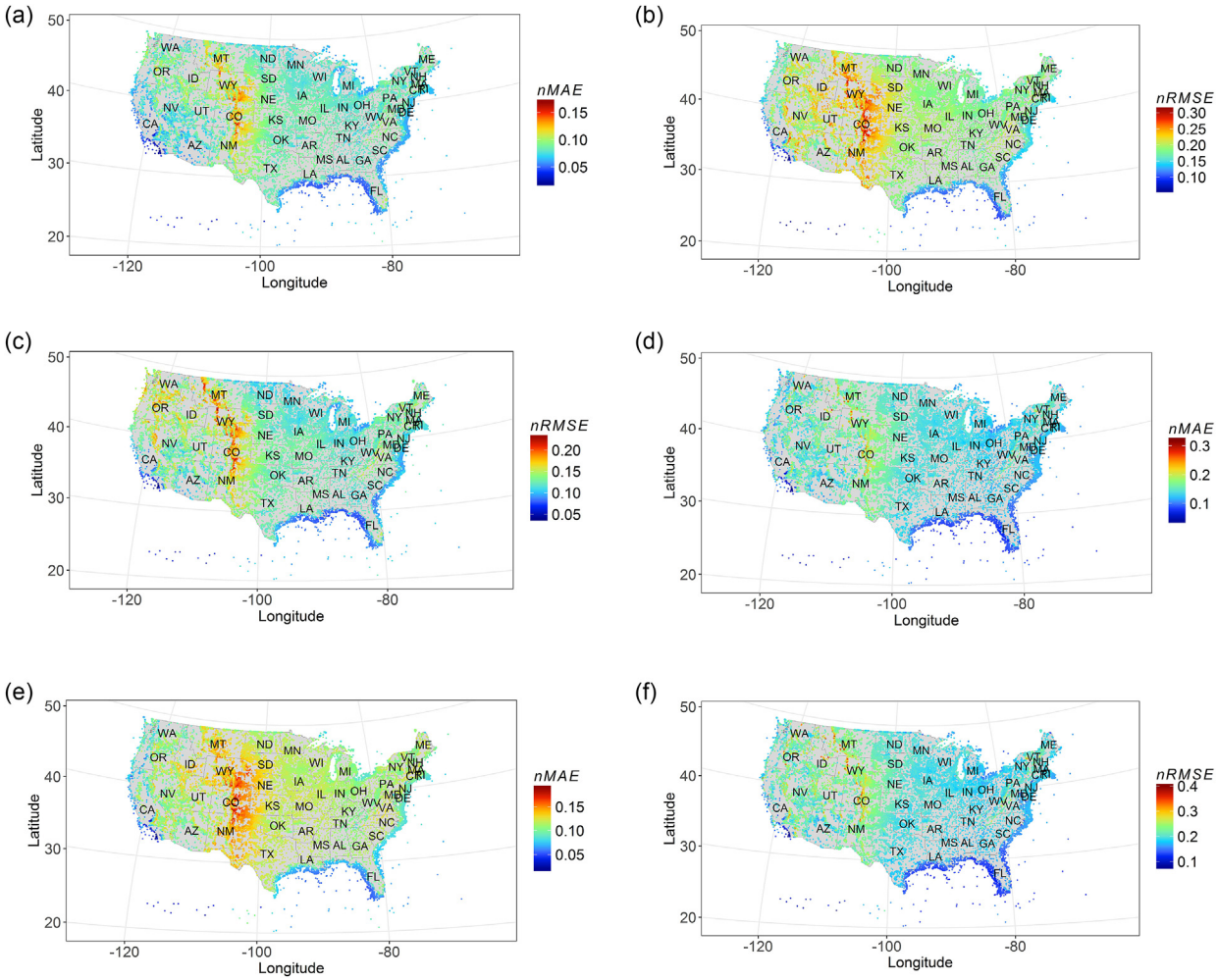

Fig. 5. Geospatial distributions of forecasting evaluation metrics. (a) Geospatial distribution of 1-h ahead forecasting nMAE; (b) Geospatial distribution of 1-h ahead forecasting nRMSE; (c) Geospatial distribution of 4-h ahead forecasting nMAE; (d) Geospatial distribution of 4-h ahead forecasting nRMSE; (e) Geospatial distribution of 1-day ahead forecasting nMAE; (f) Geospatial distribution of 1-day ahead forecasting nRMSE.

forecasting time horizons (i.e., 6-h and 1-day ahead) in WIND Toolkit are utilized. Statistical short-term (1-h to 4-h ahead) forecasts are created by a machine learning based multi-model forecasting framework (M3). The entire 126,000 + locations are investigated and five states belonging to different Independent System Operators (ISOs) are selected for case studies, which are Texas (TX), Oklahoma (OK), Wyoming (WY), Kansas (KS), and New York (NY). Texas is the leading state in terms of both cumulative and newly installed wind capacity. Oklahoma and Kansas are two states with high wind penetration. Wyoming has one of the highest wind power potential in the U.S. New York has 1,749 MW installed wind, leading the Northeast in overall wind energy capacity [28]; and the largest offshore wind farm in the U.S. is planning to be constructed in New York. Therefore, these states are selected for case studies. Table 1 summarizes the wind power status in these states, including the ISO/RTO, number of wind sites, and wind penetration.

To evaluate the diversity of the selected sites, a time series characteristic analysis (TSCA) method developed in our previous work is adopted and modified [29]. First, six characteristics are extracted from each WIND Toolkit site, which are nonlinearity, spectral entropy, wind power variability, mean wind power, mean wind speed, and wind farm capacity. These wind characteristics are expected to indicate the local weather conditions and wind farm properties of wind power series (details about the wind characteristics are described in Section 3). Then, principal component analysis (PCA) is performed to reduce the characteristic matrix dimension. It is found that the first three principal components preserve 81.1% of the information in the WIND Toolkit. The 3D and 2D projection scatter plots of WIND Toolkit sites are shown in Fig. 1. Fig. 1b–d shows that though the selected wind sites in the case studies are significantly fewer than the entire WIND Toolkit sites, the distribution areas of the case studies' scatters are close to that of the entire WIND Toolkit (datasets with different diversity values can be found in Ref. [29]). This is also supported by the diversity index (*Div*s) shown in Fig. 2, which is defined as the volume of the minimum convex polytope constructed by the scatter points [29]. It is found that the diversity of wind data is not purely dependent on the number of wind sites. For example, New York has 2,859 wind sites, which is fewer than Oklahoma and Kansas, but the dataset for the New York case is more diverse than the other two. Fig. 2 shows that the five selected cases maintain a similar level of diversity as the entire WIND Toolkit dataset.

3. Spatial analysis of wind sites

Short-term (1-h to 4-h ahead) and mid-term (6-h and 1-day ahead) wind power forecasting is performed by the M3 method [24] and the WRF model, respectively, for all the 126,000 + *wind*

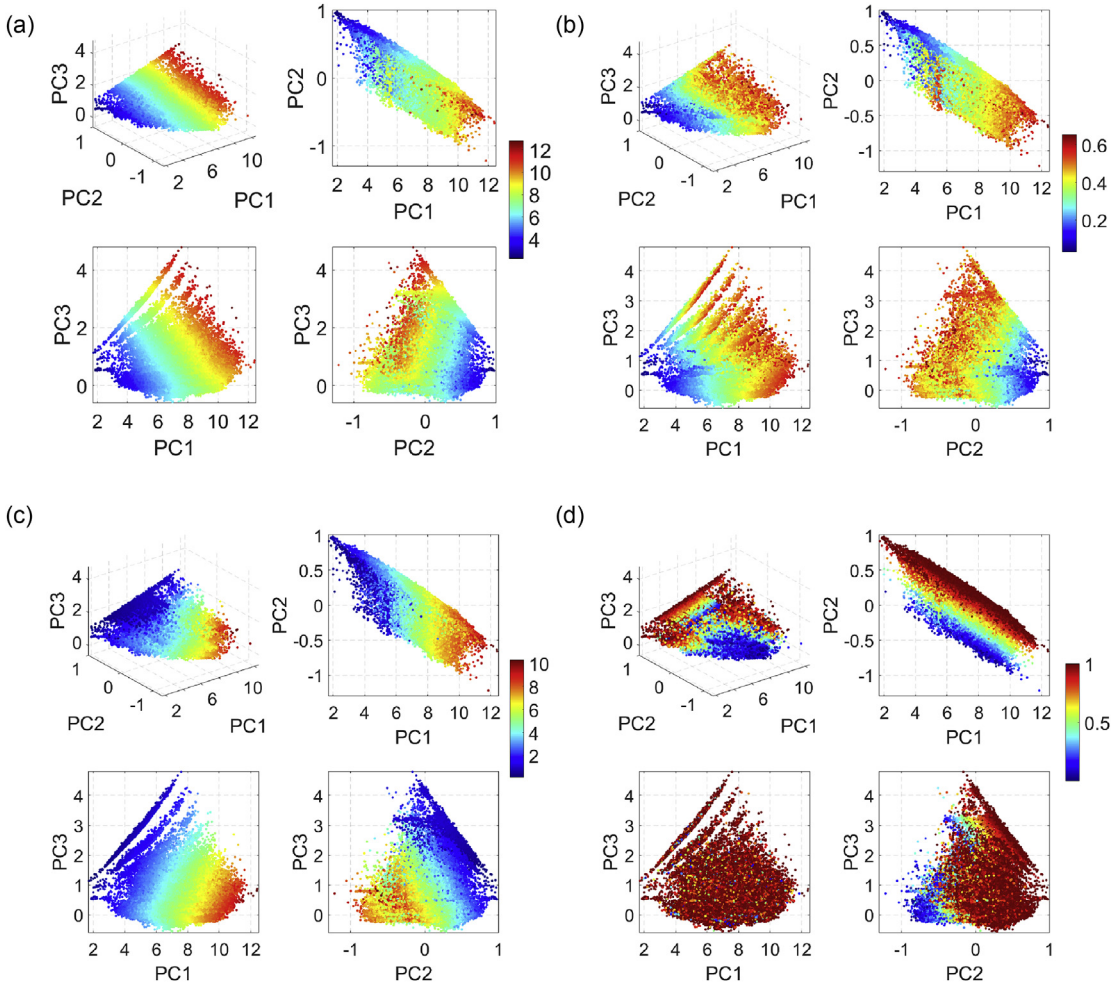

Fig. 6. Instance-spatial distributions of wind site characteristics and forecasting evaluation metrics. PCs are the principal components obtained by PCA. (a) Mean wind speed; (b) Capacity factor; (c) Mean wind power (d) Power series nonlinearity; (e) Power series entropy; (f) Wind power variability; (g) 1-h ahead forecasting nMAE; (h) 1-h ahead forecasting nRMSE.

sites. Please note that the focus of this paper is not to develop the most accurate forecasting algorithm but to investigate the impact of local weather conditions and wind farm properties on the wind site forecatability. Therefore, the impact of different forecasting algorithms on the wind site forecastability is not explored. Geospatial and instance-spatial distributions of wind sites' time series characteristics and forecasting evaluation metrics are analyzed in this section for forecastability quantification.

3.1. Wind characteristics

Local weather conditions and wind farm properties are quantified by six wind characteristics, i.e., nonlinearity, spectral entropy, the variability of wind power series, mean wind power, mean wind speed, and capacity factor. Details of the six wind characteristics are described as follows:

- Nonlinearity (α): The nonlinear nature of wind is the main challenge to accurately forecast wind power. To deal with the nonlinearity in wind time series, methods that adopt more powerful models [30] or decompose complex series into multiple simpler signals [31] have been developed in the literature. Therefore, nonlinearity is a critical feature to describe the wind characteristics. The Teräesvirta's neural network test is employed in this paper to measure the nonlinearity [32].
- Spectral entropy (H) of wind power series: Entropy is a measure of uncertainty of a random variable. In the area of renewable and load forecasting, entropy has been used as an index for time series analysis [29], an informative criterion for feature selection [33], a chaotic metric for signal preprocessing [34], and an error metric for forecasting model assessment [35]. Therefore, entropy is an important feature to quantify the wind series. The spectral entropy is calculated as [35]:

$$H = -\sum_{i=1}^{n} Pr(x_i)\log_2[Pr(x_i)] \tag{1}$$

where $Pr(x)$ is the mass probability of a random variable x, and n is the number of points in the spectrum.

- Variability (i.e., mean absolute 1-h power gradient, $\mu_{|P'|}$): the large variability of wind has introduced new challenges to power system operations, which stimulates the development of advanced scheduling approaches to security-constrained unit commitment [36], chance constrained optimal power flow [37], and others. Wind power variability could be characterized by

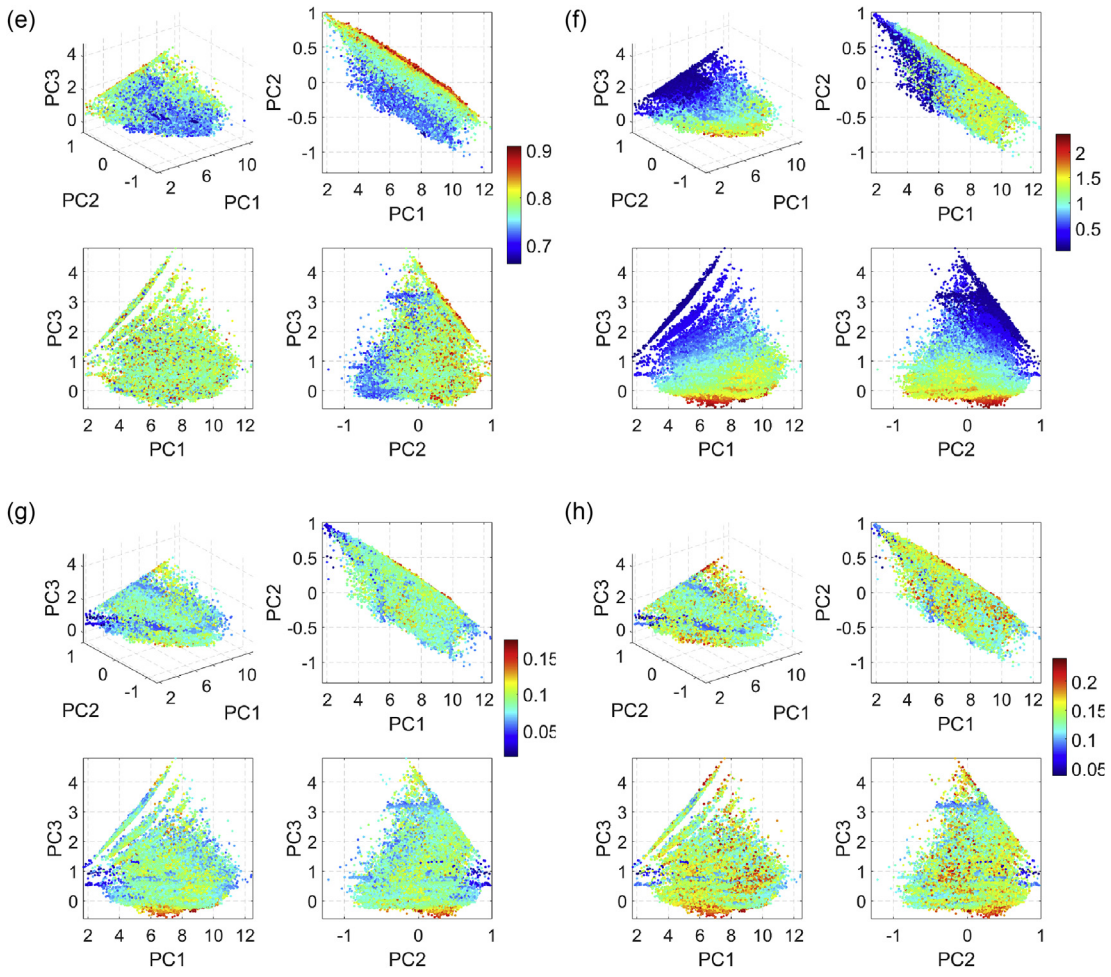

Fig. 6. (*continued*).

different metrics, such as mean absolute 1-h power gradient and variance [38]. In this paper, the mean absolute 1-h power gradient is chosen to represent the wind power variability:

$$\mu_{|P'|} = \frac{1}{n}\sum_{i=1}^{n}|P(i+1) - P(i)| \tag{2}$$

where P is the wind power and n is the size of the power vector.

- Mean wind speed (μ_S): Wind speed is the source of the potential wind power and primary source of wind power uncertainty. The mean wind speed is chosen since it is important for wind farm siting [39], wind series density fitting [40], and wind dataset reanalysis [41].
- Mean wind power (μ_P): The mean wind power is a critical index to quantify wind site properties. It presents a different trend from wind speed series due to the cubic power curve nature. The mean wind power is calculated from the WIND Toolkit wind power series based on wind speed and different power curves.
- Capacity factor (CF): Capacity factor is an essential indicator in evaluating a wind farm's efficiency [42], which is an important factor for wind farm siting, sizing, and investment [43,44]. Wind power capacity factor determines how fully the capacity of a wind farm is utilized, which is defined as the average generated power divided by the rated power:

$$CF = \frac{\frac{1}{n}\sum_{i=1}^{n}P(i)}{P_{rated}} \tag{3}$$

where P_{rated} is the rated power of a wind turbine.

3.2. Forecasting methods

Multiple hourly time-scale wind power forecasting is performed by using both statistical and numerical weather prediction (NWP) methods. A machine learning based multi-model forecasting framework (M3) developed in our previous work [24] is adopted for short-term forecasting (i.e., 1-h to 4-h ahead). The WRF model is adopted to create the 6-h and 1-day ahead forecasts. Details of the models are described below.

The M3 is a machine learning based ensemble methodology for short-term wind forecasting as shown in Fig. 3. There are two modules in the M3: a deep feature selection (DS) module and a two-layer ensemble forecasting (EF) module. The EF module consists of four machine learning algorithms in the first layer, which are artificial neural network (ANN), support vector regression (SVR), gradient boosting machine (GBM), and random forest (RF). These machine learning algorithms with several training algorithms, kernel functions, or distribution functions generate the forecasts, $\tilde{\boldsymbol{Y}}$, independently in the first layer. Specifically, three ANN models with standard back-propagation (BP), momentum-enhanced BP, and

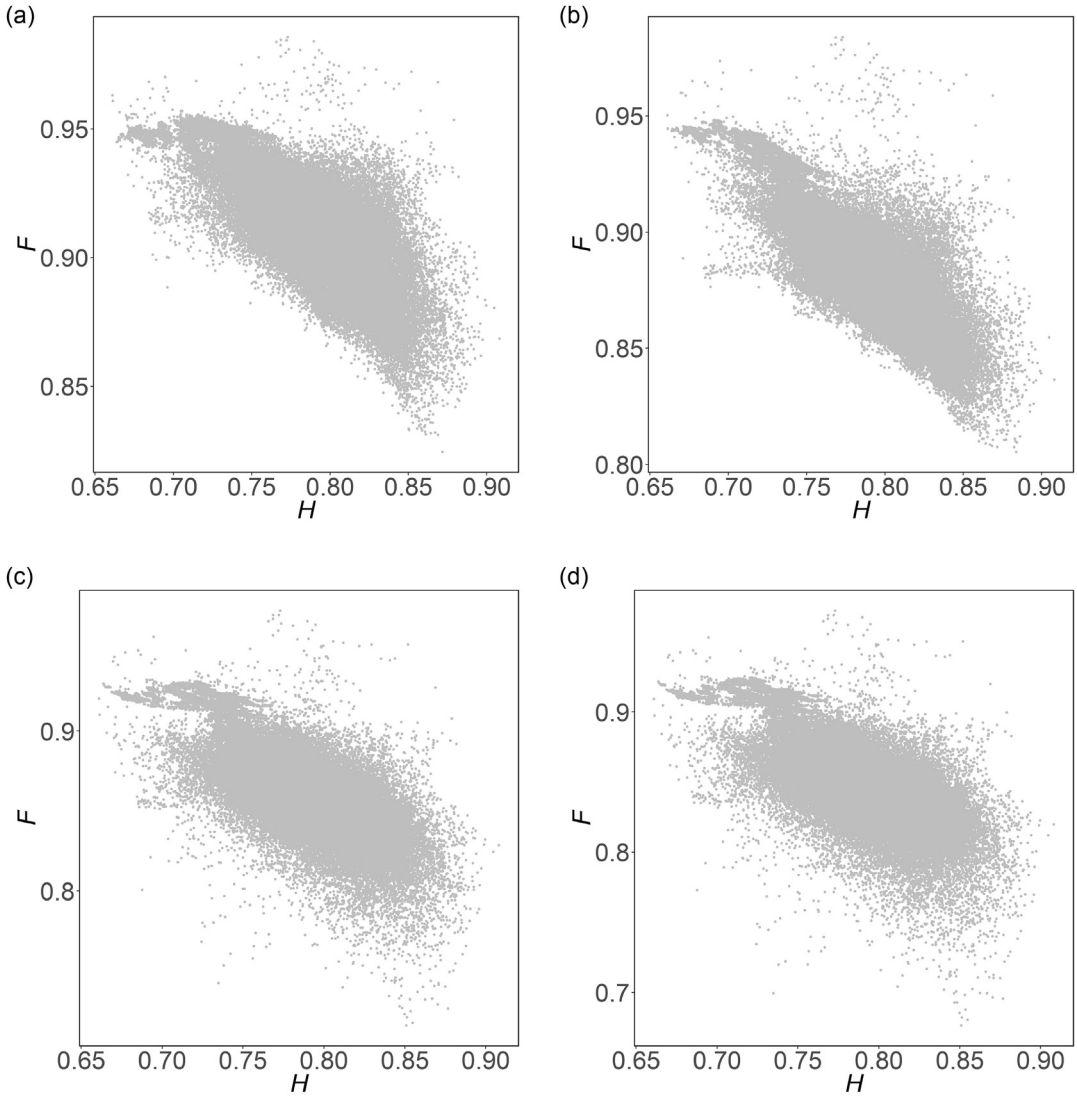

Fig. 7. The scatterplot of the forecastability (F) and the wind power series entropy (H) in different look-ahead times. (a) 1-h ahead forecasting; (b) 4-h ahead forecasting; (c) 6-h ahead forecasting; (d) 1-d ahead forecasting.

resilient BP training algorithms are selected based on their fast convergence and satisfactory performance [45–47]. The most popular kernels in SVR are used, including linear, polynomial, and radial base function kernels [48–50]. GBM models with squared, laplace, and T-distribution loss functions are empirically selected. Note that no single model is guaranteed to outperform others, and the algorithm and function selection is to diversify the behaviour of models in the first layer of M3. The inputs, $\boldsymbol{X}$, to the first-layer models are optimized by the recursive feature elimination method in the DS module. In the second layer, a blending machine learning algorithm is used to ensemble the forecasts from the first-layer models and provides the final forecasts, $\widehat{\boldsymbol{Y}}$. The optimal combination of the first-layer models is also ensured by the DS module. The M3 can be expressed as follows [51,52]:

$$\tilde{y}_{t,ij} = f_{ij}\left(\boldsymbol{x}_{t,opt}\right) \tag{4}$$

$$\widehat{y}_t = \Phi\left(\tilde{\boldsymbol{y}}_{t,opt}\right) \tag{5}$$

where t is the time index, $f_{ij}(*)$ is the ith first-layer model using kernel j, $\tilde{y}_{ij}$ is the forecasts provided by the model f_{ij}, $\boldsymbol{x}_{t,opt} \in \boldsymbol{X}$ is the optimized inputs to the first-layer models, $\tilde{\boldsymbol{y}}_{t,opt}$ is the optimal combination of the first-layer forecasts, $\widehat{y}_t$ is the final forecast at time t, and Φ is the blending algorithm in the second layer.

The WRF model is used to generate the 6-h and 1-day ahead mid-term forecasts. In the WIND Toolkit dataset, the WRF model is run with a 2-km grid. The model is initialized and forced at the boundaries with the National Oceanic and Atmospheric Administration (NOAA) Reforecast V2 Global Ensemble Forecast System (GEFS). The natural spatial-temporal correlations of different look-ahead times and between wind sites are kept in the forecasts. Details of the WRF model and WIND Toolkit can be found in Ref. [27].

3.3. Geospatial and instance-spatial distribution analysis

The M3 and the WRF models are used to generate multiple time-scale forecasts for the 126,000 + wind sites in the U.S. The geospatial distributions of the six wind characteristics are shown in Fig. 4. Interesting patterns are observed from these maps. For example, most wind series are largely non-linear in nature, as shown by dark red points in Fig. 4a. But the offshore wind sites,

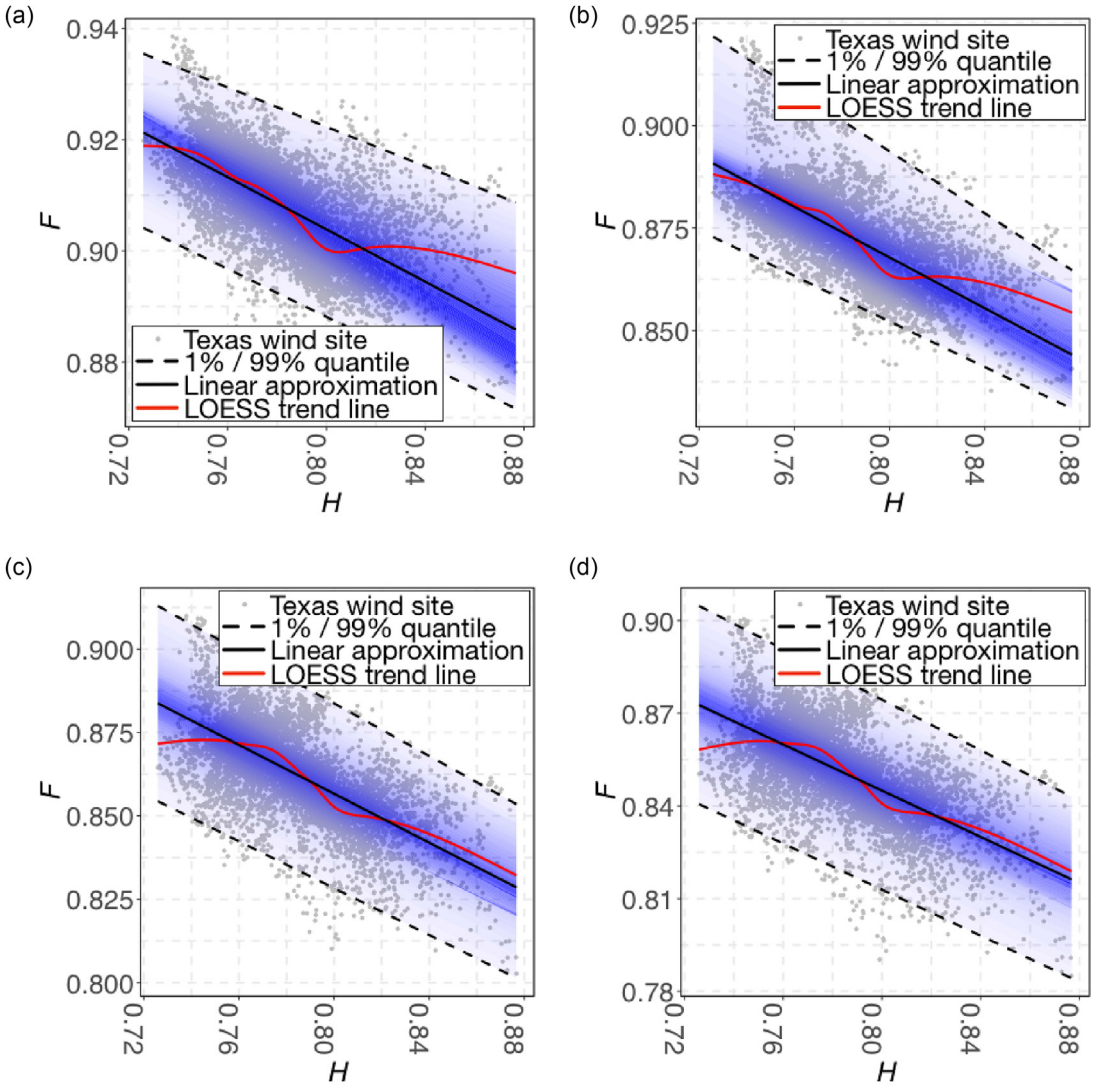

Fig. 8. Forecastability of Texas wind sites. (a) 1-h ahead forecasting; (b) 4-h ahead forecasting; (c) 6-h ahead forecasting; (d) 1-d ahead forecasting.

such as the wind sites at the Gulf of Mexico and on the East Coast, are more linear (indicated by lighter color points in Fig. 4a) than onshore sites. This is due to the more consistent and non-disturbed nature of offshore wind [53]. Figures 4b and c show the uncertainty and variability, respectively, whereas they agree with each other on several patterns. For instance, the wind speed series in the Mountain States, especially in Colorado, is more chaotic (illustrated by red and orange points in Figs. 4b and 4c) than other regions. Figures 4d-f show the geospatial distributions of three closely interconnected characteristics, i.e., mean wind power, mean wind speed, and capacity factor of all the WIND Toolkit sites. The Mountain States have higher mean wind speed (shown by green and orange points in Fig. 4e) than those in other areas due to the speed-up effect [54]. The American East presents higher wind speeds, wind power, and larger capacity factors than the American West (except for the mountain areas). This is because: (a) the wind forming at the tropical and subtropical latitudes in the northern hemisphere prefers to move toward the west-northwest; and (b) the water temperature gradient along the West Coast is lower than that along the East Coast (due to the warmer Gulf Stream). The map distributions of the forecasting normalized mean absolute errors (*nMAE*s) and normalized root mean square errors (*nRMSE*s) of different look-ahead times are shown in Fig. 5. The two error metrics are defined as:

$$nMAE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\hat{y}_t - y_t}{y_{max}}\right| \tag{6}$$

$$nRMSE = \frac{1}{y_{max}}\sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}} \tag{7}$$

where t is the time index, n is the length of forecasting data, $\hat{y}, y$, and y_{max} are the forecasted value, actual value, and maximum actual value, respectively. It is found that the offshore wind forecasts are more accurate (illustrated by the light green and blue points in Fig. 5) than the onshore wind forecasts, due to the relatively more stable characteristics of the offshore wind. Some states in the Mountain States, such as Wyoming, Colorado, and part of Montana and New Mexico, present large forecasting errors due to the large terrain roughness. By comparing the *nMAE* figures (i.e., Fig. 5a, c, and e) to the *nRMSE* figures (i.e., Fig. 5b, d, and f), even though the forecasting *nMAE*s of the wind sites in Washington, Oregon, and the boundary of California and Nevada are small, there

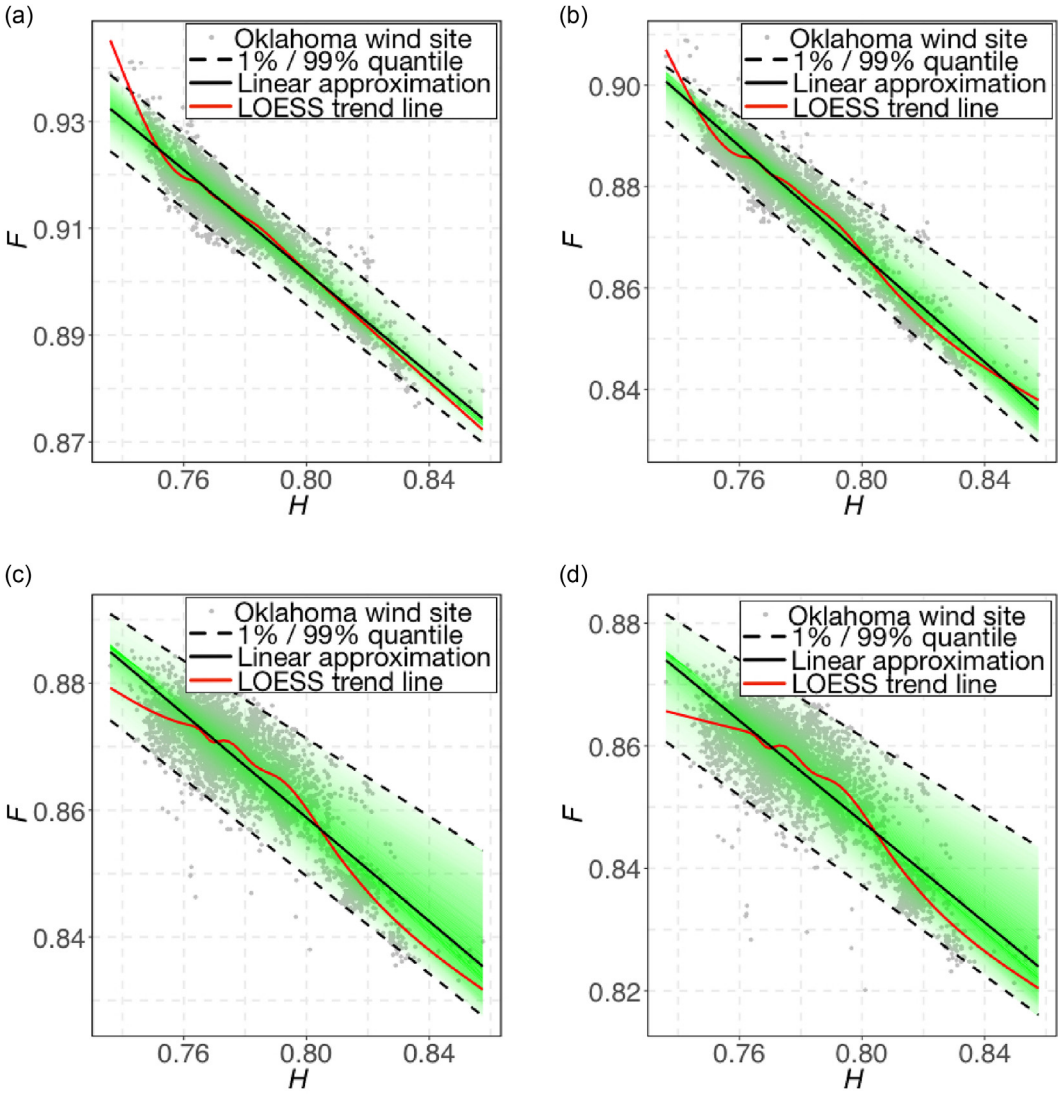

Fig. 9. Forecastability of Oklahoma wind sites.(a) 1-h ahead forecasting; (b) 4-h ahead forecasting; (c) 6-h ahead forecasting; (d) 1-d ahead forecasting.

is a higher chance to generate large forecasting errors as indicated by the *nRMSE* metric that penalizes large forecasting errors in a higher order.

Even though Figs. 4 and 5 are informative in analyzing the geospatial patterns of wind sites in the U.S., it is still challenging to compare different wind characteristics and forecasting metrics. Therefore, a 3D instance space is constructed to convert the data from a time series to a static point, while maintaining certain physical information. The scatter plots in the 3D instance space and corresponding 2D projections are shown in Fig. 6, wherein the color bar indicates the magnitude of a wind characteristic or a forecasting error metric (only considering 1-h ahead forecasting here). By comparing Figs. 6a and 6c, it is found that the color gradient directions of the mean wind power (Fig. 6c) and mean wind speed (Fig. 6a) are different, which means wind power and wind speed are not positively dependent with each other. This is because different power curves and capacities are applied in the WIND Toolkit data. The color gradient directions of the mean wind speed (Fig. 6a) and the wind farm capacity factor (Fig. 6b) are almost the same, indicating that the mean wind speed influences the capacity factor. Fig. 6d shows that the nonlinearity varies along the $-(\overrightarrow{\mathbf{P1}} + \overrightarrow{\mathbf{P2}}) \in \mathbb{R}^2_{PC1\times PC2}$ and the top left region in $\mathbb{R}^2_{PC1\times PC2}$ has large nonlinearity and thus is challenging for forecasting, vice versa. As shown in Fig. 6f, the wind power variability value increases in the $(\overrightarrow{\mathbf{P1}} - \overrightarrow{\mathbf{P3}}) \in \mathbb{R}^2_{PC1\times PC3}$ direction first and then changes to the $-\overrightarrow{\mathbf{P3}} \in \mathbb{R}^2_{PC1\times PC3}$ direction. The scatter plot in Fig. 6e is not evidently layered, whereas the entropy value decreases along the $-(\overrightarrow{\mathbf{P1}} + \overrightarrow{\mathbf{P2}}) \in \mathbb{R}^2_{PC1\times PC2}$ direction. The 1-h ahead forecasting error scatter plots are shown in Figs. 6g and 6h, which are more chaotic than wind series characteristic plots in Figs. 6a-f. By observing 2D projection spaces $\mathbb{R}^2_{PC1\times PC2}$ and $\mathbb{R}^2_{PC1\times PC3}$, forecasting errors, especially the *nMAE* is highly correlated with the entropy in Fig. 6e. Thus, the relationship between the power series entropy and the *nMAE* will be further investigated in the following section (Please note that only the univariate regression is considered in this research because the forecastability is already defined in the complex regression process).

4. Forecastability quantification

In this section, the forecastability is quantified by using the power series entropy. The scatter plots in Fig. 7 indicate a strong relationship between forecastability and entropy for all the WIND

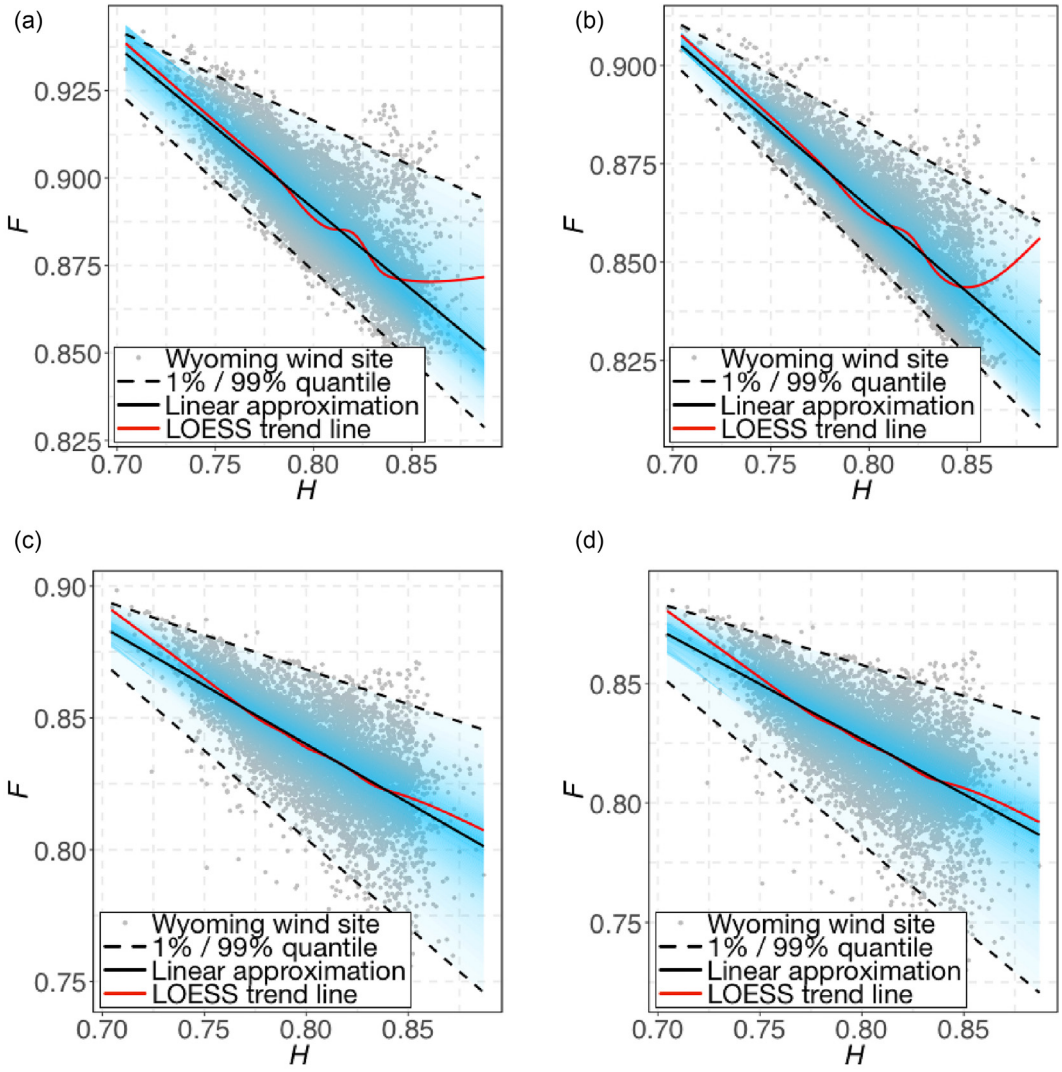

Fig. 10. Forecastability of Wyoming wind sites. (a) 1-h ahead forecasting; (b) 4-h ahead forecasting; (c) 6-h ahead forecasting; (d) 1-d ahead forecasting.

Toolkit sites. The error metric *nMAE* is one of the most widely used error metrics in wind forecasting and has been adopted to indicate the forecastability in the literature [5]. In this paper, the forecastability is defined as $F = 1 - nMAE$, and a larger F value indicates better forecastability. It is observed from Fig. 7 that wind series with smaller entropy is more forecastable. In addition, forecasting look-ahead time also affects the forecastability of time series. To be more specific, the time series forecastability is decreased with the increasing forecasting horizon. To better quantify the relationship, both linear approximation and nonlinear trend are used to map entropy to forecastability for the selected five regions. Quantile regression is adopted to show the uncertainty of the linear relationship between forecastability and entropy.

4.1. Regression methods

Three methods are used to characterize the relationship between forecastability and time series characteristics, which are linear regression, locally weighted scatterplot smoother (LOESS), and quantile regression. Linear regression is selected because the linear relationship is straightforward in practice and Figs. 6e and 6g show a similar pattern. The LOESS and quantile regression are able to characterize the nonlinear relationship and uncertainty between forecastabiltiy and time series characteristics.

The LOESS is a flexible nonparametric nonlinear regression method [55], which is widely used for depicting relationships between variables [56]. In this paper, the weighted least square is used to fit a quadratic function at each local segment, with a tricube weight function $w(x) = (1 - |x|^3)^3$. The smoothing parameter, α, is set to be 0.75 to represent proper nonlinearity of the fitting.

Quantile regression is a statistical method that can characterize uncertainty information. If the cumulative distribution function of a random variable X is $F_X(x) := Pr(X \leq x)$, the quantile function of this random variable is defined as the inverse cumulative distribution function: $Q(p) = inf\{x \in \mathbb{R} : p \leq F(x)\}$. Quantiles from 1% to 99% with an 1% interval are calculated. A detailed description about quantile regression can be referred to [57].

4.2. Forecastability of five selected cases

The forecastability of the five designed cases (i.e., Texas, Oklahoma, Wyoming, Kansas, and New York) are characterized. The quantified relationship between forecastability at different look-ahead times and the power series entropy is visualized in

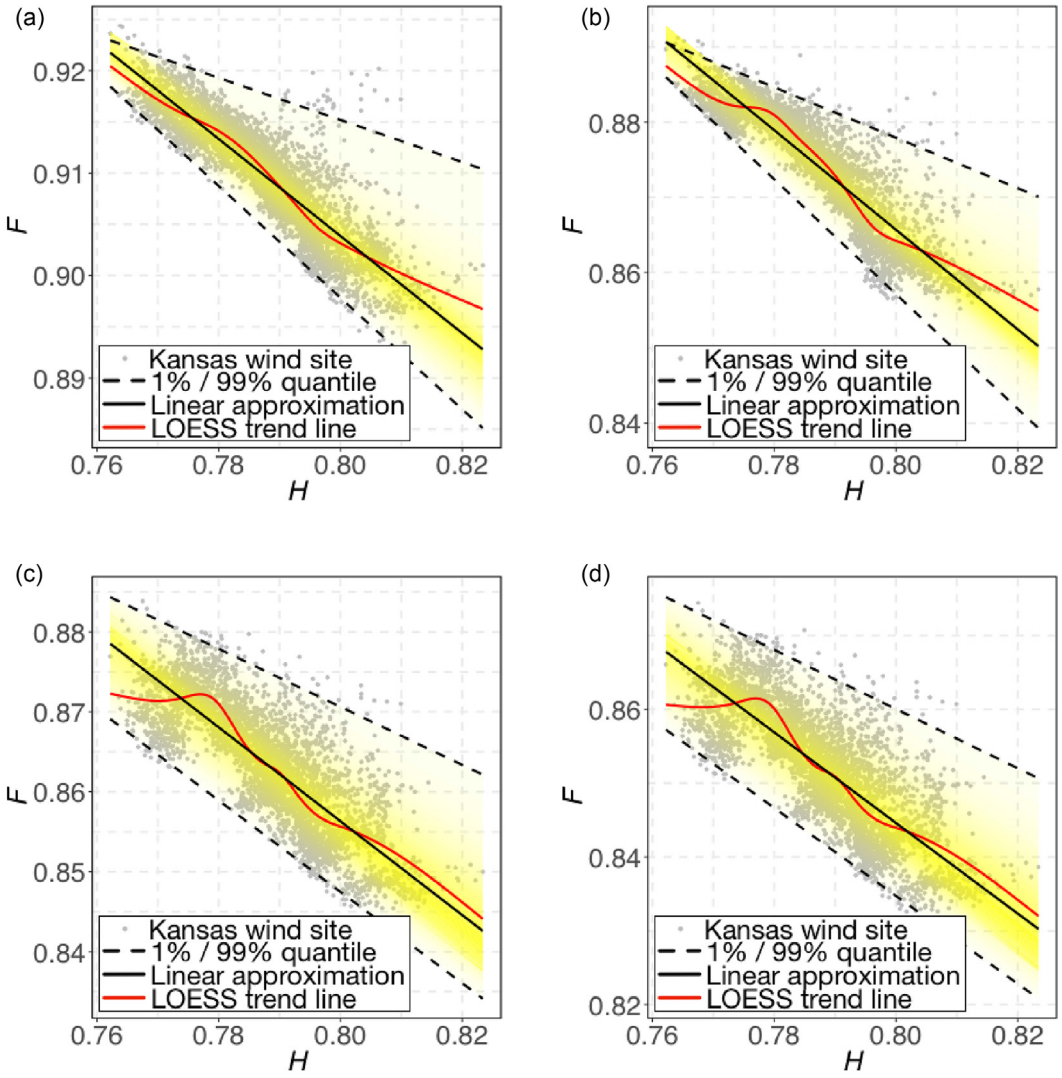

Fig. 11. Forecastability of Kansas wind sites. (a) 1-h ahead forecasting; (b) 4-h ahead forecasting; (c) 6-h ahead forecasting; (d) 1-d ahead forecasting.

Figs. 8–12. It is shown that the LOESS regression lines generally decrease with the increasing entropy for all five cases and different look-ahead times. Some increasing trends only occur at the minimum or maximum entropy areas, where the number of samples is small, as shown in Figs. 10b and 12a. By comparing the forecastability within the same look-ahead time forecasting, the uncertainty in forecastability and wind power series entropy is also highly related. For example, the uncertainty of forecastability in Figs. 10–12 (indicated by wider 1%–99% quantiles) is larger when the original wind power series has larger uncertainty (indicated by larger entropy). By comparing the different look-ahead times, the forecastability has relatively larger uncertainty in longer-term forecasting, and 1%/99% quantiles are more cone-like, as shown in Figs. 10 and 12. With the increasing look-ahead time, the fitting tends to be more nonlinear, as illustrated by the larger gaps between linear approximations and LOESS regression lines in Figs. 9–12 and the decreasing r^2 values in Table 2. The slope parameter a in Table 2 reflects the decreasing rate of the forecastability regarding to entropy.

By comparing the five cases, Texas and Oklahoma show more consistent uncertainty in forecastability than the other three cases with the same look-ahead times, since the 1% and 99% quantiles are more nearly parallel in Figs. 8 and 9 than those in Figs. 10–12. Oklahoma has the most stable forecastability compared to the other four cases, as shown by the narrowest quantiles. Additionally, Oklahoma presents the most linear relationship between the forecastability and the wind power series entropy in different look-ahead times, as shown by the largest r^2 values in Table 2. The relationship between forecastability and entropy of New York wind sites is relatively linear and stable in the 1-h ahead forecasting and becomes variable and nonlinear in the longer-term forecasting, as shown in Fig. 12. Overall, the forecastability is successfully quantified by the power series entropy in different look-ahead times, though the observations vary among the five selected cases.

5. Conclusion

This paper characterized the forecastability of wind series at 126,000 + sites in the U.S., which are extremely useful at different stages of a wind energy project, from wind resource assessment, wind farm design, to long-term planning of wind integration. The geospatial and instance spatial distributions of local weather conditions, wind farm properties, and forecasting errors of 126,000 + wind sites in the U.S. were first analysed and interesting patterns were observed. For example, the wind in certain regions, such as the Gulf of Mexico, is easier to forecast due to the smaller

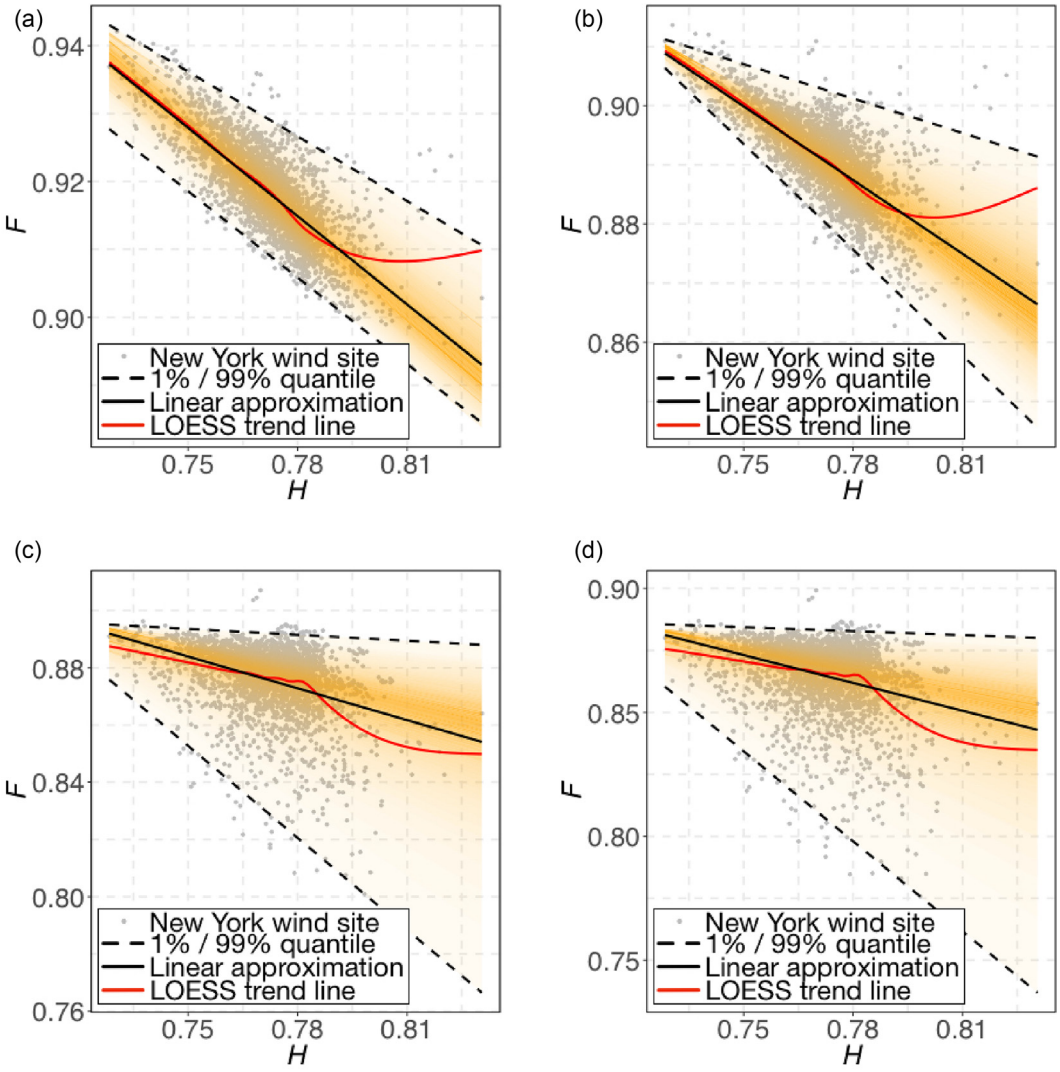

Fig. 12. Forecastability of New York wind sites. (a) 1-h ahead forecasting; (b) 4-h ahead forecasting; (c) 6-h ahead forecasting; (d) 1-d ahead forecasting.

Table 2
Linear approximation parameters.

State	1-h ahead		4-h ahead		6-h ahead		1-d ahead	
	a	r^2	a	r^2	a	r^2	a	r^2
Texas	−0.48	0.50	−0.31	0.46	−0.37	0.43	−0.37	0.39
Oklahoma	−0.55	0.90	−0.53	0.90	−0.41	0.75	−0.41	0.70
Wyoming	−0.46	0.67	−0.43	0.74	−0.45	0.50	−0.46	0.44
Kansas	−0.47	0.77	−0.66	0.82	−0.59	0.65	−0.61	0.62
New York	−0.43	0.61	−0.42	0.53	−0.37	0.13	−0.37	0.10

Note: a is the slope of the linear approximation. r^2 is the coefficient of determination of the linear approximation.

nonlinearity, entropy, and variability. To quantify the relationship between forecastability and wind series characteristics, forecasts in different look-ahead times are generated for all WIND Toolkit sites by using the M3 statistical model and WRF NWP model. We found the spectral entropy of a wind power series could be used to quantify the forecastability of a wind location. Three regression methods (i.e., linear approximation, LOESS nonparametric trend, and quantile regression) were applied to characterize the relationship between forecastability and spectral entropy, and uncertainties in forecastability were also quantified. The characterized forecastability could provide valuable information to both wind farm developers and power system operators. Future research will investigate the financial models using wind site forecastability to provide wind farm investors decision-making support.

Acknowledgement

This work was supported by the National Renewable Energy Laboratory under Subcontract No. XGJ-6-62183-01 (under the U.S. Department of Energy Prime Contract No. DE-AC36-08GO28308). This work was authored in part by Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Wind Energy Technology Office. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

References

[1] R. Wiser, M. Bolinger, 2016 wind Technologies Market Report, Tech. rep., Lawrence Berkeley National Laboratory, 2016.
[2] G. Kariniotakis, Renewable Energy Forecasting: from Models to Applications, Elsevier-Woodhead Publishing, 2017.
[3] N. Cutululis, L.M. Faiella, S. Otterson, B. Barahona, J. Dobschinski, Report on Design Tool on Variability and Predictability, Tech. rep., Technical University of Denmark, 2013.
[4] A. Botterud, Z. Zhi, J. Wang, R.J. Bessa, H. Keko, J. Mendes, J. Sumaili, V. Miranda, Use of Wind Power Forecasting in Operational Decisions, Tech. rep., Argonne National Laboratory (ANL), 2011.
[5] R. Girard, K. Laquaine, G. Kariniotakis, Assessment of wind power predictability as a decision factor in the investment phase of wind farms, Appl. Energy 101 (2013) 609–617.
[6] C. Feng, J. Zhang, Wind power and ramp forecasting for grid integration, in: Advanced Wind Turbine Technology, Springer, 2018, pp. 299–315.
[7] M. Cui, J. Zhang, C. Feng, A.R. Florita, Y. Sun, B.-M. Hodge, Characterizing and analyzing ramping events in wind power, solar power, load, and netload, Renew. Energy 111 (2017) 227–244.
[8] M.R. Milligan, A.H. Miller, F. Chapman, Estimating the Economic Value of Wind Forecasting to Utilities, National Renewable Energy Laboratory Golden, CO, 1995.
[9] M. Cui, C. Feng, Z. Wang, J. Zhang, Q. Wang, A. Florita, V. Krishnan, B.-M. Hodge, Probabilistic wind power ramp forecasting based on a scenario generation method, in: Proc. IEEE Power Energy Soc. Gen. Meeting, Chicago, IL, USA, 2017, pp. 1–5.
[10] J. Zhang, B.-M. Hodge, Forecastability as a design criterion in wind resource assessment, in: 8th International Conference on Foundations of Computer-aided Process Design, 2014, pp. 1–7.
[11] J. Tastu, P. Pinson, E. Kotwa, H. Madsen, H.A. Nielsen, Spatio-temporal analysis and modeling of short-term wind power forecast errors, Wind Energy 14 (1) (2011) 43–60.
[12] B.-M. Hodge, M. Milligan, Wind power forecasting error distributions over multiple timescales, in: Power and Energy Society General Meeting, 2011 IEEE, IEEE, 2011, pp. 1–8.
[13] H. Bludszuweit, J.A. Domínguez-Navarro, A. Llombart, Statistical analysis of wind power forecast error, IEEE Trans. Power Syst. 23 (3) (2008) 983–991.
[14] D. Wang, H. Luo, O. Grunder, Y. Lin, Multi-step ahead wind speed forecasting using an improved wavelet neural network combining variational mode decomposition and phase space reconstruction, Renew. Energy 113 (2017) 1345–1358.
[15] A. Lahouar, J.B.H. Slama, Hour-ahead wind power forecast based on random forests, Renew. Energy 109 (2017) 529–541.
[16] I.G. Damousis, M.C. Alexiadis, J.B. Theocharis, P.S. Dokopoulos, A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation, IEEE Trans. Energy Convers. 19 (2) (2004) 352–361.
[17] F. Liu, R. Li, Y. Li, Y. Cao, D. Panasetsky, D. Sidorov, Short-term wind power forecasting based on ts fuzzy model, in: Power and Energy Engineering Conference (APPEEC), 2016 IEEE PES Asia-Pacific, IEEE, 2016, pp. 414–418.
[18] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, E. Feitosa, A review on the young history of the wind power short-term prediction, Renew. Sustain. Energy Rev. 12 (6) (2008) 1725–1744.
[19] M. Lei, L. Shiyan, J. Chuanwen, L. Hongling, Z. Yan, A review on the forecasting of wind speed and generated power, Renew. Sustain. Energy Rev. 13 (4) (2009) 915–920.
[20] A.M. Foley, P.G. Leahy, A. Marvuglia, E.J. McKeogh, Current methods and advances in forecasting of wind power generation, Renew. Energy 37 (1) (2012) 1–8.
[21] X. Zhao, S. Wang, T. Li, Review of evaluation criteria and main methods of wind power forecasting, Energy Proced. 12 (2011) 761–769.
[22] I. Okumus, A. Dinler, Current status of wind energy forecasting and a hybrid method for hourly predictions, Energy Convers. Manag. 123 (2016) 362–371.
[23] C. Feng, M. Cui, M. Lee, J. Zhang, B.-M. Hodge, S. Lu, H.F. Hamann, Short-term global horizontal irradiance forecasting based on sky imaging and pattern recognition, in: Power & Energy Society General Meeting, 2017 IEEE, IEEE, 2017.
[24] C. Feng, M. Cui, B.-M. Hodge, J. Zhang, A data-driven multi-model methodology with deep feature selection for short-term wind forecasting, Appl. Energy 190 (2017) 1245–1257.
[25] H. Wang, G. Wang, G. Li, J. Peng, Y. Liu, Deep belief network based deterministic and probabilistic wind speed forecasting approach, Appl. Energy 182 (2016) 80–93.
[26] C. Feng, J. Zhang, Short-term load forecasting with different aggregation strategies, in: ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference (IDETC/CIE 2018), American Society of Mechanical Engineers, 2018.
[27] C. Draxl, A. Clifton, B.-M. Hodge, J. McCaa, The wind integration national dataset (WIND) Toolkit, Appl. Energy 151 (2015) 355–366.
[28] R. Baranowski, F. Oteri, I. Baring-Gould, S. Tegen, 2016 state of Wind Development in the united states by Region, Tech. rep., NREL (National Renewable Energy Laboratory (NREL), Golden, CO (United States), 2017.
[29] C. Feng, E.K. Chartan, B.-M. Hodge, J. Zhang, Characterizing time series data diversity for wind forecasting, in: Big Data Computing Applications and Technologies (BDCAT), 2017 IEEE/ACM 4th International Conference on, IEEE, 2017.
[30] H. Samet, F. Marzbani, Quantizing the deterministic nonlinearity in wind speed time series, Renew. Sustain. Energy Rev. 39 (2014) 1143–1154.
[31] Y. Ren, P. Suganthan, N. Srikanth, A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods, IEEE Trans. Sustain. Energy 6 (1) (2015) 236–244.
[32] T. Teräsvirta, Power properties of linearity tests for time series, Stud. Nonlinear Dynam. Econom. 1 (1).
[33] S. Jurado, À. Nebot, F. Mugica, N. Avellana, Hybrid methodologies for electricity load forecasting: entropy-based feature selection with machine learning and soft computing techniques, Energy 86 (2015) 276–291.
[34] W. Sun, Y. Wang, Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network, Energy Convers. Manag. 157 (2018) 1–12.
[35] J. Zhang, A. Florita, B.-M. Hodge, S. Lu, H.F. Hamann, V. Banunarayanan, A.M. Brockway, A suite of metrics for assessing the performance of solar power forecasting, Sol. Energy 111 (2015) 157–175.
[36] C. Shao, X. Wang, M. Shahidehpour, X. Wang, B. Wang, Security-constrained unit commitment with flexible uncertainty set for variable wind power, IEEE Trans. Sustain. Energy 8 (3) (2017) 1237–1246.
[37] W.-S. Tan, M. Shaaban, M.P. Abdullah, Chance-constrained programming for day-ahead scheduling of variable wind power amongst conventional generation mix and energy storage, IET Renew. Power Gener. 11 (14) (2017) 1785–1793.
[38] J. Dobschinski, How good is my forecast? comparability of wind power forecast errors, in: Proceedings of the 13th International Workshop on Large Scale Integration of Wind Power into Power Systems as Well as on Transmission Networks for Offshore Wind Farms, Berlin, Germany, vol. 1113, 2014.
[39] B. Belabes, A. Youcefi, O. Guerri, M. Djamai, A. Kaabeche, Evaluation of wind energy potential and estimation of cost using wind energy turbines for electricity generation in north of Algeria, Renew. Sustain. Energy Rev. 51 (2015) 1245–1255.
[40] S. Pishgar-Komleh, A. Keyhani, P. Sefeedpari, Wind speed and power density analysis based on weibull and Rayleigh distributions (a case study: Firouzkooh county of Iran), Renew. Sustain. Energy Rev. 42 (2015) 313–322.
[41] D. Cannon, D. Brayshaw, J. Methven, P. Coker, D. Lenaghan, Using reanalysis data to quantify extreme wind power generation statistics: a 33 year case study in great britain, Renew. Energy 75 (2015) 767–778.
[42] T.-P. Chang, F.-J. Liu, H.-H. Ko, S.-P. Cheng, L.-C. Sun, S.-C. Kuo, Comparative analysis on power curve models of wind turbine generator in estimating capacity factor, Energy 73 (2014) 88–95.
[43] M. Trenkel-Lopez, P. Matthews, Method for designing a high capacity factor wide area virtual wind farm, IET Renew. Power Gener..
[44] H. Cetinay, F.A. Kuipers, A.N. Guven, Optimal siting and sizing of wind farms, Renew. Energy 101 (2017) 51–58.
[45] O. A. Adeoti, P. A. Osanaiye, Effect of training algorithms on the performance of ann for pattern recognition of bivariate process, Int. J. Comput. Appl. 69 (20).
[46] Ö. Kişi, E. Uncuoğlu, Comparison of Three Back-propagation Training Algorithms for Two Case Studies.
[47] G. Li, J. Shi, On comparing three artificial neural networks for wind speed forecasting, Appl. Energy 87 (7) (2010) 2313–2320.
[48] J. Zhou, J. Shi, G. Li, Fine tuning support vector machines for short-term wind speed forecasting, Energy Convers. Manag. 52 (4) (2011) 1990–1998, https://doi.org/10.1016/j.enconman.2010.11.007.
[49] M. Aizerman, E.M. Braverman, L. Rozonoer, Theoretical foundations of potential function method in pattern recognition, Autom. Rem. Contr. 25 (6) (1964) 917–936.
[50] V. Vapnik, A.Y. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, Theor. Probab. Appl. 16 (2) (1971) 264.
[51] C. Feng, M. Cui, B.-M. Hodge, S. Lu, H. F. Hamann, J. Zhang, An Unsupervised Clustering-based Short-term Solar Forecasting Methodology Using Multi-model Machine Learning Blending, arXiv preprint arXiv:1805.04193.
[52] C. Feng, J. Zhang, Hourly-similarity based solar forecasting using multi-model machine learning blending, in: IEEE PES General Meeting 2018, IEEE PES, 2018.
[53] B. Snyder, M.J. Kaiser, Ecological and economic cost-benefit analysis of offshore wind energy, Renew. Energy 34 (6) (2009) 1567–1578.
[54] K. Rønenes, P.-Å. Krogstad, Wind tunnel simulation of terrain effects on wind farm siting, Wind Energy 12 (4) (2009) 391–410.
[55] W.S. Cleveland, Robust locally weighted regression and smoothing scatterplots, J. Am. Stat. Assoc. 74 (368) (1979) 829–836.
[56] W.G. Jacoby, Loess:: a nonparametric, graphical tool for depicting relationships between variables, Elect. Stud. 19 (4) (2000) 577–613.
[57] Quantile Regression, Available at: https://en.wikipedia.org/wiki/Quantile_regression, (Accessed 16 May 2018).