Renewable Energy 69 (2014) 74–81

Contents lists available at ScienceDirect

Renewable Energy

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

Validity of stationary probabilistic models for wind speed records of varying duration

Sanjay R. Arwade [a,*], Massimiliano Gioffrè [b]

[a] Dept. of Civil & Env. Engg., University of Massachusetts, 130 Natural Resources Rd., Amherst, MA 01003, USA
[b] Dept. of Civil & Env. Engg., University of Perugia, Via G. Duranti 93, 06125 Perugia, Italy

ARTICLE INFO

Article history:
Received 13 August 2013
Accepted 6 March 2014
Available online 1 April 2014

Keywords:
Mean wind speed
Stationarity
Stochastic process
Wind energy

ABSTRACT

A method for assessing the degree of non-stationarity in annual wind speed records is presented. The method uses quantitative tests on the wind speed records to assess the length of the period over which an assumption of stationarity in the wind record can be considered to provide reasonable engineering accuracy. The tests evaluate stationarity in second moment properties and marginal distribution. Numerical examples are provided for three offshore sites along the Atlantic Coast of the United States—one off of Virginia, and two off of Maine. The examples illustrate that an assumption of stationarity over a period of one week is largely justified, but that such an assumption over periods of one month is certainly not. Assuming stationarity over a period of a week can lead to errors in model values of the second moment properties of 2%–3% whereas the assumption applied to a monthlong period can lead to error greater than 10%. Examination of the persistence of marginal distribution reveals that, although true stationarity in marginal distribution persists for a few days at most, there exist two 'seasons', winter and summer, during which the marginal distribution remains relatively consistent, with rapid changes in marginal distribution occurring near the beginning and ends of these seasons. Results are found to be largely consistent across the three sites investigated as numerical examples. The methods and results presented here may be useful to those investigating the potential for offshore wind energy development using stochastic process theory to study wind speed or power production since stationary stochastic models provide simpler and more accessible predictions of quantities such as probabilities of exceedance of threshold values, upcrossing rates, and residence times.

1. Introduction

The characteristics of time varying wind speed records have importance to many fields of engineering, particularly structural engineering in which wind can generate design-controlling loads and wind energy engineering in which the wind determines energy production and structural loading on the blades, mechanical equipment and support structure. In this paper we address the period of time over which an offshore wind record can reasonably be considered to be stationary with respect to its marginal distribution and second moment properties. The question of stationarity has, we believe, primary relevance to the modeling of the power production of an offshore wind turbine as a stochastic process. Although models for both stationary and non-stationary stochastic processes are well developed and robust [1,2], models for stationary processes are much simpler and more accessible to engineers not well grounded in the mathematics of stochastic processes. Therefore, it is highly advantageous to be able to use stationary models for time varying parameters of engineering systems whenever possible, provided that one is keenly aware of the approximation inherent in applying a stationary model to an underlying physical process which has time-varying, that is to say non-stationary, parameters.

The goal of this paper is to introduce a framework for quantitatively assessing whether an assumption of stationarity is appropriate for offshore wind records over periods of time longer than one day—in other words, neglecting the well known diurnal variation of wind speed. Diurnal variations are important for performing short term analysis that leads, for example, to estimates of extreme loads on structures, and in fact simulations of not longer than 1 h are recommended to avoid the introduction of non-stationarity to the analysis of wind turbines [3,4]. The magnitude

* Corresponding author. Tel.: +1 413 577 0926.
E-mail addresses: arwade@umass.edu (S.R. Arwade), massimiliano.gioffre@unipg.it (M. Gioffrè).

http://dx.doi.org/10.1016/j.renene.2014.03.016
0960-1481/

of the diurnal variation, is small compared to the seasonal fluctuations that are primarily addressed here, and furthermore, engineers interested in the long-term performance of wind turbines with respect to power generation are likely to be interested in performance over time scales of weeks, months, and even years. The question of the time period of interest is closely related to the spectral content of the wind speed time history, a question addressed in the seminal work of Van der Hoven [5]. Spectral content, though not a primary topic of this paper, is treated in this paper to clearly illustrate the periodicities that are present in the data. The framework that is presented here is applied directly to three offshore sites along the Atlantic Coast of the United States, and some conclusions are drawn regarding the appropriate periods of time for use of stationary models for the wind speed.

The work presented in this paper distinguishes itself from the extensive body of literature regarding the statistical characterization and probabilistic modeling of wind speed records in concentrating on medium term (days to months) characteristics of the wind speed and on core statistics such as the gross marginal distribution or second moment properties rather than the extreme values generated over short periods of the wind record. In the extreme value work most effort has been devoted to estimation of distribution with particular emphasis on modeling the upper tails of the distribution and determining the effect of averaging period or recording frequency on such estimates [6–9]. Also, this work treats offshore sites rather than onshore sites, with offshore sites considered particularly important to the growth of the US wind energy industry.

Medium-term variability of wind records has been discussed in the literature, and some key examples include studies of inter-annual variability of wind characteristics [10], month to month variations [11], and seasonal variations [12]. While these papers address issues similar to those addressed here, they are focussed on assessing the variability of the wind speed rather than providing a quantitative framework in which to decide upon a time interval over which an assumption of stationarity is valid. Furthermore, they all address on-shore sites.

Another key set of studies with relevance to this work is represented in the large body of literature regarding the appropriate choice of distribution for wind speed modeling [6,13,14]. Here the emphasis is not on the choice of distribution, per se, but rather on the persistence of a particular model distribution in time. In the context of wind energy engineering, these studies are related to the so-called 'measure-correlate-predict' (MCP) approach to assessing sites for wind energy development [15]. MCP studies have tended to focus on wind data aggregated on an annual basis rather than the medium-term lengths studied here, and also on the length of time for which data should be collected to properly assess a site [16]. In fact, the approaches presented here could inform future MCP work by providing quantitative support to the choice of observation/recording time.

It is important to distinguish the work presented here, which address the period of time over which a stationary model is appropriate for wind records, from research which has established appropriate simulation lengths for analyzing the performance of wind turbines [3,4] (10 min is the standard), and from questions regarding the details of how wind speed records should be recorded (averaging time and recording frequency) in order to properly estimate distributions. Finally, the medium-term (days to months) nature of this work is emphasized to contrast it with the important topic of long term non-stationarity in meteorological data brought on by global climate change [17].

The remainder of the paper is structured as follows: First the analysis approach is described including the ways in which wind records can be aggregated over different time periods and the statistical tests used for assessing assumptions of stationarity; then the characteristics of the three sites selected as case studies are described; results of the case studies, or examples, are presented both in terms of stationarity with respect to marginal distribution and second moment properties; concluding remarks summarize the key methods and findings of the paper.

2. Analysis approach

The most commonly available types of long term wind measurements report a version of the wind speed at intervals ranging from roughly a minute to an hour—the data presented here consist of hourly measurements. Since our interest lies primarily in characteristics of wind speed records at time scales greater than one day and ranging up to one year, we neglect the diurnal non-stationarities that lead to the presence of harmonics in wind speed records with periods of 12 h and 24 h and aggregate the wind data (see the following section for details) into data sets representing the collections of measured wind speeds in individual 24 h (one day), 168 h (seven days) and 672 h (28 days) periods. We adopt the 672 h period as representative of a month since 672 h is nearly an integer divisor of 8760 h, the length of a (non-leap) year and is nearly an integer multiple of the 168 h, the number of hours in a calendar week. Using the usual calendar months complicates matters significantly due to the varying length of the calendar months and would obscure the key approaches and results of this paper.

Consider the sequence of random variables $\{V_i\}$, $i = 1,\ldots,8760$ that represent hourly wind speeds at a particular location during the course of a year. What can be obtained from instruments are values of $\{v_{ij}\}$, $i = 1,\ldots,8760$, $j = 1,\ldots,n_y$ where n_y is the number of years for which measurements are available and v is written in lower case since v_{ij} represents a realization, or sample, of the random variable V_i. The collection $\{v_{ij}\}$, $j = 1,\ldots,n_y$ for fixed i can be viewed as an ensemble of samples of V_i.

Now define daily, weekly, and monthly aggregations of the random variables V_i as follows. Let $\{V_{d,i}\}$, $i = 1,\ldots,365$ be a sequence of random variables that is defined such that $V_{d,i}$ represents the hourly wind speed V_k for any $k \in [24(i-1)+1, 24i]$. That is, $V_{d,i}$ characterizes the hourly wind speed during day i of the year, and the collection of realizations for fixed i is defined such that $\{v_{d,ij}\} \equiv \{v_{kl}\}$, $j = 1,\ldots,24n_y$, $k \in [24(i-1)+1, 24i]$, $l = 1,\ldots,n_y$. For the weekly aggregations, let $\{V_{w,i}\}$, $i = 1,\ldots,365$ be a sequence of random variables that is defined such that $V_{w,i}$ represents the hourly wind speed V_k for any $k \in [24(i-4), 24(i+3)]$ with yearly periodicity of the wind speed statistics giving $k = 8760 - k$, $k < 1$ and $k = k - 8760$, $k > 8760$. The variable $V_{w,i}$ characterizes the hourly wind speed during the 168 h period centered on day i of the year such that the collection of realizations for fixed i is $\{v_{w,ij}\} \equiv \{v_{kl}\}$, $j = 1,\ldots,168n_y$, $k \in [24(i-4)+1, 24(i+3)]$, $l = 1,\ldots,n_y$ with periodicity for $k < 1$ and $k > 8760$ defined as previously. Finally, the monthly aggregations are defined by $\{V_{m,i}\}$, $i = 1,\ldots,365$, a sequence of random variables that is defined such that $V_{m,i}$ represents the hourly wind speed V_k for any $k \in [24(i-14)+1, 24(i+14)]$ with yearly periodicity of the wind speed statistics defined as for the weekly aggregations. The variable $V_{m,i}$ characterizes the hourly wind speed during the 672 h period centered on day i of the year such that the collection of realizations for fixed i is $\{v_{m,ij}\} \equiv \{v_{kl}\}$, $j = 1,\ldots,672n_y$, $k \in [24(i-14)+1, 24(i+14)]$, $l = 1,\ldots,n_y$ with periodicity for $k < 1$ and $k > 8760$ defined as previously.

The core approach to evaluating the rate of temporal change in the probabilistic characteristics of wind speed records presented in this paper involves testing the hypothesis that pairs of random variables $\{V_{d,i}, V_{d,j}\}$, $\{V_{w,i}, V_{w,j}\}$, and $\{V_{m,i}, V_{m,j}\}$ come from the same underlying distribution. The two-sample Kolmogorov–Smirnov (KS2) test is used here to evaluate the null hypothesis. In the KS2 test, the test statistic is $D = \sup_x(F_1(x) - F_2(x))$ where $F_1(x)$ and $F_2(x)$ are the cumulative distribution functions of the two random variables being tested. For example, if $F_1(x) = F_{Vw,25}(v)$ and $F_2(x) = F_{Vw,50}(v)$, the KS2 test would evaluate the null hypothesis that the wind speed on days 25 and 50 can be modeled by the same underlying distribution. The test is conducted at a significance level α and corresponding confidence level $1 - \alpha$, and an important statistic is p which is the probability of observing a test statistics D that is equal to or larger than that calculated for the samples given the null hypothesis. That is, $p \in [0,1]$ is the minimum significance level, corresponding to the maximum confidence level, at which the null hypothesis (that the distributions are the same) can be rejected. Therefore large values of p indicate a low confidence that the distributions differ, and low values of p indicate that the distributions differ with high confidence.

While the KS2 test provides a method for evaluating temporal changes in the marginal distribution of the wind speed, there are numerous situations in which the main interest of an analyst or modeler may be only in the second moment properties (mean and variance) of the wind speed. Consider the sequence $\{\mu_i\}$, $i = 1,\ldots,8760$ in which $\mu_i = E[V_i]$, and the corresponding sequence $\{\sigma_i\}$, $i = 1,\ldots,8760$ in which $\sigma_i^2 = E[V_i^2] - \mu_i^2$ which define the hourly mean and variance of the wind speed sequence. In practice the entries of these sequences will be estimated from data by, for example, $\widehat{\mu}_i = n_y^{-1}\sum_{j=1}^{n_y} v_{ij}$ and $\widehat{\sigma}_i^2 = (n_y - 1)^{-1}\sum_{j=1}^{n_y} v_{ij}^2 - \widehat{\mu}_i^2$.

In this paper, models are fit to the sequences $\widehat{\mu}_i$ and $\widehat{\sigma}_i$ and the time-gradients of the models are used to assess the degree of temporal non-stationarity present in the wind records. The first model used is meant to represent the periodic, annual fluctuations in the mean and standard deviation of the wind speed, namely,

$$
\begin{aligned}
\mu_h(t) &= \mu_w + b_1\cos\left(\frac{2\pi t + 2\pi c_1}{8760}\right) \\
\sigma_h(t) &= \sigma_w + b_2\cos\left(\frac{2\pi t + 2\pi c_2}{8760}\right)
\end{aligned}
\tag{1}
$$

where $\mu_w = 8760^{-1}\sum_{i=1}^{8760}\widehat{\mu}_i$ and $\sigma_w = 8760^{-1}\sum_{i=1}^{8760}\widehat{\sigma}_i$ are the temporal mean values of the mean and standard deviation of the wind speed, b_1 and b_2 are the amplitudes of the fluctuations, and c_1 and c_2 are the phase shifts. The parameters b_1, b_2, c_1, c_2 are determined for a particular data set through a nonlinear least squares regression. The time parameter t in the above is expressed in hours.

The second type of model used here is what might be called a local linear model and is essentially a linear regression to the sequences $\{\widehat{\mu}_i\}$ and $\{\widehat{\sigma}_i\}$ over a fixed interval of time. Specifically, the local linear models considered here have the form

$$
\begin{aligned}
\mu_{l,i}(t) &= d_{1,i} + e_{1,i}t \\
\sigma_{l,i}(t) &= d_{2,i} + e_{2,i}t
\end{aligned}
\tag{2}
$$

where the parameters $d_{1,i}$, $e_{1,i}$, $d_{2,i}$, $e_{2,i}$ are determined by a least squares regression to the sequences $\left\{\widehat{\mu}_j\right\}, j\in[i-336, i+336]$ and $\{\widehat{\sigma}_j\}, j\in[i-336, i+336]$. That is, the local linear models are linear fits to the estimates of the mean and standard deviation of the wind speed over a 672 h (one month) period centered on day i. In principle one could define local linear models over intervals other than 672 h. As will be discussed in the following sections, the 672 h interval has been found to provide a good balance between reduction in estimation error due to a large enough sample size on which to perform regression and a short enough interval that a local linear model is logical given the magnitude of the second derivative of the mean wind speed and standard deviation of the wind speed.

The models introduced have significant similarities with the so-called generalized autoregressive conditional heteroschedastic (GARCH) models for time varying mean and volatility of time series that have been applied to wind speed records [18–20]. The emphasis of this paper is on using simple but reasonable models to evaluate the period of validity of an assumption of stationarity, rather than on assessing the quality of the models themselves. Therefore, while GARCH models would reasonably be expected to provide better fits to and predictions of the wind speed, they are unlikely to lead to significantly different conclusions regarding periods of stationarity. One possible difference that might arise if GARCH models were used would be that, by virtue of the additional terms included in GARCH models, different periods of stationarity might be found at different times of the year. This would be in interesting question for further investigation.

In assessing whether an assumption of temporally local mean-square stationarity over some time interval would be admissible in analysis, the rate of change of the second moment properties is the quantity of primary interest. It is useful to define normalized versions of these rates of change that are essentially the first derivatives of the mean and standard deviation of the wind speed. Define these normalized quantities by

$$
\begin{aligned}
S_{\mu_h,p,i} &= \frac{d\mu_h(t)}{dt}\Big|_{t=i}\frac{T_p}{\mu_w} \approx \frac{\mu_h(i+T_p/2) - \mu_h(i-T_p/2)}{T_p}\frac{T_p}{\mu_w} \\
S_{\sigma_h,p,i} &= \frac{d\sigma_h(t)}{dt}\Big|_{t=i}\frac{T_p}{\sigma_w} \approx \frac{\sigma_h(i+T_p/2) - \sigma_h(i-T_p/2)}{T_p}\frac{T_p}{\mu_w} \\
S_{\mu_l,p,i} &= e_{1,i}\frac{T_p}{\mu_w} \\
S_{\sigma_l,p,i} &= e_{2,i}\frac{T_p}{\sigma_w}
\end{aligned}
\tag{3}
$$

where $p \in \{w,m\}$ and $T_p \in \{168h, 672h\}$ corresponding to weekly and monthly periods, the subscripts $\{h,l\}$ indicate the normalized derivative of the harmonic model and local linear models centered at day i respectively. Furthermore, one can define the difference in these normalized slopes

$$
\begin{aligned}
E_{\mu,p,i} &= \left|S_{\mu_h,p,i} - S_{\mu_l,p,i}\right| \\
E_{\sigma,p,i} &= \left|S_{\sigma_h,p,i} - S_{\sigma_l,p,i}\right|
\end{aligned}
\tag{4}
$$

for use in evaluating the quality of the harmonic model.

3. Data sources and characteristics

The government of the United States, through the National Oceanic and Atmospheric Administration, maintains a network of meteorological and ocean monitoring stations [21] through the National Data Buoy Center (NDBC), and the data used in this study is drawn from the NDBC database. Three locations were selected for examination in this study (see Table 1) based on the length of historical meteorological data available and their locations. Two sites have been selected that are geographically close to one

Table 1
Location characteristics of wind data source sites.

Station name	Anemometer			
	Type	Elevation	Location	Distances
Matinicus Rock	Island	22.9 m	43.783 N 68.855 W	107 km/945 km
East Hue & Cry Rock	Buoy	5 m	43.531 N 70.144 W	107 km/867 km
Virginia Beach	Buoy	5 m	36.611 N 74.842 W	945 km/867 km

Table 2
Wind data record characteristics.

Station name	Averaging interval	Total years	Thinned years
Matinicus Rock	2 min	28	19
East Hue & Cry Rock	8 min	30	18
Virginia Beach	8 min	22	12

another, while the third site is distant from the first two while remaining within the region of the US Atlantic coast where wind energy development is being actively pursued. Note that no attempt has been made to adjust the data to account for the different anemometer heights or the fact that one of the stations is on a small island. Although elevation and local topography affects the average wind speed and turbulence characteristics, the main interest of this paper is in the rate of change of the statistical characteristics of the wind speed over days, weeks, and months, and those temporal changes are not likely to be affected by such factors as elevation and topography.

Table 2 gives some important characteristics of the data. The 'thinned' number of years of data represents the number of years for which the wind speed record as downloaded had less than 10% missing or corrupted data. The averaging interval represents the length of time over which wind speed is averaged, and then reported once hourly. For example, at Matinicus Rock, the reported hourly wind speeds represent the average of the wind speed over the last 2 min of each hour. Again, the difference in averaging interval is not expected to dramatically affect investigations regarding statistical stationarity over days, weeks, and months. Data from years in which the fraction of good data is less than 0.90 have been neglected, and therefore n_y in the definitions of the previous section is always equal to the number of thinned years.

During the years retained by the thinning process, there remains the possibility of up to 10% of the data being missing or otherwise corrupted. In such cases, linear interpolation between the previous and next good data points has been used to infill the data set.

Fig. 1 shows a single year of data from Matinicus Rock along with the histogram and the Fourier spectrum generated by treating all wind measurements recorded at Matinicus Rock as samples of a single random variable. The histogram neglects non-stationarity and is meant to simply give an idea of the overall distribution shape—significant positive skewness—of the wind data. The spectrum shows clearly that this data contains significant periodicity at the 1 year, 1 day, and 12 h periods. Recall that the 1 day and 12 h periodicities are neglected here since primary interest lies in the longer term behavior of the wind speed. Data for the other sites are qualitatively similar and are therefore not shown graphically since aggregate statistics are not the focus of this paper. The aggregate statistics shown in Table 3 show fundamentally similar characteristics, and the higher values of each statistic at Matinicus Rock can be plausibly attributed to the higher elevation of the measuring station and the effect of the local island topography.

4. Results and discussion

This section presents the results of the investigation into the statistical stationarity of the three wind speed records described above. Tests for stationarity in distribution are discussed first, followed by discussion of tests for stationarity in mean square sense.

4.1. Distribution tests

Consider first the results of the KS2 test conducted on the samples $(\{v_{m,i}\},\{v_{m,j}\}), i,j = 1,\ldots,365$. Such a test returns a 1 if the null hypothesis that $V_{m,i}$ and $V_{m,j}$ have the same distribution is rejected with 0.95 confidence. In effect, then, the 365^2 tests provide a pairwise evaluation of the similarity of the monthly aggregate distribution of wind speed for month-long intervals centered on all possible pairs of days in the year. By definition the test will not be able to reject the null hypothesis at any level of confidence for $i = j$. In Fig. 2 the results of these KS2 tests are shown graphically, with a white pixel at location (i,j) in the figure signifying that the

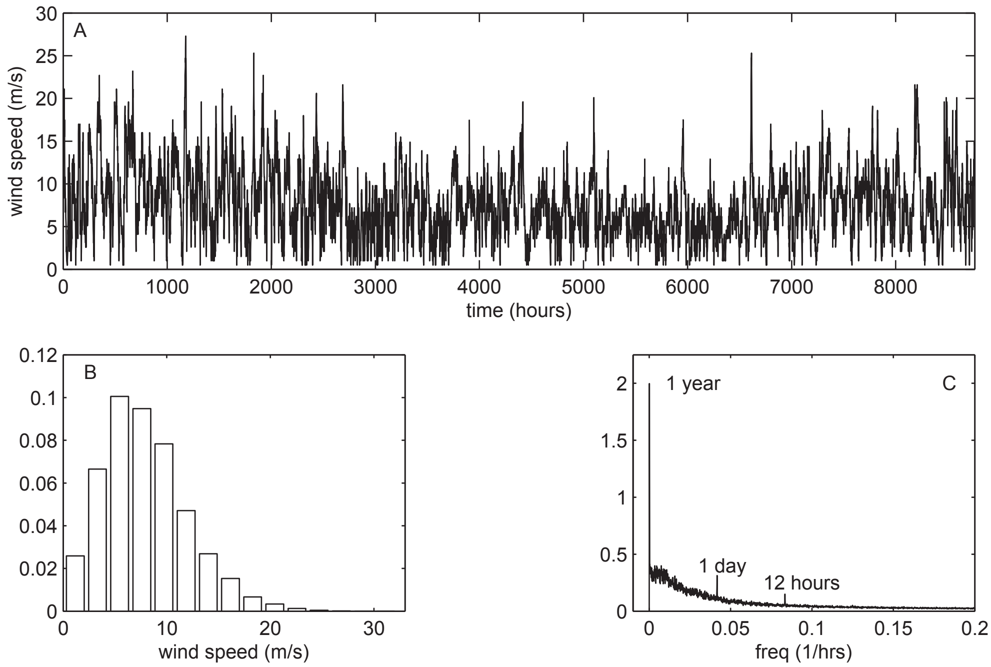

Fig. 1. A: One year wind speed record from Matinicus Rock, Maine. B: Aggregate histogram of Matinicus Rock wind speed data. C: Fourier spectrum of Matinicus Rock wind speed data.

Table 3
Aggregate statistics of wind records.

Location	Mean (m/s)	Std. dev. (m/s)	Skewness	Kurtosis
Matinicus Rock	7.9	4.1	0.74	3.6
East Hue	5.6	3.2	0.63	3.2
Virginia Beach	6.1	3.4	0.64	3.0

distributions $V_{m,i}$ and $V_{m,j}$ are likely to be different from one another. As an example, if one wanted to identify whether the wind speed distribution on January 31 and Feb 28 were plausibly similar—this would imply stationarity between January 31 and February 28—one would look to the point in the figure corresponding to day 31 (January 31) along the horizontal axis and 59 (February 28) along the vertical axis. If the pixel at that point in the figure is black then the KS2 test has not rejected the hypothesis that the distributions are the same, and the distributions can be said to be plausibly similar. In contrast, if the pixel is white, the KS2 test has rejected the hypothesis of similar distribution and some degree of non stationarity has been shown.

The diagonal along the line $i = j$ indicates, as would be expected, that monthly aggregates centered on neighboring days are likely to have the same distribution. The width of the band of black pixels around $i = j$ indicates the number of days over which the distribution of the wind speed, estimated from a monthlong data series, could be assumed to be constant. If the wind speed process were stationary, we would expect all (or nearly all due to estimation error) of the pixels to be black. For the results shown, this period of stationarity varies from as little as 1 day to as many as 20 days, with an average of approximately 13 days. One must keep in mind that the collections of samples $\{v_{m,i}\}$ and $\{v_{m,j}\}$ contain data in common if $|i - j| < 28$. The results appear qualitatively similar for the three sites, and the scattering of findings of similar distribution away from the diagonal are viewed as chance occurrences. For essentially all of the cases where $|i - j| > 28$ distribution similarity is rejected with 0.95 confidence. In summary, the data shown in Fig. 2 leads to the following conclusion: "if a one month series of wind speed data can be collected at a particular site, the distribution estimated from that data should be considered to be reflective only of the period of measurement or a period shifted forward or backward in time by not more than several days."

Statistical tests similar to those described immediately above, but on pairs of weekly data aggregations $\{v_{w,i}\}$ and $\{v_{w,j}\}$ and daily data aggregations $\{v_{d,i}\}$ and $\{v_{d,j}\}$ were performed. Fig. 3 shows the results (following the same graphical conventions as Fig. 2) for the weekly aggregations of data. A similar structure to the results as that in Fig. 2 for the monthly aggregations appears in the weekly comparisons with significantly more similarity findings (black pixels) away from the diagonal and weak evidence of a secondary structure in the data appearing as a band of modest density of similarity findings running from the upper left to the lower right. This secondary structure will be discussed in more detail in the context of the daily aggregate data presented next. The width of the main diagonal band (corresponding to the number of days over which the distribution of weekly data aggregations remains nearly constant) in this case ranges from approximately 3 days to approximately 25 days, with an average of approximately 7 days. Since the case where $|i - j| > 7$ for weekly aggregations corresponds to a comparison of weeks that share no data in common, these results indicate a stronger temporal persistence of the distribution of weekly aggregate data than monthly aggregate data. Specifically, it is likely justified to use a distribution estimated from an ensemble of one week observations to model wind speeds during adjacent weeks. As is the case for the monthly aggregations, the findings are qualitatively similar for each of the three sites.

The comparisons of daily aggregations of data (Fig. 4) reveal the deeper structure in the temporal variation of the distribution of wind speed. Recalling that black pixels in the KS2 matrices indicate the likelihood of distribution similarity for the days i and j corresponding to the location of that pixel in the matrix, the results show two distinct blocks of time in which there is a large fraction of findings of similar distribution. Consider first the Matinicus Rock data. The dark block at the center of the images indicates that during a 'summer' season stretching from approximately day 110 (late April) to day 280 (early October) there is a high likelihood that the daily distribution of wind speed cannot be distinguished at 0.95 confidence level. Noting the periodicity of the matrices—i and $i + 365$ would denote the same day of the year—the dark blocks at the corners of the image form a single region that stretches roughly from day 280 to day 110 of the year. This corresponds roughly to a 'winter' season in the data. What is striking about the results is that the boundaries of these blocks are quite clear. For example, the distribution of wind speeds at say 115 is likely to be similar to the distribution of wind speeds for days forward to day 280, but backward only to day 110. There also does not appear to be a gradual diminution of the likelihood of distribution similarity near the boundaries of these blocks. The structure is most pronounced in the Matinicus Rock data (Fig. 5), though it is still obviously present in the East Hue data. Though the structure is evident in the Virginia Beach data, the boundaries of the summer and winter blocks are relatively poorly defined. Given this description of the daily aggregate data, the traces of a similar structure can be seen in the weekly aggregate data of Fig. 3. The apparent 'noisiness' of the data—the presence of many rejections of distribution similarity denoted by white pixels—can be attributed to the sensitivity of the KS2 test to sample size and to statistical estimation error. As the aggregation period shortens from months to weeks to days the number of samples in each data set shrinks from $672n_y$ to $168n_y$ to $24n_y$, meaning both that the KS2 test loses confidence and that

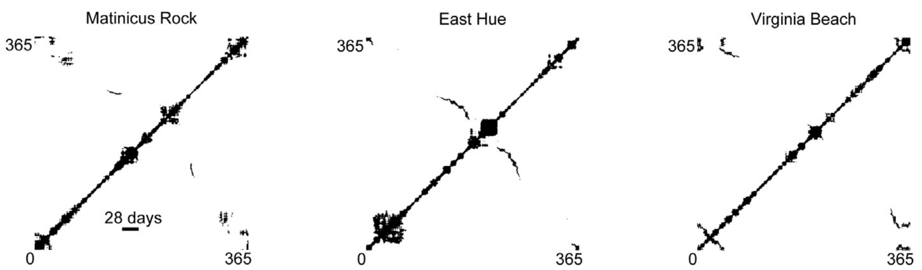

Fig. 2. Pairwise KS2 tests results on distributions estimated from monthlong intervals of wind speed data. White pixels represent rejection of the null hypothesis of similar distribution at confidence level 0.95, and therefore black pixels can be interpreted as indicating a likelihood of distribution similarity. A 28 day scalebar is included in the leftmost frame of the figure.

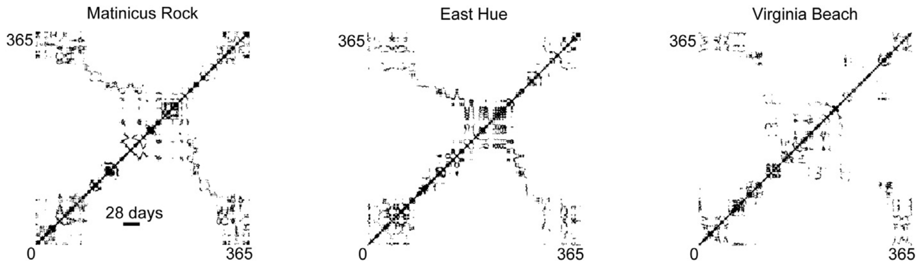

Fig. 3. Pairwise KS2 tests results on distributions estimated from weeklong intervals of wind speed data. White pixels represent rejection of the null hypothesis of similar distribution at confidence level 0.95, and therefore black pixels can be interpreted as indicating a likelihood of distribution similarity. A 28 day scalebar is included in the leftmost frame of the figure.

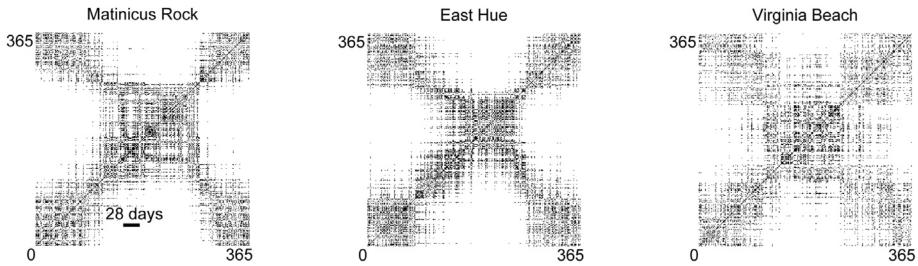

Fig. 4. Pairwise KS2 tests results on distributions estimated from daylong intervals of wind speed data. White pixels represent rejection of the null hypothesis of similar distribution at confidence level 0.95, and therefore black pixels can be interpreted as indicating a likelihood of distribution similarity. A 28 day scalebar is included in the leftmost frame of the figure.

simple estimation error may lead to a rejection of distribution similarity. Nevertheless, the strong concentration of findings of distribution similarity within the two seasonal time blocks is strong evidence for the presence of such an underlying structure in the data. In summary, it appears that there are two periods of the year demarcated roughly by days 110 and 280, during which the distribution of the wind speed can be reasonably assumed to be stationary if a modest degree of approximation or error in results stemming from that assumption can be tolerated. This finding has strong support for sites off the coast of Maine, and weaker support for the Virginia site.

4.2. Mean square stationarity

This section describes the results of investigation into the time scales over which the second moment properties of the wind records can be assumed to be nearly stationary. The approach involves estimation of the normalized rates of change of the mean and variance of the wind speed as represented by the harmonic and local linear models (Eqs. (1) and (2)). Table 4 gives the parameters and confidence intervals for the harmonic model. For a graphical representation of the relationship between the various models and the data for a characteristics month of the Matinicus Rock data see Fig. 6. In the figure, the harmonic model and local linear models are shown in solid lines and the dashed lines demarcate the amount of change over periods of one month (for the local linear model) and one week (for the harmonic model) in the model prediction of the mean wind speed. To obtain the normalized rates of change (Eq. (3)) one must further divide these amounts of change by the aggregate mean of the wind speed μ_W. Note also that to preserve clarity, and for illustrative purposes, the harmonic model is used to illustrate the weeklong change and the local linear model the monthlong change. Each of the weeklong and monthlong normalized rates of change have been calculated and are reported for both sets of models. Similar calculations are made for the standard deviation of the wind speed.

The results of this analysis are shown in Tables 5 and 6 and should be interpreted in the following way: the numerical entries represent the fraction of the aggregate mean (or standard deviation) wind speed by which the harmonic or local linear model changes over the monthlong or weeklong period examined. For example, at Matinicus Rock $\text{mean}[|S_{\mu_h,m}|] = 0.076$ and

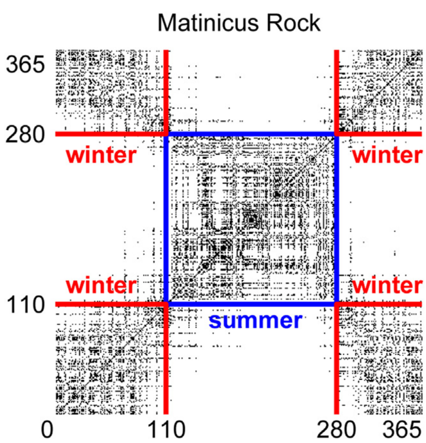

Fig. 5. Detail of Matinicus Rock pairwise daily KS2 comparisons showing seasonality.

Table 4
Harmonic model parameters. For each location and parameter the best fit and 95% confidence intervals are given.

Location	Mean		Standard dev.	
	Amplitude b_1 (m/s)	Phase (hrs)	Amplitude (m/s)	Phase (hrs)
Matinicus Rock	1.97 ± 0.028	−400 ± 20	1.04 ± 0.019	−514 ± 26
East Hue	1.62 ± 0.025	−253 ± 20	0.73 ± 0.15	−248 ± 30
Virginia Beach	1.62 ± 0.029	−292 ± 25	0.64 ± 0.020	51 ± 43

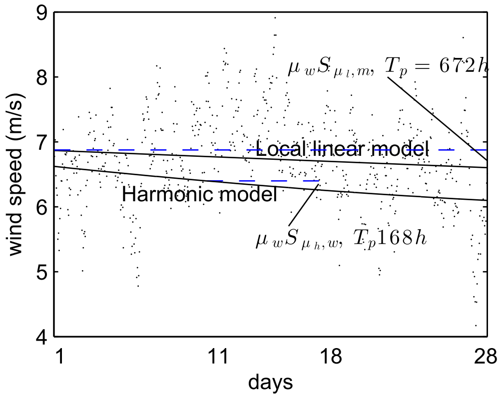

Fig. 6. Graphical representation of time varying mean wind speed models and definitions of the normalized rates of change. Definitions are analogous for the standard deviation data. Data points in the figure represent the ensemble average ($n_y = 19$ years for Matinicus Rock) of the hourly wind speed data.

$\max[|S_{\mu_h,m}|] = 0.12$ meaning that the harmonic model, on average, predicts a 7.6% change in the mean wind speed over the course of a 672 h period, and that the maximum of all these monthly changes is 12%. Another way to interpret this example is that an assumption of mean square stationarity for 672 h period would result in an average error in mean wind speed of 7.6% and a maximum error in wind speed of 12%.

The first key observation from the results is that the 672 h (monthlong) period appears to be too long to justify an assumption of mean square stationarity in the wind speed. Normalized rates of change for the 672 h period are on the order of 0.08–0.12, which is a larger error than would normally be acceptable in a stochastic analysis. For the 168 h (weeklong) periods, however, the normalized rates of change are on the order of 0.02–0.03, degrees of divergence from stationarity that would seem largely acceptable in light of the presence of other sources of uncertainty and error in the analysis of engineering systems that depend on wind speed.

Furthermore, the harmonic model consistently delivers lower mean and maximum normalized rates of change of the second moment properties if compared to the local linear model. Without being able to decouple the inherent non-stationarity in the data from the estimation error of the local linear model it is difficult to arrive at a strong conclusion regarding the relative merits of the models in this context. However, an investigation of a local linear model fit to weeklong instead of monthlong data sets found a sharp increase in the mean normalized rates of change that was inconsistent with the actual seasonal non-stationarity and consistent with large estimation error. Therefore, it appears better to attach primary weight to the results obtained from the harmonic model.

Finally, there is strong consistency in the results across the three sites selected, with the Virginia Beach site showing somewhat stronger non-stationarity in the local linear model results. This is most likely due to the smaller number of years included in the Virginia Beach data set (12 versus 18 or 19 for the Maine sites) leading to greater estimation error.

Table 5
Statistics of normalized slope values for harmonic and local linear models for the mean wind speed at three sites.

Location	$\lvert S_{\mu_h,m}\rvert$		$\lvert S_{\mu_h,w}\rvert$		$\lvert S_{\mu_l,m}\rvert$		$\lvert S_{\mu_l,w}\rvert$	
	Mean	Max	Mean	Max	Mean	Max	Mean	Max
Matinicus Rock	0.076	0.12	0.019	0.030	0.089	0.28	0.022	0.070
East Hue	0.089	0.14	0.022	0.035	0.11	0.31	0.028	0.077
Virginia Beach	0.081	0.13	0.020	0.032	0.13	0.40	0.033	0.10

Table 6
Statistics of normalized slope values for harmonic and local linear models for the standard deviation of wind speed at three sites.

Location	$\lvert S_{\sigma_h,m}\rvert$		$\lvert S_{\sigma_h,w}\rvert$		$\lvert S_{\sigma_l,m}\rvert$		$\lvert S_{\sigma_l,w}\rvert$	
	Mean	Max	Mean	Max	Mean	Max	Mean	Max
Matinicus Rock	0.078	0.12	0.019	0.030	0.097	0.26	0.024	0.066
East Hue	0.069	0.11	0.017	0.027	0.089	0.33	0.022	0.084
Virginia Beach	0.058	0.092	0.015	0.023	0.12	0.47	0.029	0.12

5. Conclusions

This paper has addressed the issue of statistical stationarity in offshore wind speed records with an emphasis on fluctuations at time scales longer than one day. Motivated by the need for wind speed models that can be used in the assessment of the offshore wind energy resource and the generating potential of sites identified for wind energy development, the paper seeks to identify time scales over which the marginal distribution or second moment properties of the wind speed record can be assumed to be stationary, or time-invariant. The well known diurnal variation of wind speed is neglected in this study.

The paper presents a framework that is based on statistic tests for the similarity of distributions to evaluate the period of time over which the marginal distribution of the wind speed can be assumed stationary, and a similar framework for evaluating second moment stationarity that is based on fitting mathematical models to the non-stationary data and evaluating the rate of change of the second moment properties predicted by such models.

Key findings are that an assumption of stationarity of the wind speed process over a time period of approximately a week is reasonably well justified, but that such an assumption for a monthlong period would lead to large prediction errors. Although reasonably strict notions of stationarity can be supported only over period up to one week, the data show a distinct division of the year into two period of time (roughly winter and summer) during which the marginal distribution of the wind speed is relatively consistent. The marginal distribution changes abruptly at the ends of such periods, but relatively gradually within the periods. Finally, the paper has presented numerical examples based on data collected at three sites off the Atlantic coast of the United States, and finds highly consistent results across the three sites, which range from the states of Maine to Virginia.

Acknowledgments

Arwade would like to acknowledge the financial support of the US National Science Foundation through grant CMMI-1234656.

References

[1] Vanmarcke E. Random fields: analysis and synthesis. Cambridge: MIT Press; 1983.
[2] Grigoriu M. Applied non-Gaussian processes. New Jersey: Prentice Hall; 1995.
[3] I.-I. E. Commission, IEC 61400-3 Ed.1.2009 Wind turbines-part 3: design requirements for offshore wind turbines. International Electrotechnical Commission, Geneva, Switzerland, 2009.
[4] I.-I. E. Commission, IEC 61400–1 Ed.3.2005 Wind turbines-part 1: design requirements, International Electrotechnical Commission, Geneva, Switzerland, 2009.
[5] Van der Hoven I. Power spectrum of horizontal wind speed in the frequency range from 0.007 to 900 cycles per hour. J Meteorol 1957;14:160–4.
[6] Carta JA, Ramirez P, Velazquez S. A review of wind speed probability distributions used in wind energy analysis case studies in the canary islands. Renew Sustain Energy Rev 2009;13(5):933–55. pT: J; UT: WOS: 000266683800001.

[7] Coles S, Casson E. Extreme value modelling of hurricane wind speeds. Struct Saf 1998;20(3):283–96. pT: J; UT: WOS:000077100600006.
[8] Heckert N, Simiu E, Whalen T. Estimates of hurricane wind speeds by "peaks over threshold" method. J Struct Eng ASCE 1998;124(4):445–9. pT: J; UT: WOS:000072681700012.
[9] Seguro J, Lambert T. Modern estimation of the parameters of the weibull wind speed distribution for wind energy analysis. J Wind Eng Ind Aerodyn 2000;85(1):75–84. pT: J; UT: WOS:000085917100005.
[10] Carta JA, Velazquez S, Matias JM. Use of bayesian networks classifiers for long-term mean wind turbine energy output estimation at a potential wind energy conversion site. Energy Convers Manag 2011;52(2):1137–49. pT: J; UT: WOS: 000285485100039.
[11] Justus C, Mani K, Mikhail A. Interannual and month-to-month variations of wind speed. J Appl Meteorol 1979;18(7):913–20. pT: J; UT: WOS: A1979HF00900009.
[12] Morgan V. Statistical distributions of wind parameters at sydney, australia. Renew Energy 1995;6(1):39–47. pT: J; UT: WOS: A1995QV59700007.
[13] Ramirez P, Carta J. Influence of the data sampling interval in the estimation of the parameters of the weibull wind speed probability density distribution: a case study. Energy Convers Manag 2005;46(15–16):2419–38. pT: J; UT: WOS: 000229413300007.
[14] Morgan EC, Lackner M, Vogel RM, Baise LG. Probability distributions for offshore wind speeds. Energy Convers Manag 2011;52(1):15–26. pT: J; UT: WOS:000284746800003.
[15] Rogers A, Rogers J, Manwell J. Comparison of the performance of four measure-correlate-predict algorithms. J Wind Eng Ind Aerodyn 2005;93(3): 243–64. pT: J; UT: WOS:000227965400003.
[16] Lackner MA, Rogers AL, Manwell JF, McGowan JG. A new method for improved hub height mean wind speed estimates using short-term hub height data. Renew Energy 2010;35(10):2340–7. pT: J; UT: WOS:000278664500024.
[17] Pryor SC, Barthelmie RJ. Climate change impacts on wind energy: a review. Renew Sustain Energy Rev 2010;14(1):430–7. pT: J; UT: WOS: 000271279100031.
[18] Liu H, Erdem E, Shi J. Comprehensive evaluation of armagarch(-m) approaches for modeling the mean and volatility of wind speed. Appl Energy 2011;88:724–32.
[19] Ewing B, Kruse J, Schroeder J. Time series analysis of wind speed with time-varying turbulence. Environmetrics 2006;17:119–27.
[20] Taylor J, McSherry P. Wind power density forecasting using ensemble predictions and time series models. IEEE Trans Energy Conserv 2009;24:775–82.
[21] NOAA, Ndbc, National Data Buoy Center. http://www.ndbc.noaa.gov/station_page.php?station=mism1.