



## The effect of averaging, sampling, and time series length on wind power density estimations

Gross, Markus; Magar, Vanesa; Peña, Alfredo

*Published in:*  
Sustainability (Switzerland)

*Link to article, DOI:*  
[10.3390/SU12083431](https://doi.org/10.3390/SU12083431)

*Publication date:*  
2020

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Gross, M., Magar, V., & Peña, A. (2020). The effect of averaging, sampling, and time series length on wind power density estimations. *Sustainability (Switzerland)*, 12(8), Article 3431. <https://doi.org/10.3390/SU12083431>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Article

# The Effect of Averaging, Sampling, and Time Series Length on Wind Power Density Estimations

Markus Gross <sup>1,\*</sup>, Vanesa Magar <sup>1,†</sup> and Alfredo Peña <sup>2,†</sup>

<sup>1</sup> CICESE, Carretera Ensenada-Tijuana No. 3918, Zona Playitas, CP. 22860 Ensenada, B.C., Mexico; vmagar@cicese.edu.mx

<sup>2</sup> Department of Wind Energy, Technical University of Denmark, Frederiksborgvej 399, 4000 Roskilde, Denmark; aldi@dtu.dk

\* Correspondence: mgross@cicese.mx

† These authors contributed equally to this work.

Received: 2 March 2020; Accepted: 13 April 2020; Published: 22 April 2020



**Abstract:** The Wind Power Density (WPD) is widely used for wind resource characterization. However, there is a significant level of uncertainty associated with its estimation. Here, we analyze the effect of sampling frequencies, averaging periods, and the length of time series on the WPD estimation. We perform this analysis using four approaches. First, we analytically evaluate the impact of assuming that the WPD can simply be computed from the cube of the mean wind speed. Second, the wind speed time series from two meteorological stations are used to assess the effect of sampling and averaging on the WPD. Third, we use numerical weather prediction model outputs and observational data to demonstrate that the error in the WPD estimate is also dependent on the length of the time series. Finally, artificial time series are generated to control the characteristics of the wind speed distribution, and we analyze the sensitivity of the WPD to variations of these characteristics. The WPD estimation error is expressed mathematically using a numerical-data-driven model. This numerical-data-driven model can then be used to predict the WPD estimation errors at other sites. We demonstrate that substantial errors can be introduced by choosing too short time series. Furthermore, averaging leads to an underestimation of the WPD. The error introduced by sampling is strongly site-dependent.

**Keywords:** wind power density; wind speed time series; site characterization; numerical-data-driven model

## 1. Introduction

The need to replace fossil fuels and the continuously increasing demand for energy drive the harvesting of renewable energy sources. Amongst those, wind energy is currently one of the most efficient sources. Before new sites can be used for the development of wind energy projects, a resource assessment has to be carried out to identify potential sites and to predict the available energy. For such assessments, the wind power density is a useful metric because it characterizes the energy content of the wind climatology. It is a relatively simple metric, and hence, it can easily be computed for large areas, guiding the exploration of potential sites. It is also useful when attempting to predict future changes in wind resources using simulations of future climate scenarios.

While widely used and despite the understanding that averaging is detrimental and that high sampling frequencies and long time series improve the Wind Power Density (WPD) estimate, it is not well known how long the time series should be, what sampling/averaging is permissive, and what errors to expect in WPD estimations. Here, we address these questions.

The WPD is expressed as:

$$\text{WPD} = \frac{1}{2} \frac{1}{N} \sum_{i=1}^N \rho_i v_i^3, \quad (1)$$

where  $v$ ,  $\rho$ , and  $N$  denote the wind speed, air density, and the number of samples, respectively. It is generally computed from a time series of wind speed observations. WPD estimates depend on the length of the time series, averaging effects that reduce the weights of the velocity fluctuations, and sampling effects, i.e., the wind speed variability might not be adequately represented by the sampling used. Wind speed time series are discrete and are characterized by their sampling frequency and averaging period. If they are derived from model outputs, they may be “instantaneous” model realizations, meaning that they represent the model state at a particular time. Wind speed observations are generally recorded using cup and sonic anemometers operating at frequencies within the range 1–20 Hz, and the reported data are usually the time series of the mean values within each 10 min period; thus, 10 min is the usual sampling rate. Satellite data generally have a much lower sampling frequency, of around two measurements per day. Despite their low sampling frequency, satellite data have been used in various studies [1–8]. A review of the various data sources from remote sensing, reanalysis, and mesoscale models was presented by Wang et al. [9].

Mesoscale models are frequently used in wind resource characterization [10–12]. However, sampling in mesoscale model outputs is generally performed at low frequencies. Hourly data are commonly considered high frequency, and six hourly data are common, due to the massive amounts of data generated by these models. Lower sampling frequencies are probably not very useful for wind energy applications, as the details of the diurnal cycle would undoubtedly be missing [13,14].

Data derived from climate models are also frequently used for WPD estimation [15–20]. Climate models often only provide low sampling and generally generate temporally averaged outputs because this preserves the climate information without the storage burden. For wind energy, however, this is problematic, as we will show. The extent of the errors introduced by the sampling frequency in the WPD estimate is not well known. This study illustrates the effect of sampling frequency and averaging on the WPD estimate. A method that estimates the error and relates it to three wind speed time series statistical characteristics, namely the Weibull shape parameter, the mean wind speed, and the exponential autocorrelation decay base, is outlined here.

The paper is organized as follows. In Section 2.1, we introduce the observational, the model, and the artificial datasets and the three observational sites used in this study to illustrate the problem. In Section 2.2, we introduce the error norm applied throughout this publication. In Section 3, we then present the results of the analysis, first in general analytic terms, by evaluating the impact of averaging. We then focus on sampling frequency and time-series length, using observational time series and model data. This section concludes with the introduction of an artificial time series, from which a numerical-data-driven model is constructed to generalize the analysis. We provide this model as a Supplementary Materials to the article, which can be used by the reader to evaluate errors at other case sites of interest for wind energy resource exploitation. This model is compared to the error obtained by analyzing the observations. Lastly, in Section 4, we present some conclusions.

## 2. Data and Methodology

The analysis was performed using four approaches. In the first approach, we compared the errors in the WPD estimates when assuming a Gaussian and a Weibull distribution for the wind speed. In the second approach, we used observational data from meteorological stations and model outputs to analyze the effect of sampling at three different sites. Generally, the analysis of observational data has been limited because long wind speed time series are scarce or proprietary. Moreover, the findings are site-dependent and cannot be directly transferred to other sites. In the third approach, we used data from a numerical weather prediction model to analyze the effect of the length of the time series. Finally, in the fourth approach, we derived a numerical-data-driven model from the results of the

error analyses based on the artificial datasets, and the performance of this model was compared to the observational data from the third site.

### 2.1. Data

Three observational datasets were analyzed. The first observational dataset consisted of wind speed measurements at Østerild [21], Northern Denmark, taken over one year at a 106 m height. The second observational dataset consisted of measurements at La Rumorosa, Mexico. At La Rumorosa, which will be referred to as the M01 mast data, wind speeds were measured over one year at an 80 m height. The third observational dataset consisted of wind speed measurements at Høvsøre [22], Western Denmark, taken over a 14 year period at a 100 m height. All measurements were averaged and recorded every 10 min. Data were obtained from simulations using the Weather Research and Forecasting (WRF) model with the configuration of the Mexican Wind Atlas project [23] to explore the dependence of the WPD on the length of wind speed time series. Here, 10 m instantaneous winds, sampled at a frequency of one hour and evaluated at the nearest grid point to the Rumorosa M01 mast location, were used as a case study.

### 2.2. Methods

Meteorological data from the three sites were analyzed. Furthermore, artificial time series were generated. The analysis using these artificial time series provided the error due to sampling and averaging, which depended on the time series characteristics. These artificial data could then be evaluated, here using five-dimensional linear interpolation, to compute the error in WPD estimations for a range of time series characteristics, thus potentially representing a range of sites. This interpolation is the numerical-data-driven model. These artificial hourly wind speed time series were generated through a Markov random walk method combined with a Weibull distribution of the wind speed and a transition matrix following Veers and McEnerney [24]. These artificial hourly wind speed time series depended on the mean wind speed, the Weibull shape parameter, and the autocorrelation decay base. Sixty (60) wind speed time series with a sampling frequency of one hour and a 20 year length formed the basis of the artificial time series analysis. The 60 time series were comprised of the full permutation set for three parameters with the following values: a Weibull shape parameter of 1.9, 2.0, 2.2, or 2.4; a mean wind speed of 6.0, 10.0, or 14.0 m/s; and an exponential autocorrelation decay base of 1.4, 1.5, 1.6, 2.0, or 2.4.

The observational and artificial data were analyzed using different averaging periods,  $t_1$ , sampling intervals,  $t_2$ , and time series lengths,  $l$ . In the following, this is expressed as  $\bar{v}^{t_1}|_{t_2}$ , where the averaging period,  $t_1$ , appears as a superscript to the right of the averaged quantity, indicated with an over-line, and the sampling interval,  $t_2$ , appears as a subscript to the vertical bar, commonly used as “evaluated at.” The error in estimating the WPD, i.e., the WPD error, is here the difference between the theoretically optimum time series of several decades in length and high sampling frequency. Hence, the WPD error is defined as a function of  $t_1$ ,  $t_2$ , and  $l$ :

$$\epsilon(t_s, t_1, t_2, l) = \frac{1}{l/t_2} \sum_{t=t_s}^{t=t_s+l} \left( \bar{v}_t^{t_1} |_{t_2} \right)^3 / \frac{1}{l_{\text{ref}}/t_{2,\text{ref}}} \sum \left( \bar{v}^{t_1, \text{ref}} |_{t_{2,\text{ref}}} \right)^3, \quad (2)$$

and:

$$\epsilon(t_1, t_2, l) = \mathcal{M} [\mathcal{W} \{ \mathcal{A} (\epsilon(t_s, t_1, t_2, l)) \}], \quad (3)$$

where the subscript ref denotes the respective reference or best approximation, with time series length  $l_{\text{ref}}$ , averaging period  $t_{1,\text{ref}}$ , and sampling interval  $t_{2,\text{ref}}$ .  $\mathcal{M}$ ,  $\mathcal{W}$ , and  $\mathcal{A}$  are the monotone, windowing, and absolute operators. Unless explicitly stated, the operators are defined as follows. The absolute operator:

$$\mathcal{A}(x) = 1 + |x - 1|, \quad (4)$$

and the monotone operator:

$$\mathcal{M}(x) = \max(\max(x(t_1, t_2, l), x(t_1, t_2, l + \delta l)), x(t_1, t_2 - \delta t, l)), \quad (5)$$

with  $\delta t$  and  $\delta l$  denoting the step in the sampling interval and time series length used when evaluating the errors, respectively. In this study,  $\delta l$  was one year and  $\delta t$  one hour. The monotone operator enforces the fact that with a shorter time series length, the error has to be larger, and equally, a longer sampling period will also increase the error.

For each  $l \neq l_{ref}$ , the period of the time series with the largest error has to be used. Not using the largest error would lead to a potentially large underestimation and non-monotone behavior, as illustrated in Figure 1. To achieve this, a sliding window was applied through the windowing operator:

$$\mathcal{W}(x) = \begin{cases} x(t_s = 0) & \text{if } l \equiv l_{ref} \\ \max(x(t_s = ndt)) & \text{for all realizable } n \in \mathbb{N}, \end{cases} \quad (6)$$

with  $dt = 30$  days. Two implementations, using a sliding window search and only using the first  $l$  years, i.e., for  $n = 0$ ,  $\mathcal{W}(x) = x(t_s = 0)$ , are contrasted in Figure 1. The use of the windowing operator was motivated by the following three immediate observations that could be made by comparing the graphs in that figure. First, the error introduced by not using a sliding window analysis can be observed in the top left panel. The error for a five year time series was nearly as low as the error using the full 20 years. While this observation was valid for this illustrative synthetic dataset and the particular period analyzed, the occurrence of this minimum did not help derive an estimate of potential errors that could be incurred by using too short time series. The second observation can be made comparing the top-right with the top-left panel. In the top right panel, the sliding window transforms the surface into the expected monotonically declining surface, and the maximum error observed jumps from 1.08 (top left) to 1.5 (top right). The final observation can be made by comparing the top row with the bottom row. In the top row, the error does not change significantly with sampling frequency. The behavior in the bottom row is substantially different, showing an error of 30% compared to 12% with hourly sampling and a one year time series length, for example.

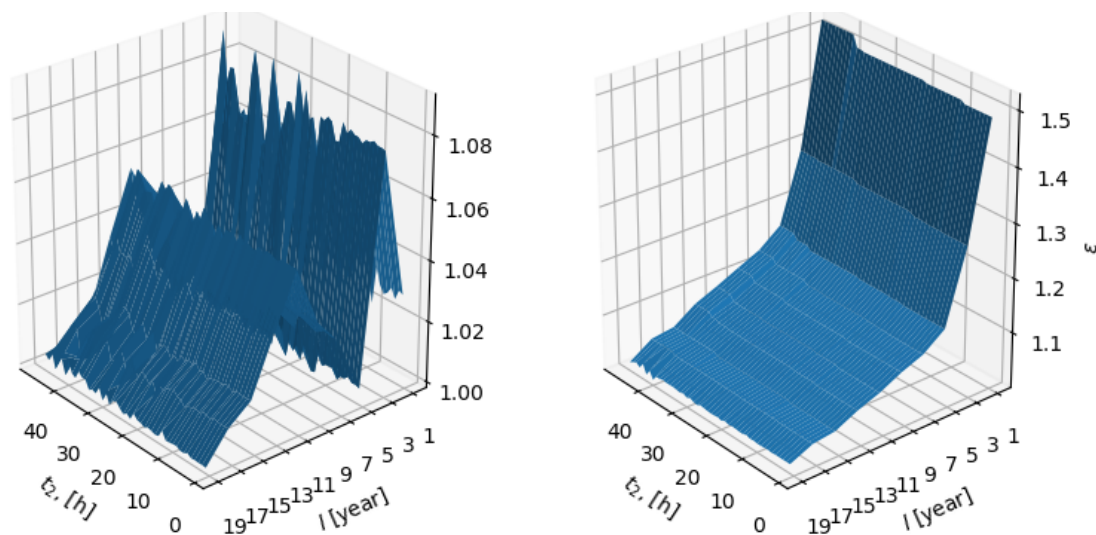
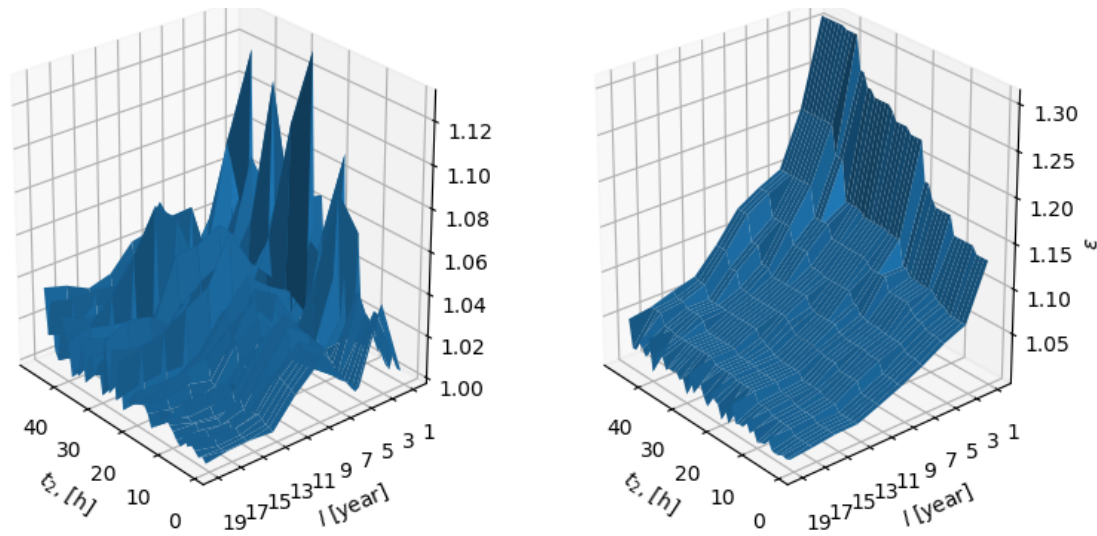


Figure 1. Cont.



**Figure 1.** Comparison of the absolute error using the first  $l$  years of the artificial time series for the error estimate (left) and using the period of largest error (right), for two different configurations: Weibull shape parameter of 2.6 (all frames), mean wind speed of 14 m/s (top frames) and 6 m/s (bottom frames), and autocorrelation decay base of 2.4 (top frames) and 1.6 (bottom frames). The figures on the right were also monotonized according to Equation (5).  $l_{ref} = 20$  years,  $t_{1,ref} = 0$  h, and  $t_{2,ref} = 1$  h in all panels.

### 3. Results and Discussion

We first illustrate the error of estimating the WPD as the cube of the mean wind speed and highlight how the asymmetry of the wind speed distribution amplifies this error. Then, using the observational data, we demonstrate the strong detrimental impact of averaging on WPD estimates. Due to the generally short time series available from observational data, we demonstrate using long time series from model outputs that seasonal changes strongly impact the WPD estimates. Due to the limitations of data availability, artificial wind speed time series were generated with different characteristics, mimicking different site climatologies. Apart from the expected dependence on the Weibull shape parameter, we also found a strong dependence on the autocorrelation decay base and the time series length.

#### 3.1. The Problem of the Cube of the Mean Wind Speed for WPD Calculation

Assuming that changes in air density are negligible, the error of estimating the WPD from the cube of the mean wind speed,  $\bar{v}^3$ , can be estimated as:

$$\epsilon = \frac{\bar{v}^3}{\int_0^\infty v^3 p(v) dv}, \tag{7}$$

where  $p(v)$  is the probability distribution (pdf) of the wind speed. For a Gaussian pdf, the error is:

$$\epsilon_G = \frac{\bar{v}^3}{\bar{v}^3 + 3\bar{v}\sigma_v^2} = \frac{(\bar{v}/\sigma_v)^2}{(\bar{v}/\sigma_v)^2 + 3} \tag{8}$$

where  $\sigma_v^2$  is the variance of the pdf. For such a distribution, the error is a function of both the mean wind speed and the variance of the pdf (see Figure 2, left). For a Weibull distribution, which is widely used to represent wind speed distributions [25–33], the error in the WPD estimation is:

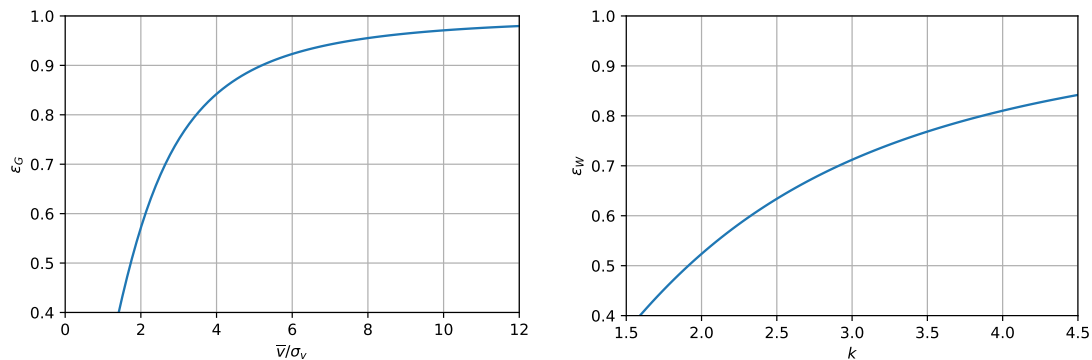
$$\epsilon_W = \frac{\bar{v}^3}{\lambda^3 \Gamma(1 + 3/k)}, \tag{9}$$



where  $\lambda$ ,  $k$ , and  $\Gamma$  are the scale, shape parameter, and Gamma function, respectively. Since for a Weibull distribution:

$$\bar{v} = \lambda \Gamma \left( 1 + \frac{1}{k} \right),$$

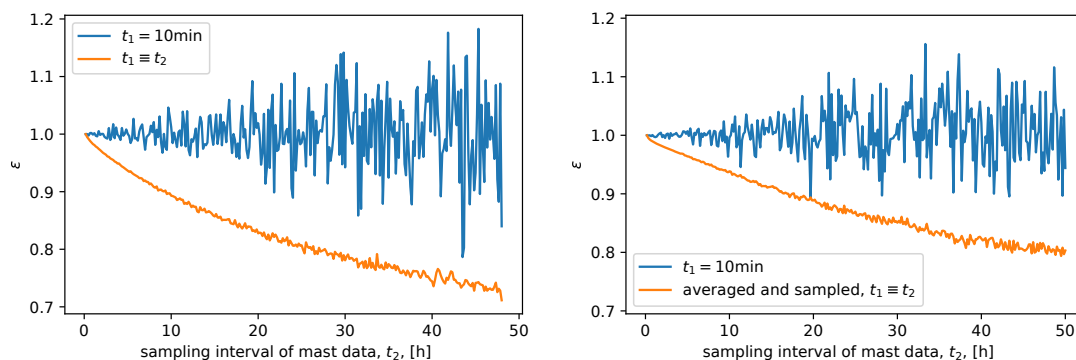
$\varepsilon_w$  is only dependent on  $k$ , as shown in Figure 2, right. From these plots, it is clear that the error is non-linear in the shape parameter and that the broader the wind speed distribution (smaller value of  $k$ ), the larger the error.



**Figure 2.** Averaging errors for Gaussian (left) and Weibull (right) wind speed distributions.

### 3.2. Impact of Sampling and Averaging

Using data from the M01 and the Østerild masts, we now analyze the dependence of the WPD on sampling and averaging by constructing subsets from the observational datasets using different sampling frequencies and averaging periods. Figure 3 shows the error with different sampling frequencies and averaging periods at the two masts, from 0 up to nearly 50 h to mimic the worst-case scenario.



**Figure 3.** Impact of sampling and combined sampling and averaging on the Wind Power Density (WPD) error using data from the M01 mast (left frame) and Østerild (right frame).  $l_{\text{ref}} = 1$  year and  $t_{1,\text{ref}} = t_{2,\text{ref}} = 10$  min.  $\mathcal{A}(x) = x$ ,  $\mathcal{M}(x) = x$ .

The combined sampling and averaging error was always below one, where one represents here the best possible estimation of WPD, i.e., using the wind speed time series constructed from 10 min averages and sampled every 10 min [34]. Using 24 h sampling/averaging periods resulted in an error between 10 and 15% for both observational datasets. If only the sampling frequency were changed, the error could be either larger or smaller than one. The reason for that was that the sampling selected a particular subset of data points from the distribution. This selection may result in one of three outcomes. The subset gave a good representation of the original wind speed distribution, leading to a similar WPD estimation and a small sampling error; the subset was biased towards higher wind

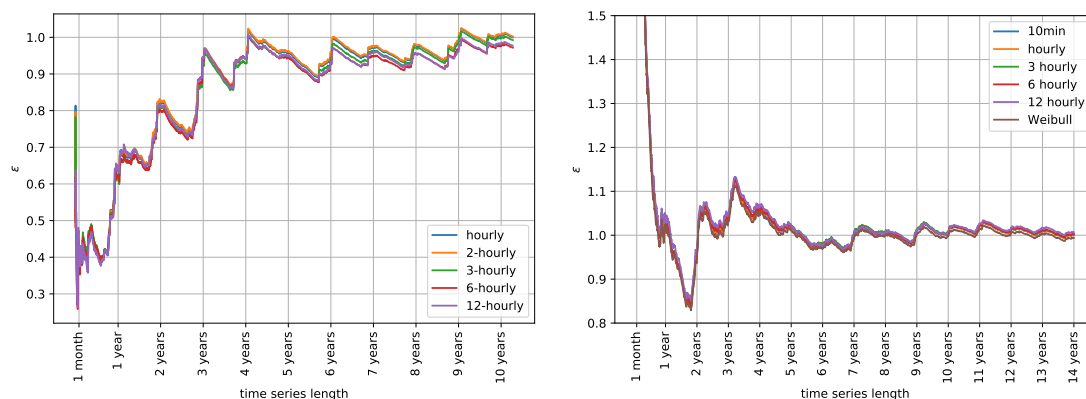
speeds, leading to an overestimation of the WPD and a sampling error over one, or the subset was biased towards lower wind speeds, leading to an underestimation of the WPD and a sampling error below one. If the sampling interval was small, around one or two hours, the error was below 1%. A 6 h sampling interval increased the error to approximately 2.5%, and for a 24 h interval, it increased to 10–15%. These two sites showed similar behavior for the WPD estimation error. We now need to address two further questions: first, What is the impact of the length of the time series on the WPD estimation, and second, what is the impact on the wind speed distribution characteristics on the WPD estimation?

### 3.3. Impact of the Length of the Time Series

The impact of the time series length was difficult to estimate because long time series are rare, and the error measure intrinsically assumed that the long time series used as the reference was indeed close to the exact WPD. We first estimated the WPD error using model data, then used artificial time series, and finally, compared the error obtained from the artificial time series with the error obtained from a long observational time series.

#### 3.3.1. Impact of the Length of the Time Series Using Model and Observational Data

In order to assess the impact of the length of the time series on WPD estimates, we analyzed hourly instantaneous WRF model outputs of wind speeds at a 10 m height for La Rumorosa over ten years from the grid point closest to the M01 mast and 14 years of measurements from the Høvsøre met mast. For different sampling frequencies, in our example 1, 2, 3, 6, and 12 h data, the error in the estimation of the WPD,  $\epsilon$ , is shown in Figure 4. The longer sampling interval introduced a bias, which was most probably a result of the impact of the diurnal cycle on the wind speed, which, with increasing sampling period, was less and less represented. The figure also shows that for the sampling intervals tested, we needed at least four years of data to reduce the error below 10% for the WRF data and 3.5 years for the Høvsøre data. Furthermore, the intra-annual cycle of the wind speed influenced the error, as one could find seasonal fluctuations, which, even with a nearly 10 year time series length, were around 5% for the WRF data. It is worth noting that only two sites were tested; at other sites, the required time series length to have errors below 10% may vary, depending on the local variability of the winds and climatology. The error behavior at other sites could be analyzed using the numerical-data-driven model presented next.

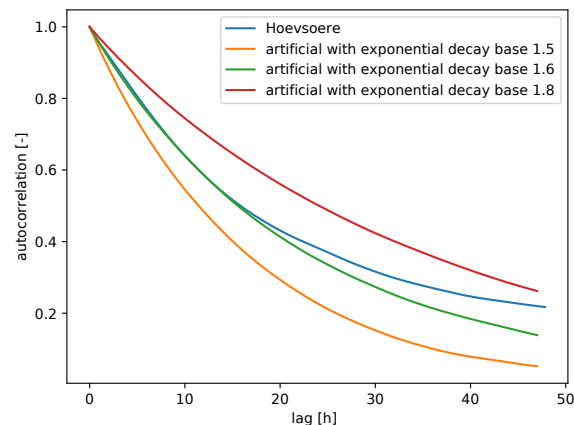


**Figure 4.** Impact of time series length on WPD estimates. Left frame: using wind speed at a 10 m height from a WRF simulation at the M01 met mast ( $l_{\text{ref}} = 10$  years,  $t_{1,\text{ref}} = 0$  h, and  $t_{2,\text{ref}} = 1$  h). Right frame: using wind speed at a 116.5 m height from the observations of the Høvsøre met mast ( $l_{\text{ref}} = 14$  years,  $t_{1,\text{ref}} = 0$  min, and  $t_{2,\text{ref}} = 10$  min). The line “Weibull” in the legend denotes the error estimation fitting the Weibull distribution to the 10 min time series. The maximum error of the WPD estimation for the Høvsøre met mast is 3.35 for 10 min sampling, 3.41 for 12 h sampling, and 3.39 for the Weibull fit, respectively, for a time series length of one month.  $\mathcal{A}(x) = x$ ,  $\mathcal{M}(x) = x$ ,  $\mathcal{W}(x) = x(t_s = 0)$ .



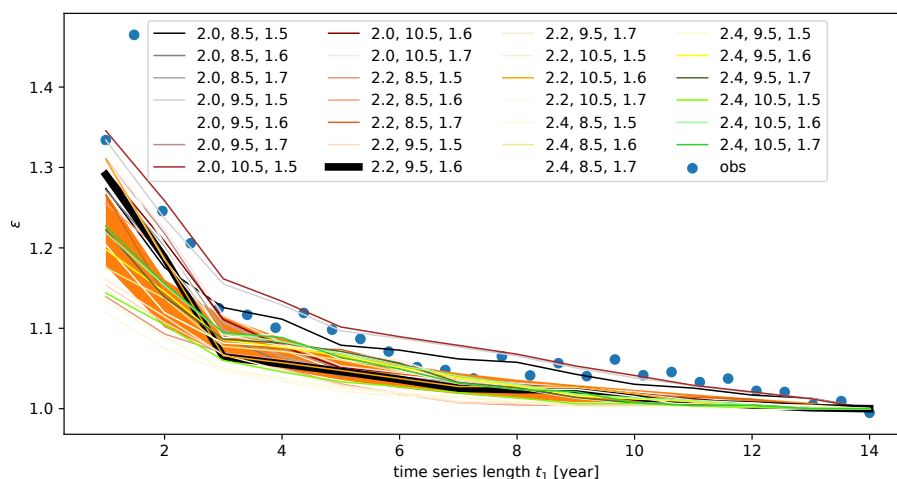
### 3.3.2. Impact of the Time Series Length Using an Artificial Time Series

The realism of the error derived from the artificial time series can be demonstrated by comparing the numerical-data-driven model (see Appendix A) with observations from Høvsøre. The observations from Høvsøre were reduced to hourly sampling to match the sampling of the artificial time series. At Høvsøre, the 14 year time series had a Weibull shape parameter of 2.19 and a mean wind speed of 9.47 m/s. By using these two parameters, the autocorrelation decay became a free parameter. Figure 5 shows the autocorrelation decay of several artificial time series together with that estimated from the Høvsøre data. An autocorrelation decay of 1.6 fit the observations well, at least within the range 0–20 h.



**Figure 5.** Autocorrelation decay as a function of the lag time for the Høvsøre observations and artificial time series.

In order to estimate the sensitivity of the predicted error to the input parameters, the numerical-data-driven model was evaluated at  $k = 2.0, 2.2$ , and  $2.4$ ,  $\bar{v} = 8.5, 9.5$ , and  $10.5$  m/s, and exponential autocorrelation decay bases of  $1.5, 1.6$ , and  $1.7$ . The resulting error estimates are shown alongside the observational data in Figure 6.



**Figure 6.** Model predictions for the Høvsøre site alongside observational data with one hour sampling. The closest fit of the model parameters with the characteristics of the observations is indicated with the thick black line. Observations are shown as blue markers. The legend labels state the shape factor, mean wind speed, and exponential decay base, respectively. The sampling frequency does not change the results significantly (not shown). The orange shading shows the mean  $\varepsilon \pm$  one standard deviation (see Section 3.3.3 for details).

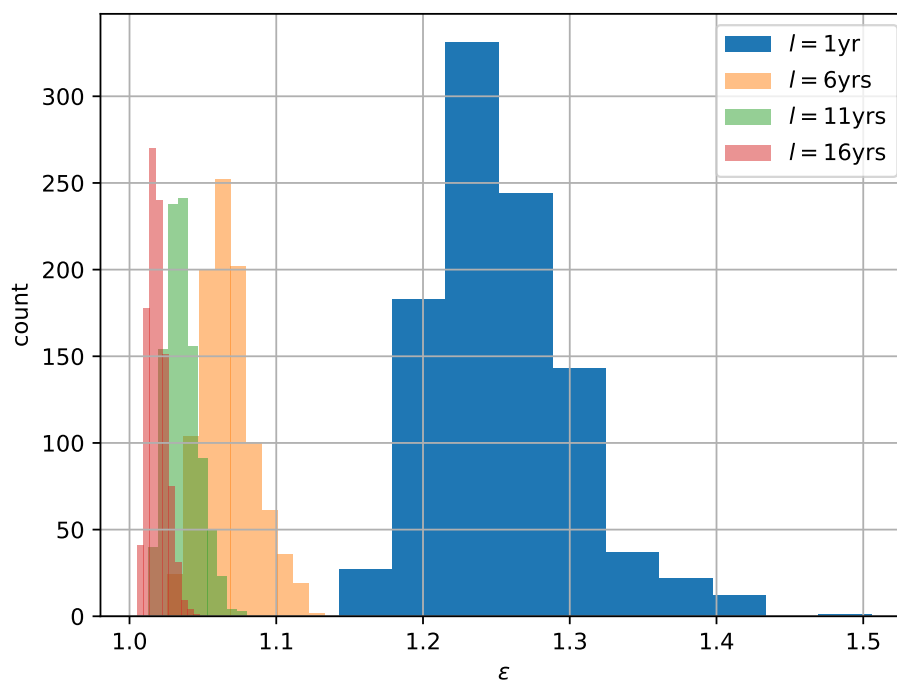
It can be seen that for a length of one year,  $\varepsilon$  was in the range of 1.12 to 1.35. The observations confirmed an  $\varepsilon$  of 1.33. The model error reduced faster than that estimated from the observational data. This accelerated error reduction was probably due to the lack of climatological effects within one year, which were not present in the artificial time series. In general, climatological variability in Northern Europe can lead to inter-decadal variations in wind energy of up to 30% and inter-annual variations in power; mean relative standard deviations of approximately 13% over 22 years have been reported [35]. After three years, the model dropped below 5% error, which was achieved by the observations after seven years. This result highlighted that the model provided a lower bound of the likely WPD error.

### 3.3.3. Variability of the WPD Error Using Synthetic Data

In the previous sections, one realization of a synthetic wind speed time series per set of wind speed characteristics was used to estimate the WPD error. However, the time series were generated using a stochastic process, and hence, these time series also inherited a stochastic behavior. The WPD error could thus exhibit such stochastic behavior, and here, we explore this by generating 1000 time series with a length of 20 years, which were computed with the Høvsøre site parameters (Weibull shape parameter of 2.19, mean wind speed of 9.47 m/s, and autocorrelation decay of 1.6).

The resulting histograms of  $\varepsilon$ , based on 20 year long time series, are shown in Figure 7. The standard deviations of  $\varepsilon$  for the 1, 6, 11, and 16 year long time series were 0.048, 0.018, 0.010, and 0.006, respectively. The mean of  $\varepsilon$  decreased from 1.25 for  $l = 1$  year to 1.07, 1.04, and 1.02, for  $l = 6, 11,$  and 16 years, respectively.

These results for the distribution of  $\varepsilon$  reiterated that for short time series, the errors were potentially large. For one year time series, the errors reached up to approximately 40%, and the uncertainty, here defined as the spread of the predictions made by a thousand time series evaluations, was equally large, ranging from 15 to 40%. The use of longer time series improved the maximum observed errors, and the uncertainty was reduced to values around 5%.



**Figure 7.** Histograms of  $\varepsilon$  for the synthetic time series generated using the Høvsøre parameters for four different lengths (different colors as indicated in the legend). The Python source codes for this analysis are provided in the Supplementary Materials to this article (also see Appendix B).

#### 4. Conclusions

The findings of this study clearly showed that a sufficient length of the wind speed time series used for useful wind resource estimation at a site was essential. Several years of wind speed time series data were required. Using four years of wind speed time series could be sufficient for some locations, e.g., the Høvsøre site presented here, but for other sites, the same length could still bear a 25% error in the final WPD estimate.

This study was limited by the fact that the analysis was only performed at three locations. Other locations with different wind climatologies and seasonal changes are likely to show slightly different length requirements. Furthermore, at the hub height, where modern large wind turbines operate, the winds are likely to show less variability. The study, however, highlighted important characteristics inherent to the estimation of WPD, which were expected to be transferable to other sites, and some of these limitations were rectified by the use of artificial time series data.

The estimation of the WPD at a site is vital to guide the exploration and development of wind energy. The examples presented in this work helped to quantify the uncertainties in WPD estimations, their strong dependence on the source data, and how they were pre-processed. The results herein are particularly relevant for the studies attempting to evaluate the impact of future climate scenarios on wind resources. If the evaluation periods in those climate impact studies are too short, the sampling is coarse, and potentially inappropriate averaging is used, any sensitivity to climate change may well be orders of magnitude smaller than the error incurred by the limitations mentioned above.

It is also instructive to see the impact of the exponential decay base on the WPD error. If there were no dependence on the exponential decay, then a one year time series of hourly data would yield the same error as a two year time series with 2 h sampling. However, due to the temporal evolution of the wind, in contrast to a purely stochastic process, the frequency of the sampling not only affected the amount of data over a given period, but also the representation of its fundamental characteristics.

The presented study started with an analysis of the impact of the cube of the wind speed on its discrete mean. It was shown how averaging added substantial underestimates (>10%) when the averaging periods were too long (>10 h). We also showed how this was in part due to the distribution of the wind speeds, which followed a Weibull distribution closely. This effect was non-linear and could, therefore, not easily be corrected. We also showed that the length of the time series was also crucial. Too short time series would introduce errors due to seasonal and inter-decadal oscillations.

This study presented of a numerical-data-driven model, which could be used to analyze different sites of interest and their sensitivity to sampling and time series length. This model was shown to provide good results with slight underestimates of the WPD. Despite this tendency to underestimate, the results were nevertheless beneficial, as they could guide us to find the minimum requirements for both sampling frequencies and time series length. In other words, the numerical-data-driven model could be used to optimize the time series length requirement, to avoid the error incurred by analyzing a too short time series being substantially larger than an application-dependent upper limit.

The stochastic analysis presented in Section 3.3.3, however, urges caution concerning the short time series and the numerical-data-driven model, as there, the distribution of  $\varepsilon$  could be quite broad. This problem will be investigated further in future work, alongside the evaluation of the sensitivity to the site parameters and possible ways of how to integrate more site characteristics into the model and construct a system that can reliably predict time series requirements for a large variety of sites. This study demonstrated the need, motivation, and first step towards this longer term aim.

**Supplementary Materials:** The following are available at <http://www.mdpi.com/2071-1050/12/8/3431/s1>.

**Author Contributions:** Conceptualization, M.G.; methodology, M.G.; software, M.G. and A.P.; validation, M.G. and A.P.; formal analysis, M.G. and A.P.; investigation, M.G., A.P., and V.M.; resources, A.P.; data curation, M.G.; writing, original draft preparation, M.G.; writing, review and editing, M.G., A.P., and V.M.; visualization, M.G.; funding acquisition, A.P. All authors read and agreed to the published version of the manuscript.

**Funding:** The presented work was partly funded by the Ministry of Foreign Affairs of Denmark and administered by the Danida Fellowship Centre through the “Multi-scale and Model-Chain Evaluation of Wind Atlases” (MEWA) project.

**Acknowledgments:** The authors would like to thank the four anonymous referees for their constructive comments.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A. Usage of the Numerical-Data-Driven Model

The numerical-data-driven model is provided as Supplemental Data in the form of a netcdf file. This file can be used to evaluate the error in WPD for desired parameter spaces as illustrated below using Python (the source code is also provided in the Supplemental Data File as datamodel.py).

```
# define model input parameters, here, as example the parameters of the
# strong black line in Figure 6
shapes = [2.2] # Weibull shape parameter
meanWinds = [9.5] # mean wind speed (at hub height) in meter per second
expDecayBases = [1.6] # exponential autocorrelation decay base
sampling = [1] # sampling period in hours
years = range(1,14) # time series length

import numpy as np
# convert the one dimensional arrays into 5 dimensional arrays for interpolation
grid_f, grid_y, grid_s, grid_m, grid_e = np.meshgrid(sampling, years, shapes, meanWinds, \
expDecayBases)

from netCDF4 import Dataset as Ds
# load the data
ncf = Ds("datamodel.nc")
from scipy.interpolate import griddata as gd
# evaluate using linear interpolation
epsilon = gd(ncf.variables["vars"][:,:],
ncf.variables["values"][:,],
(grid_f, grid_y, grid_s, grid_m, grid_e),
method='linear')
# print the resulting vector of epsilon (as plotted in Figure 6)
print (epsilon)
```

## Appendix B. Program Code for the Stochastic Analysis of Section 3.3.3

The results presented in Section 3.3.3 can be reproduced by executing the three Python scripts provided in the Supplemental Data File in the sequence:

- generateTs.py, to generate the synthetic time series,
- evaluateErr.py, to perform the sliding window error analysis, and
- analyzeErr.py, to plot the histograms and compute standard deviations and means.

## References

1. Zheng, C.W.; Pan, J. Assessment of the global ocean wind energy resource. *Renew. Sustain. Energy Rev.* **2014**, *33*, 382–391. [[CrossRef](#)]
2. Gadad, S.; Deka, P.C. Offshore wind power resource assessment using Oceansat-2 scatterometer data at a regional scale. *Appl. Energy* **2016**, *176*, 157–170. [[CrossRef](#)]
3. Karagali, I.; Peña, A.; Badger, M.; Hasager, C.B. Wind characteristics in the North and Baltic Seas from the QuikSCAT satellite. *Wind Energy* **2014**, *17*, 123–140. [[CrossRef](#)]

4. Kim, J.Y.; Kim, H.G.; Kang, Y.H. Offshore Wind Speed Forecasting: The Correlation between Satellite-Observed Monthly Sea Surface Temperature and Wind Speed over the Seas around the Korean Peninsula. *Energies* **2017**, *10*, 994. [[CrossRef](#)]
5. Pimenta, F.; Kempton, W.; Garvine, R. Combining meteorological stations and satellite data to evaluate the offshore wind power resource of southeastern Brazil. *Renew. Energy* **2008**, *33*, 2375–2387. [[CrossRef](#)]
6. Doubrawa, P.; Barthelmie, R.J.; Pryor, S.C.; Hasager, C.B.; Badger, M.; Karagali, I. Satellite winds as a tool for offshore wind resource assessment: The great lakes wind atlas. *Remote Sens. Environ.* **2015**, *168*, 349–359. [[CrossRef](#)]
7. Hasager, C.B.; Barthelmie, R.J.; Christiansen, M.B.; Nielsen, M.; Pryor, S.C. Quantifying offshore wind resources from satellite wind maps: Study area the north sea. *Wind Energy* **2006**, *9*, 63–74. [[CrossRef](#)]
8. Hasager, C.B.; Badger, M.; Peña, A.; Larsen, X.G.; Bingol, F. SAR-Based Wind Resource Statistics in the Baltic Sea. *Remote Sens.* **2011**, *3*, 117–144. [[CrossRef](#)]
9. Wang, Y.H.; Walter, R.K.; White, C.; Farr, H.; Ruttenberg, B.I. Assessment of surface wind datasets for estimating offshore wind energy along the central California coast. *Renew. Energy* **2019**, *133*, 343–353. [[CrossRef](#)]
10. Carvalho, D.; Rocha, A.; Gómez-Gesteira, M.; Santos, C.S. WRF wind simulation and wind energy production estimates forced by different reanalyses: Comparison with observed data for portugal. *Appl. Energy* **2014**, *117*, 116–126. [[CrossRef](#)]
11. Fueyo, N.; Sanz, Y.; Rodrigues, M.; Montañés, C.; Dopazo, C. High resolution modelling of the on-shore technical wind energy potential in Spain. *Wind Energy* **2010**, *13*, 717–726. [[CrossRef](#)]
12. Dvorak, M.J.; Archer, C.L.; Jacobson, M.Z. California offshore wind energy potential. *Renew. Energy* **2010**, *35*, 1244–1254. [[CrossRef](#)]
13. He, Y.; McFarlane, N.A.; Monahan, A.H. The Influence of Boundary Layer Processes on the Diurnal Variation of the Climatological Near-Surface Wind Speed Probability Distribution over Land. *J. Clim.* **2012**, *25*, 6441–6458. [[CrossRef](#)]
14. Barcons, J.; Avila, M.; Folch, A. Diurnal cycle RANS simulations applied to wind resource assessment. *Wind Energy* **2019**, *22*, 269–282. [[CrossRef](#)]
15. Li, D.; Geyer, B.; Bisling, P. A model-based climatology analysis of wind power resources at 100-m height over the Bohai Sea and the Yellow Sea. *Appl. Energy* **2016**, *179*, 575–589. [[CrossRef](#)]
16. Matthew, O.J.; Ohunakin, O.S. Simulating the effects of climate change and afforestation on wind power potential in Nigeria. *Sustain. Energy Technol. Assess.* **2017**, *22*, 41–54. [[CrossRef](#)]
17. Wang, M.; Ullrich, P.; Millstein, D. Datasets on hub-height wind speed comparisons for wind farms in California. *Data Brief* **2018**, *19*, 214–221. [[CrossRef](#)]
18. Wang, M.; Ullrich, P.; Millstein, D. The future of wind energy in California: Future projections with the variable-resolution cesm. *Renew. Energy* **2018**, *127*, 242–257. [[CrossRef](#)]
19. Reboita, M.S.; Amaro, T.R.; de Souza, M.R. Winds: Intensity and power density simulated by RegCM4 over South America in present and future climate. *Clim. Dyn.* **2018**, *51*, 187–205. [[CrossRef](#)]
20. Fant, C.; Gunturu, B.; Schlosser, A. Characterizing wind power resource reliability in southern Africa. *Appl. Energy* **2016**, *161*, 565–573. [[CrossRef](#)]
21. Peña, A. Østerild: A natural laboratory for atmospheric turbulence. *J. Renew. Sustain. Energy* **2019**, *11*, 063302. [[CrossRef](#)]
22. Peña, A.; Floors, R.; Sathe, A.; Gryning, S.E.; Wagner, R.; Courtney, M.S.; Larsén, X.G.; Hahmann, A.N.; Hasager, C.B. Ten years of boundary-layer and wind-power meteorology at Høvsøre, Denmark. *Bound.-Layer Meteorol.* **2016**, *158*, 1–26. [[CrossRef](#)]
23. Hahmann, A.N.; Peña, A.; Hansen, J.C. *WRF Mesoscale Pre-Run for the Wind Atlas of Mexico*; Technical Report E-0126; DTU Wind Energy: Kgs. Lyngby, Denmark, 2016.
24. Veers, P.S.; McNerney, G.M. *Markov Method for Simulating Non-Gaussian Wind Speed Time Series*; Technical Report SAND84; Sandia National Labs.: Albuquerque, NM, USA, 1985.
25. Shoaib, M.; Siddiqui, I.; Rehman, S.; Rehman, S.U.; Khan, S.; Lashin, A. Comparison of Wind Energy Generation Using the Maximum Entropy Principle and the Weibull Distribution Function. *Energies* **2016**, *9*, 842. [[CrossRef](#)]

26. 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*, 2419–2438. [[CrossRef](#)]
27. Ahmed, A.S. Wind resource assessment and economics of electric generation at four locations in Sinai Peninsula, Egypt. *J. Clean. Prod.* **2018**, *183*, 1170–1183. [[CrossRef](#)]
28. Shu, Z.R.; Li, Q.S.; Chan, P.W. Investigation of offshore wind energy potential in Hong Kong based on Weibull distribution function. *Appl. Energy* **2015**, *156*, 362–373. [[CrossRef](#)]
29. Mohammadi, K.; Alavi, O.; Mostafaeipour, A.; Goudarzi, N.; Jalilvand, M. Assessing different parameters estimation methods of Weibull distribution to compute wind power density. *Energy Convers. Manag.* **2016**, *108*, 322–335. [[CrossRef](#)]
30. Shu, Z.R.; Li, Q.S.; Chan, P.W. Statistical analysis of wind characteristics and wind energy potential in Hong Kong. *Energy Convers. Manag.* **2015**, *101*, 644–657. [[CrossRef](#)]
31. Shoaib, M.; Siddiqui, I.; Rehman, S.; Rehman, S.U.; Khan, S. Speed Distribution Analysis Based on Maximum Entropy Principle and Weibull Distribution Function. *Environ. Prog. Sustain. Energy* **2017**, *36*, 1480–1489. [[CrossRef](#)]
32. Chaurasiya, P.K.; Ahmed, S.; Warudkar, V. Study of different parameters estimation methods of Weibull distribution to determine wind power density using ground based Doppler SODAR instrument. *Alex. Eng. J.* **2018**, *57*, 2299–2311. [[CrossRef](#)]
33. Rehman, S.; Alam, A.M.M.; Meyer, J.P.; Al-Hadhrani, L.M. Wind speed characteristics and resource assessment using weibull parameters. *Int. J. Green Energy* **2012**, *9*, 800–814. [[CrossRef](#)]
34. IEC. *Power Performance Measurements of Electricity Producing Wind Turbines*; Standard IEC 61400-12-1; International Electrotechnical Commission: Geneva, Switzerland, 2005.
35. Petersen, E.L.; Mortensen, N.G.; Landberg, L.; Højstrup, J.; Frank, H.P. Wind power meteorology. Part I: Climate and turbulence. *Wind Energy* **1998**, *1*, 2–22. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).