



Article

Enhancing Long-Term Wind Power Forecasting by Using an Intelligent Statistical Treatment for Wind Resource Data

Monica Borunda ^{1,2,3,*}, Adrián Ramírez ³, Raul Garduno ⁴, Carlos García-Beltrán ¹  and Rito Mijarez ⁴ 

¹ Centro Nacional de Investigación y Desarrollo Tecnológico, Tecnológico Nacional de México, Cuernavaca 62490, Morelos, Mexico; carlos.gb@cenidet.tecnm.mx

² Consejo Nacional de Humanidades, Ciencias y Tecnologías, Mexico City 03940, Mexico

³ Faculty of Science, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico; felos@ciencias.unam.mx

⁴ Instituto Nacional de Electricidad y Energías Limpias, Cuernavaca 62490, Morelos, Mexico; rgarduno@ineel.mx (R.G.); rmijarez@ineel.mx (R.M.)

* Correspondence: monica.bp@cenidet.tecnm.mx; Tel.: +52-55-4939-1608

Abstract: Wind power is an important energy source that can be used to supply clean energy and meet current energy needs. Despite its advantages in terms of zero emissions, its main drawback is its intermittency. Deterministic approaches to forecast wind power generation based on the annual average wind speed are usually used; however, statistical treatments are more appropriate. In this paper, an intelligent statistical methodology to forecast annual wind power is proposed. The seasonality of wind is determined via a clustering analysis of monthly wind speed probabilistic distribution functions (PDFs) throughout n years. Subsequently, a methodology to build the wind resource typical year (WRTY) for the $n + 1$ year is introduced to characterize the resource into the so-called statistical seasons (SSs). Then, the wind energy produced at each SS is calculated using its PDFs. Finally, the forecasted annual energy for the $n + 1$ year is given as the sum of the produced energies in the SSs. A wind farm in Mexico is chosen as a case study. The SSs, WRTY, and seasonal and annual generated energies are estimated and validated. Additionally, the forecasted annual wind energy for the $n + 1$ year is calculated deterministically from the n year. The results are compared with the measured data, and the former are more accurate.

Keywords: forecasting; wind power generation; machine learning; clustering; Weibull PDFs; statistical seasonality; wind resource typical year; energy yield



Citation: Borunda, M.; Ramírez, A.; Garduno, R.; García-Beltrán, C.; Mijarez, R. Enhancing Long-Term Wind Power Forecasting by Using an Intelligent Statistical Treatment for Wind Resource Data. *Energies* **2023**, *16*, 7915. <https://doi.org/10.3390/en16237915>

Academic Editor: Davide Astolfi

Received: 31 October 2023

Revised: 28 November 2023

Accepted: 30 November 2023

Published: 4 December 2023



Copyright: © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Renewable energy integration is among the key actions proposed for climate change mitigation [1]. Hydraulic power, wind power, and solar power are the major renewable energy alternatives to supply clean energy. Regarding wind power, it accounted for 24% of the electricity generated with renewables in 2018 [2], and it accounted for 93.6 GW of new wind power installed capacity, bringing the total to 837 GW in 2021, which meant that it remained the second-largest alternative source of electric energy [3]. Despite these impressive numbers, the annual wind capacity still needs to increase from about 75 GW in 2022 to 350 GW in 2030 to achieve the 1.5 °C and Net Zero climate goals for 2030 [3]. According to [4], major efforts should concentrate on facilitating, permitting, and gaining public support, identifying suitable sites, decreasing costs, and reducing project development. Recommended courses of action include developing advanced solutions for wind power grid integration, as well as improved methodologies for resource assessment and system expansion planning. This paper deals with the development of novel methodologies for assessing wind speed and power resources.

The power available in a horizontal flow of wind to be transformed with a wind turbine generator (WTG) is proportional to the area swept by the rotor, the air density at the

hub level, and the cube of the wind speed, which means that wind power is greatly affected by wind speed [5]. For any working WTG at a given location, the area swept by the rotor is invariable, and the air density is nearly constant, but the wind speed can vary enormously; consequently, the power output and the energy generated over time by any WTG change accordingly. The wind speed is very sensitive to the atmospheric temperature and pressure, wind gusts and turbulence, time of the year, and land rugosity, among other factors, which can fluctuate widely from place to place for very short periods (seconds, minutes, and tens of minutes), as well as for longer times (hours, days, and months) [6]. Short-term wind speed variations present random behavior called intermittency, meanwhile, long-term wind speed variations exhibit cyclical patterns, defined as variability.

Wind speed variations cause uncontrolled deficiencies or excesses of wind power and produce stress on power generation systems [7]. In general, the intermittency and variability of wind speed and power are the major challenges in terms of integrating wind power into electric power systems [8]. The main consequences are the following:

- a. Maintaining electricity supply and system stability, since the integration of wind generators with uncontrollable and variable output decreases the system inertia required to back up immediately the loss of power of failed generators, thus affecting the reliability and flexibility of the power system [9].
- b. Maintaining the cost-effective operation of the power system, since the compensation of wind power variations may require burning fuel oil or gas at conventional plants, importing electricity from other power systems at higher costs, integrating expensive energy storage means, etc. [10].
- c. Maintaining low CO₂ emissions, since a lack or surplus of generation from wind facilities may require the flexible operation of fossil-fueled power generators emitting more greenhouse gases than necessary if the operation is not optimal [11].

In this regard, the prediction of wind resources and the estimation of wind power can be used to schedule and maximize the contribution of wind power in power system operation, mitigating the undesirable effects of wind speed variation. Thus, accurate forecasts can deliver remarkable economic, technical, and environmental impacts on the operation of power systems.

The usefulness of wind forecasting in the operation of power systems depends on the time scale in the future for which forecasts are generated, named the forecast horizon. Forecast horizons range from a few seconds to several years. On the other hand, the forecast period refers to the frequency of the input data used for the prediction. Table 1 shows the forecast horizons, the length of time for which predictions are generated, the frequency of the data used for the prediction, and their applications. It is worth noting that different authors may slightly differ in the forecast horizon used, but the time scales remain the same in all classifications [12,13]. It is important to remark that the larger the forecast horizon, the less accurate the prediction. Thus, long-term forecasting is a more difficult and less studied topic than medium- or short-term forecasting. As noted in the following, there are not many that have been developed concerning long-term wind power forecasting. Thus, a wide window of opportunities has been opened for new methodologies that provide more accurate results.

In this work, we are keenly interested in long-term wind power forecasting. In particular, the main contribution is a novel hybrid methodology that uses concepts from the fields of statistics and artificial intelligence for one-year-ahead wind power forecasting. The obtained results are presented with their corresponding probability that they will exceed predefined energy thresholds useful for risk analysis. The results obtained from this approach are compared and have been found to be superior to the ones obtained when using the traditional deterministic method, which consists of using the annual average wind speed. The former ones are more accurate. Additionally, it is possible to forecast energy generation in the statistical seasons of a year. Thus, the proposed simple yet powerful and precise method provides an improvement in long-term wind power forecasting, which leads to better decision making for operation, management, dispatch

planning, operation optimization, resource assessment, site selection, cost estimation, feasibility analysis, system expansion planning, bankable documentation, and financial investments. Furthermore, as a plus, a methodology is proposed to construct a wind resource in a typical year to characterize the wind resource at any given site (there are some restrictions on the dimension of the database, which are mentioned in Section 2.3). This is also very important since there are currently no typical meteorological year databases for all locations on the globe.

Table 1. Forecasts at different forecast horizons and periods and their corresponding applications.

Forecast	Forecast Horizon	Forecast Period	Application
Very short-term	1 min to 1 h ahead	30 s, 1 min, or 10 min	Wind turbine control, real-time grid operation, and electricity market compensation.
Short-term	1 h to 1 day ahead	10 min or 1 h	Load dispatch, load following, feedback voltage and power control, and protection to preserve physical integrity and operational security in the electricity market.
Medium-term	1 day to 1 week ahead	10 min or 1 h	Day-ahead electricity market, economic dispatch, unit commitment and maintenance.
Long-term	1 week to years ahead	1 h, 1 day, 1 month, or 1 year	Operation management, dispatch planning, optimal operation, resource assessment, site selection, cost and feasibility analysis, system expansion planning, bankable documentation, and financial investments [14,15].

State of the Art

Wind power forecasting methods are mainly grouped into four categories [16]:

- Physical prediction methods convert meteorological variables (temperature, pressure, humidity, etc.) and geomorphic conditions (land roughness, topography, obstacles, etc.) at the sites of interest into wind speed forecasts through the development of multivariate models based on mathematical equations of the physical processes [17]. Once wind speed is predicted, wind power is estimated using the speed–power curve of the wind turbines, either using the curves provided by the manufacturer for new projects or those derived from measurements, if already available. Finally, transfer functions from the physical variables to wind power are determined to use predicted meteorological conditions to forecast wind power.
- Statistical prediction methods are based on historical data models that relate wind speed or power to the values of meteorological variables [18]. Statistical prediction models follow two steps: (1) a wind speed prediction model is designed using curve fitting; (2) the model parameters are refined using the current predicted data and early previous data values [13]. Statistical models can be (a) time series (linear, non-linear), (b) structural (Kalman filter), or (c) black box models [19,20]. Time series models based on mathematical equations, linear or non-linear, are the most used. Time series models are composed of two parts: an autoregressive and a moving average part to consider the persistent behavior of the wind, and a transformation to include the bias effect caused by other meteorological variables forecasts. Popular statistical models include the Auto-Regressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA), and Seasonal Auto-Regressive Integrated Moving Average (SARIMA). Other statistical methods use the Kalman filter to predict the wind speed or power. These methods modify the weights of recursive equations during the process to achieve high-precision predictions that overcome the poor forecasting precision of low-order time series models. However, difficulties arise in establishing the state and measurement equations of the Kalman filter. At present, statistical prediction methods consider the use of meteorological forecasts from different meteorological offices as input, as well as the optimal use of spatially distributed measurement data either for prediction error correction or for issuing warnings of a potentially large uncertainty.

A comparative study between a physical method, using a downscaling approach of Numerical Weather Prediction (NWP) models, and a statistical method, using time series-based models, for wind speed and power short-term forecasting can be found in [21].

- Principally, Artificial Intelligence (AI) prediction methods are based on Machine Learning (ML), which is focused on solving practical problems, shifting from the symbolic approaches of traditional AI towards using approaches borrowed from statistics, fuzzy logic, and probability theory [21]. Currently, the major objectives of ML are to classify data using non-linear models, not written through the use of a simple mathematical relationship, and to make predictions using those models. ML includes techniques such as Artificial Neural Networks (ANNs), Neuro-Fuzzy Systems (NFSs), Support Vector Machines (SVMs), decision trees, Bayesian networks, belief functions, regression analysis, Wavelet Analysis, Gaussian processes, Genetic Algorithms (GAs), evolutionary optimization (EO) and Singular Spectrum Analysis (SSA) [22]. It is worth highlighting that these models usually need big data sets to be trained. It is common to find missing data in data sets; however, machine learning tools can be used to complete them [23].
- Hybrid methods use at least two different methods to improve the forecasting performance and reduce the prediction error. For instance, promising results for very short and short-term wind speed forecasting have been achieved in [18,24–28]. An interesting methodology based on probabilistic forecasting and machine learning techniques has been constructed for very short-term wind power forecasting with a high level of reliability and accuracy [29].

Table 2 shows some of the most recent surveys reported in the literature for wind power forecasting using these methods.

Even though several works have been undertaken on this subject, it is a complex topic with many issues to consider. In particular, wind resources are characterized by their intermittent nature. The wind speed fluctuates continuously, and as a result, the power from a wind turbine or a wind farm varies. Specifically, wind speed may vary in the short or long term. Very short-term wind speed variations last from several seconds to several minutes, such as turbulence and gusts. We refer to short-term wind power variations to wind power fluctuations ranging from minutes to a few hours, which is reflected in the variation in the ramping of the output power of a wind turbine or wind farm. Long-term variations cover daily, seasonal, or yearly wind speed changes that affect wind power production, resulting in inter-annual, seasonal, monthly, and diurnal variations in output power levels.

Table 2. Recent surveys for wind power forecasting using different prediction methods, physical, statistical, artificial intelligence-based and hybrid, reported in the literature. Years of publication and references are included.

Reference	Year	Physical	Statistical	AI	Hybrid	Title
[30]	2010	X				Review of the use of numerical weather prediction (NWP) models for wind energy assessment
[13]	2010	X	X			Forecasting methods with different time horizons
[31]	2010	X	X			Wind power forecasting & prediction methods
[32]	2011	X	X	X	X	Current methods and advances in forecasting of wind power generation
[33]	2011	X	X			A review of wind power forecasting models.
[16]	2011	X	X	X	X	Review of evaluation criteria and main methods of wind power forecasting
[34]	2013	X	X	X		A detailed literature review on wind forecasting

Table 2. Cont.

Reference	Year	Physical	Statistical	AI	Hybrid	Title
[35]	2013		X			Wind power forecasting: A review of statistical models
[36]	2014	X	X	X	X	A literature review of wind forecasting methods
[37]	2015	X	X	X		A review and evaluation of current wind power prediction technologies
[38]	2015	X	X	X	X	Ensemble methods for wind and solar power forecasting—A state of the art review
[39]	2016				X	A comparative study of wind power forecasting techniques—A review
[40]	2017	X	X	X	X	Wind power forecasting—a review of the state of the art.
[41]	2019			X	X	Deterministic wind energy forecasting: a review of intelligent predictors and auxiliary methods
[42]	2020	X	X	X	X	A critical review of wind power forecasting methods-past, present and future
[43]	2020		X	X	X	A review of wind speed and wind power forecasting techniques
[44]	2020			X		Wind power forecasting using machine learning: state of the art, trends and challenges
[45]	2021			X		A review of wind speed and wind power forecasting with deep neural networks
[46]	2021	X	X			Wind power forecasting
[47]	2022			X		New developments in wind energy forecasting with artificial intelligence and big data: a scientometric insight
[48]	2023			X	X	Artificial intelligence in wind speed forecasting: a review
[49]	2023			X		Recent advances in data-driven prediction for wind power
[50]	2023	X	X	X	X	A review of modern wind power generation forecasting technologies

As mentioned, long-term wind availability is largely influenced by local weather conditions, seasonal variations, and time span variations and, thus, is characterized by strong fluctuations, uncertainty, and intermittency. Hence, it is very important, but also very complex, to count on good long-term wind power forecasting to match the electricity demand, which also depends heavily on the same factors. Thus, long-term wind power forecasting may be addressed by studying wind speed seasonality. Some studies addressing this approach for wind power forecasting have been reported in the literature [51]. These studies use either physical, statistical, A.I., or hybrid methodologies [18,52]. Some of them are described in the following.

In 2009, Fan et al. developed a two-stage hybrid network with Bayesian clustering using dynamics and support vector regression for the generation forecasting of a wind farm [53]. They considered the wind speed time series non-stationarity due to the multiple seasonality and found a correlation between the wind speed and each season of the year at the site.

In 2011, the stochastic and seasonality pattern of wind power was tackled by designing a combined Autoregressive Fractionally Integrated Moving Average (ARFIMA) and GARCH model [54]. Additionally, a novel hybrid wind speed forecasting method based on a back propagation neural network was reported [55]. They eliminated seasonal effects using a seasonal exponential adjustment and characterized four seasonal wind data. They forecasted the daily average wind speed one year ahead in a location in China with acceptable mean absolute errors.

In 2012, Troccoli et al. analyzed long-term trends in wind observations [56]. They found that a qualitative link could be established between features in the linear trends and some atmospheric indicators. They also found that the magnitude of the trend is sensitive to the period selected.

In 2013, Doblas-Reyes et al. claimed that seasonal climate forecasts occupy an intermediate zone between weather forecasting and climate projections [57]. They presented an overview of the state of the art in global seasonal predictability and forecasting and concluded with a list of challenges for researchers interested in seasonal forecasting to focus on in the future.

In 2015, Saroha et al. remarked that wind power generation is highly associated with nature and multiple seasonality aspects [37]. They concluded that it is not an easy task to design a perfect prediction model.

In 2016, Azimi et al. developed a hybrid model for wind energy by employing the K-means clustering algorithm to group wind data [58]. They proposed a methodology for selecting the optimal clusters, the data of which will be fed into a perceptron-type neural network to enable the wind power forecasting process for future time intervals of 1, 24, and 48 h ahead. Additionally, Grigonyte et al. presented a short-term wind speed forecasting Autoregressive Integrated Moving Average (ARIMA) model and found that the accuracy increases when finding wind weather seasons [59]. They found results with good accuracy in 6–8 h ahead wind speed forecasting.

In 2017, Yatiyana et al. introduced an Autoregressive Integrated Moving Average (ARIMA) method to consider the seasonal and trending changes in their wind power generation forecasting hybrid model [60]. They forecasted the wind speed and direction in Western Australia.

In 2018, Lledó et al. studied the impact of seasonality and other factors on wind speeds [61]. They showed that the interannual variability of wind speed in the USA is dominated by El Niño/Southern Oscillation and by sea surface temperature variation in the Pacific. Then, in 2019, they found that renewable generation from hydro, solar, and wind power installations is sensitive to seasonal or multiannual climate oscillations and long-term trends [62]. They found a methodology to produce seasonal predictions of a wind turbine capacity factor to compute wind power generation. They validated their methodology with data from the new European Centre for Medium-Range Weather Forecasts (ECMWF) seasonal forecast system (System 4) [63] in Europe in winter.

Additionally, in 2018, Mariotti et al. focused on the need for a good weather–climate forecasting methodology in the subseasonal-to-seasonal prediction gap, which ranges from 2 weeks to a season [64]. They also presented the progress in this direction achieved by the Subseasonal-to-Seasonal (S2S) Prediction Project of the World Climate Research Programme (WCRP) and the World Weather Research Programme (WWRP).

In 2022, Tyass et al. presented a Seasonal Autoregressive Integrated Moving Average (SARIMA) model for short-term wind speed forecasting [65]. They validated their model with data from a location in Morocco by computing RMSE and MAPE errors. Their results achieved excellent forecasting accuracy. Additionally, Tawn et al. proposed a methodology that utilizes the ECMWF S2S (Season-to-Season) climatological model dataset and historical data from the ECMWF's ERA5 model for wind speed estimation [66]. Forecasts are shown to improve climatology at all sites. The results were very accurate for 1 week ahead and fairly accurate up to 6 weeks ahead.

In 2023, Sulagna et al. presented a statistical analysis model of wind power generation forecasting for Western India [67]. They looked for seasonality and trends in some weather factors. They found annual seasonality in wind power generation useful to one-day-ahead wind power prediction. Additionally, in this year, Mesa-Jimenez et al. conducted a study where the annual behavior of wind is modeled by employing probabilistic Bayesian inference models and Markov Chain Monte Carlo models [68]. They identified seasonal patterns if the wind speed follows a probability distribution composed of beta distributions within each season. Subsequently, these distributions were utilized to estimate the energy

produced in each season. In the same year, Magaña-Gonzalez et al. performed an analysis of the seasonal variability of the wind and solar resources in Mexico [69]. They explored wind and solar resources using experimental and ERA5 data. They identified a bias effect on the power estimations. They also found that the capacity factor of wind turbines is affected by bias-correction methods compared to that of photovoltaics.

As mentioned above and highlighted from the previous works, the wind speed seasonality is an important aspect to take into consideration for wind resource analysis that is not often considered. Seasonality is demonstrated by the change in environmental and weather conditions, such as light, humidity, temperature, rainfall, wind, and so on throughout the year [70]. Seasonality is defined as “a time-based cycle of systematic, regular fluctuation within a fixed pattern that could be described by peak timing, amplitude, and interval”. Thus, seasonality may be the key ingredient to enhance weather prediction, particularly, wind forecasting. Moreover, weather and wind seasonality depend on the location of the site of interest and local meteorological and topological conditions.

The objective of this work is to perform long-term wind power forecasting. For this purpose, an intelligent statistical method is presented. The one-year-ahead forecast of the produced annual wind energy requires two main components: (a) finding the statistical seasonality of the site of interest using an n -year wind speed database, which is performed using statistics and Artificial Intelligence; (b) constructing the Wind Resource Typical Year for the $n + 1$ year using the n -year database, which is performed using statistics. Then, the annual and seasonal forecasted wind energies can be estimated.

There may be many ways to find the weather and the wind resource seasonality. The obvious one corresponds to the seasons resulting from the axial parallelism of Earth's tilted orbit around the Sun [71]. However, depending on the distance of the site to the equator, these stations are noticeable to a greater or lesser extent, but there may be different and more precise ways to find a pattern in the wind speed at a given site. In this work, we propose the use of Machine Learning to find patterns in the wind speed data at a given site. Using clustering analysis [72], it is possible to group a large amount of wind speed data into clusters with similar characteristics. Thus, we introduce the definition of the Statistical Seasonality (SS) of the wind resource at the site, with each cluster corresponding to a statistical season.

Based on previous wind data processing, the statistical seasons for wind are determined using AI tools. The point is to construct groups of natural months that behave similarly according to their statistical wind activity, that is, with similar wind speed PDFs. With this aim, data clustering analysis using the K-means method is carried out in three steps over the defining parameters (shape and scale) of the 12 monthly PDFs of the n years and Silhouette performance is evaluated to determine the optimal number of wind statistical seasons. Note that wind statistical seasons do not necessarily correspond to natural seasons, even in the number of months they contain; furthermore, a wind statistical season may not be composed of subsequent natural months, that is, may include disjointed natural months.

Once the statistical seasons of wind are found, their corresponding PDFs are calculated for the $n + 1$ year through the construction of the Wind Resource Typical Year. The annual production of wind power for the $n + 1$ year can be calculated by adding up the produced energies at every season of the $n + 1$ year. This yield is compared against that calculated with the single annual PDF. Improved performance is expected using the PDFs of statistical wind seasons. To validate this approach, the wind data of n years are used to obtain the required PDFs and a comparison is made against the power obtained with the data of the year $n + 1$. Furthermore, such PDFs are advantageously used to directly predict the Annual Energy Production (AEP) at different probabilities of exceedance, which are useful when analyzing the annual cost-to-benefit ratio of the wind power station.

In Section 2, the proposed methodology is described. Section 2.1 describes the statistical nature and treatment of the wind resource. Section 2.2 introduces the procedure for determining the seasonality of wind speed using a clustering analysis. Section 2.3

shows the basics of the clustering analysis and the Silhouette method to obtain the optimal number of clusters. Section 2.4 outlines the construction of the wind resource typical year. Section 2.5 presents the estimation of wind power that can be produced, and Section 2.6 shows the Mean Absolute Percentage Error (MAPE) criteria to quantify the forecasting errors. Section 3 provides a case study to show the application and validate the proposed methodology with data from a site in Mexico. Then, a pertinent discussion of the results is provided in Section 4. Finally, the main conclusions of the work are presented in Section 5 showing the core contributions and drawbacks of this work as well as future work.

2. Materials and Methods

Meteorological conditions are the main factor influencing the intermittency of renewable energy sources, such as solar and wind energy, at a given site. Thus, meteorological data are fundamental when assessing the availability of the resource. These data are accessible through weather stations, when available, or by global and regional weather models [73–75]; they usually consist of the hourly values of solar radiation, wind speed and angle, ambient temperature, humidity, and atmospheric pressure. Most sites possess big datasets of this information covering many years. Hence, it is necessary to deal with this huge amount of data with statistics. Similarly, a Typical Meteorological Year (TMY) represents the most frequent weather conditions of a site.

This work orients its efforts toward boosting the potential of wind energy by enhancing long-term wind power forecasting. For this purpose, information about wind availability is required. The wind power generated by wind turbines depends on the wind velocity, which consists of the wind speed and its direction. Most wind turbines contain the Yaw control mechanism to move the nacelle and blades according to the wind direction to capture the maximum available wind. Hence, this work focuses on wind speed for our study on long-term wind power forecasting.

Moreover, wind resources present with seasonality. This seasonality corresponds to cyclical changes in the meteorological variables, solar radiation, wind, ambient temperature, and rainfall. Hourly data can exhibit three types of seasonality: daily, weekly, and annual patterns. Therefore, wind resource statistical seasonality enhances wind power forecasting via statistical tools and Artificial Intelligence.

Considering n years of available wind speed data, the objective of this paper is to forecast the wind power generated in the $n + 1$ year by (a) finding the statistical seasonality of the resource at the site and (b) constructing the Wind Resource Typical year for the $n + 1$ year using the wind speed data of the previous n years. We propose a method for the estimation of the annual wind energy for the year $n + 1$, based on wind seasonality, which may better characterize the annual variability of the wind speed for the year $n + 1$ at any specific site of interest. It is believed that proceeding to estimate wind power in terms of the so-called wind statistical seasons, calculated from n years of wind speed data, provides an enhanced estimation of the produced energy in the year $n + 1$ against that obtained using the annual mean value of the wind speed, calculated with the data of the year n , as is currently performed. Hence, increasing the precision of forecasts will undoubtedly improve the exploitation of wind energy to provide the bulk quantities of electric energy required worldwide in the coming years.

To forecast the annual energy at year $n + 1$, and as a preliminary result, the construction of the Wind Resource Typical Year (WRTY) for the $n + 1$ year is proposed. Statistics are used to process the large amounts of wind speed data that should span n years of wind activity at the site of interest. Such data can be currently obtained for free from well-known meteorological data repositories. Consider the following notation:

$PDF_{y,m}$	PDF for the month m of the year y ,
$PDF_{cum,m}$	PDF for the month m for the cumulative years,
$PDF_{WRTY,m}$	PDF for the month m of the WRTY

where *PDF* stands for Probabilistic Distribution Function, the index y refers to the number of the year $y = 1, \dots, n$ and m refer to the month of the year $m = 1, \dots, 12$. A construction of the Wind Resource Typical Year (WRTY) for the $n + 1$ year is performed by using the different *PDFs*. First, a monthly characteristic behavior is found by gathering data from all n years, month by month. Then, 12 monthly characteristic *PDFs* are constructed by fitting *PDFs* to the gathered monthly data $PDF_{cum,m}$. On the other hand, *PDFs* are fitted to the monthly data for each year $PDF_{y,m}$ obtaining n *PDFs* for each month. Then, the month of January of the WRTY for the $n + 1$ year is formed using the data from the month of January of year y_1 , where $y_1 \in [1, \dots, n]$, such that $PDF_{cum,m=1} \approx PDF_{y=y_1,m=1} \equiv PDF_{WRTY,m=1}$; the month of February of the WRTY for the $n + 1$ year is formed using the data of February of the year y_2 , where $y_2 \in [1, \dots, n]$, such that $PDF_{cum,m=2} \approx PDF_{y=y_2,m=2} \equiv PDF_{WRTY,m=2}$; and so on, until the month of December of the WRTY for the $n + 1$ year which is formed with the data of December of the year y_{12} , where $y_{12} \in [1, \dots, n]$, such that $PDF_{cum,m=12} \approx PDF_{y=y_{12},m=12} \equiv PDF_{WRTY,m=12}$. The matching is performed with the Mean Absolute Error (MAE) criteria so that the months of the WRTY are formed as $\{January_{y_1}, February_{y_2}, \dots, November_{y_{11}}, December_{y_{12}}\}$.

The wind seasonality may be found via a clustering analysis of the distribution parameters of the $PDF_{y,m}$. The resulting groups of months, with similar parameters, correspond to statistical seasons. A *PDF* for each statistical season can be calculated using the months of WRTY for the $n + 1$ year. Finally, the annual produced wind energy at $n + 1$ year is calculated by adding the produced energies at each season. Figure 1 shows the general methodology to estimate the produced wind energy in year $n + 1$, given the wind speed data of n years on a site.

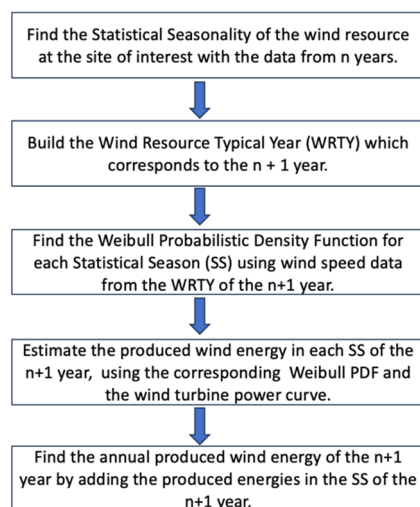


Figure 1. General methodology to estimate the wind energy on a site.

Section 2.1 describes the frequent statistical treatment of wind speed on a site. The Weibull probability distribution function, which is the most widely used to characterize the wind resource [76], is defined. Section 2.2 presents the methodology to find the wind resource statistical seasonality for year $n + 1$ using a clustering analysis with the wind speed data from n years. Section 2.3 introduces the basic theory of clustering. The construction of the Wind Resource Typical Year (WRTY) for the year $n + 1$ is described in Section 2.4. Finally, Section 2.5 outlines the estimation of the wind power generation at a given period of time. Even though this methodology considers the Weibull *PDF* to be the *PDF* that better characterizes the wind resource, it can be applied to any other *PDF*. However, as the next section describes, other *PDFs* may have one, two, three, or more parameters and are less frequently used to characterize wind resource. The greater the number of parameters, the greater the complexity of the computing process becomes.

2.1. Statistical Nature of the Wind Resource

In this study, following a statistical approach, a histogram representing the frequency of the measurements of the wind speeds per year on a site is built, and a probability distribution function (*PDF*) is fitted to the data to characterize the wind resource.

Wind speed distributions vary depending on the geographical and temporal conditions. So far, there is no model that provides a sufficient description at any site. Wind speed distributions are divided into classes: (a) parametric distribution models and (b) nonparametric distribution models. Table 3 shows some of the main wind speed distributions and their characteristics [77,78]. The nonparametric distribution models do not need to estimate the parameters of any distribution [79]. The parameters are learned from historical data. The most used nonparametric models are the Kernel Density Estimation (KDE) and the Maximum Entropy Principle (MEP) [80].

Table 3. The first column shows some of the parametric distribution models for wind speed distributions. The second column shows the reference where information about them can be found. The third column shows the number of parameters of the distribution and the last column shows their main applications in wind energy.

Distribution	Reference	# Parameters	Applications
Weibull	[76]	2	Simple form, high flexibility, and feasible computing parameters. It is the most popular <i>PDF</i> for wind speed. However, it is less efficient for low wind speeds, especially for wind speed data with small wind probability. However, it is not good for extreme winds.
3 parameter Weibull	[81]	3	Suitable for low wind speeds and small wind probability.
Rayleigh	[82]	1	It is easier to use since it has only one parameter. However, it assumes that the long-term mean wind vector is zero. Therefore, it is not used for sea winds.
Gumbel	[83]	2	It is an extreme value distribution. It is more accurate for extreme wind speeds. However, it is used to estimate the annual maximum wind speed distribution but not the monthly or daily extreme winds.
Inverse Weibull	[84]	2	An alternative to the Weibull distribution, but it provides flexibility for modeling the long-tailed right-skewed data.
Generalized extreme value	[85]	3	It is used for extreme wind speed data, but it is more difficult for computations due to the 3 parameters.
Gamma	[86]	2	It is used for fitting low wind speed data. It can be an alternative to the Weibull distribution for low ranges. However, the analytic expression is complicated.
Generalized Gamma	[87]	4	It is more flexible and good for high wind speed data. However, its analytical expression is complex due to its 4 parameters. It is suitable for European regions with different surfaces and weather conditions.
Lognormal	[88]	2	It is used for wind speed data, which change randomly.
Burr	[89]	4	It is more flexible and adaptable to wind data. It is more complex, and that makes computations more difficult. Wind speed data in Southern Italy is well-described by it.
Johnson	[90]	4	It is more flexible and adaptable to wind data. It is more complex, and that makes computations more difficult. Wind speed data measured in the Mediterranean Sea is well-modeled by it.
Kappa	[91]	4	It is more flexible and adaptable to wind data. It is more complex, and that makes computations more difficult. It is suitable for modeling onshore and offshore wind speed distribution models.
Wakeby	[91]	5	It is more flexible and adaptable to wind data. Its 5 parameters make it more complex for computing purposes. It is suitable for modeling onshore and offshore wind speed distribution models.

There are many methods used for the parameter estimation of the distribution models. The Maximum Likelihood Method (MLM), the Least Squares Method (LSM) also known as the graphic method, the Method of Moments (MOM) and the Power Density Method are the most used [78]. Once the parameters of the distribution are determined, the goodness-of-fit may be determined using some criteria. The coefficient of determination R^2 determines the similarity between the estimated and the observed data and is one of the most used. The Root Mean Square Error (RMSE) determines the accuracy of the estimated probability given the observed probability. The Kolmogorov–Smirnov (KS) and the Anderson–Darling (AD) tests determine whether a probability distribution is suitable for the wind speed data. Finally, the Chi Square Test (χ^2) verifies whether the measured wind speed data frequency is consistent with the frequency from the estimated distribution model [78].

The Weibull *PDF* is successfully used to fit the annual frequency wind speeds of many sites [92]. Therefore, in this work, for the sake of simplicity, it is used to characterize the wind resource. However, this methodology can be applied to any wind speed distribution and their parameters. It is worth mentioning that the fewer *PDF* parameters, the less complex the computation becomes. As the number of *PDF* parameters increases, the complexity of the calculations also increases.

The Weibull *PDF* is given as follows:

$$f_W(v; \lambda, k) = \begin{cases} \frac{k}{\lambda} \left(\frac{v}{\lambda}\right)^{k-1} e^{-\left(\frac{v}{\lambda}\right)^k}, & v \geq 0 \\ 0, & v < 0 \end{cases}, \quad (1)$$

where v is the wind speed and $k > 0$ and $\lambda > 0$ are the shape and scale parameters of the distribution, respectively [93]. The shape and scale parameters are related to the shape and width of the distribution, respectively. The mean wind speed, \bar{v} , of the *PDF* is given by

$$\bar{v} = \lambda \Gamma\left(1 + \frac{1}{k}\right), \quad (2)$$

and it is different to the average wind speed of the dataset. The Raleigh distribution corresponds to the special case $k = 2$. Many sites are well-characterized by either the Raleigh distribution or the more general Weibull distribution; see [94] for an example performed in India. However, other sites can be characterized with bimodal probability distributions [95,96] or different distributions.

2.2. Seasonality

A cluster analysis consists of grouping data based on their similar characteristics or attributes. It can be carried out considering one attribute, one-dimensional cluster analysis; two attributes, two-dimensional cluster analysis; or n attributes, n -dimensional cluster analysis. Even though an annual wind speed *PDF* is adequate for studying the wind resource, a more detailed analysis involves a seasonal analysis. Hence, a cluster analysis to ascertain the seasonality of the wind resource on a specific site is proposed, where the characteristics of the wind speed's *PDF* are used to find the seasonality of the wind resource.

Wind speed data from n available years are gathered every month. Then, the Weibull *PDFs* are fit to the data, and scale and shape parameters, together with the velocity of the data, are obtained for each the 12 months of the n available datasets. Next, we perform a 1D, 2D, and, 3D cluster analysis with the scale parameter; scale, and shape parameters; and scale and shape parameters and velocity of the *PDFs*, respectively. The yielded results permit the selection of the number of seasons and the months corresponding to each season. Figure 2 describes this procedure and shows that it can be extended to other *PDFs*.

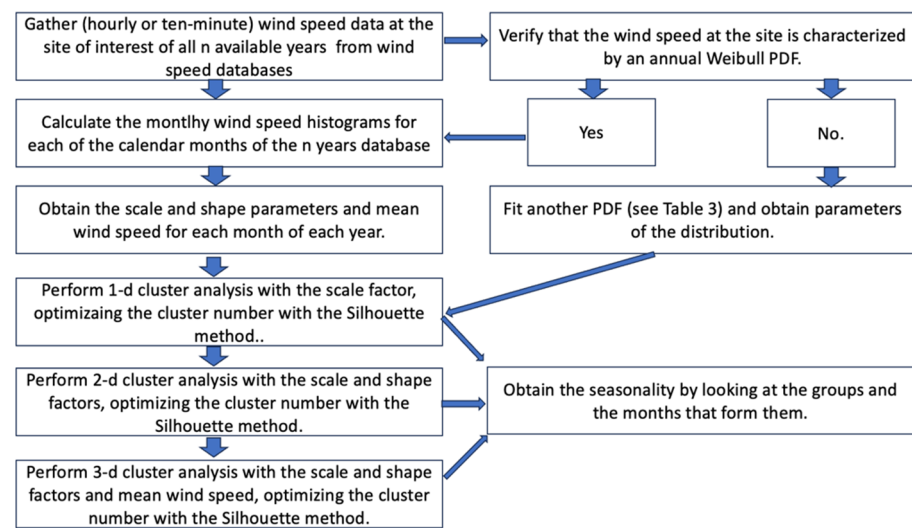


Figure 2. Methodology to determine the seasonality.

2.3. Clustering

There are many algorithms used for clustering analysis, with K-means being one of the most used unsupervised ones. K-means is based on Euclidean distance minimalization [97,98]. Given a set of m data (x_1, x_2, \dots, x_m) , where each data point is a d -dimensional vector, the K-means algorithm groups it in $k (\leq m)$ clusters $C = \{C_1, C_2, \dots, C_k\}$, such that the Euclidean distance between the objects and the mean of the points μ_i in the centroids C_i is minimized:

$$\operatorname{argmin} \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (3)$$

Qiu and Joe recommend that the minimum sample size should be of 10 times the number of variables to consider the number of clusters, $10 \times d \times k$, where d is the number of variables for equally sized clusters; otherwise, $10 \times d$ elements should contain the smallest cluster [99,100].

There are many ways to initialize the centroids. Traditionally, data points from the dataset are randomly chosen as cluster centers and the distances from the rest of the points to the centroids are calculated to find the minimum distance, which usually uses the Euclidean metric. Then, the clusters are re-centered and the distances between points and the centers are calculated again, in an iterative process, until the centroids converge to a value and the minimum distance, Equation (3), is achieved. In this work, the first centroid is randomly selected from a data point, and, then, the subsequent centroids are chosen from the remaining points based on a probability proportional to the square distance away from the nearest-neighboring centroids. Thus, the elements of a cluster are close to their centroids, but the centroids are far away from the clusters. Then, even though there are differences in the results due to centroid initialization, the optimal centroids are obtained eventually [101].

Cluster initialization depends on the number of clusters. The best number of groups is obtained by using the Silhouette method [102]. This method tests the goodness of the grouping when the dataset exhibits an intrinsic natural number of groups. A Silhouette value quantifies the similarity between objects in the same cluster compared with objects in other clusters. The distance between a data point i in the i -th cluster C_i and all other data in the same cluster are given as follows:

$$a(i) = \frac{1}{|C_i - 1|} \sum_{j \in C_i, i \neq j} d(i, j), \quad (4)$$

where the distance between the points i and j in the cluster C_i is given by $d(i, j)$. Moreover, the smallest mean distance of point i to all points in the other groups is given as follows:

$$b(i) = \min \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j). \quad (5)$$

A measure of the goodness of the grouping is given by the Silhouette score SS , which corresponds to the mean of the $s(i)$ overall data of the group, where

$$s(i) = \begin{cases} \frac{b(i) - a(i)}{\max\{a_i, b_i\}}, & \text{if } |C_i| > 1 \\ 0, & \text{if } |C_i| = 1 \end{cases} \quad (6)$$

and $-1 \leq s(i) \leq 1$. The data are well-grouped when $s(i) = 1$, and should be regrouped when $s(i) = -1$. The best number of clusters k is given by the Silhouette coefficient SC given by the maximum $s(i)$ over all data,

$$SC = \max k. \quad (7)$$

Consequently, the seasonality is detected by grouping the months according to their *WPDF* parameters and average wind speed. The first clustering analysis is a one-dimensional clustering that groups months over n years with similar scale parameters; the second is a two-dimensional clustering, grouping months with similar scale and shape parameters; the third, which is a three-dimensional clustering, groups months with similar average wind speeds and scale and shape parameters. Therefore, two-, three- or four-month groups correspond to wind resource seasonality. Depending on the height and width of their *PDFs*, the statistical seasonalities are referred to as High–High Statistical Seasonality (HHSS), High Statistical Seasonality (HSS), Low Statistical Seasonality (LSS), and Low–Low Statistical Seasonality (LLSS). This procedure is demonstrated in detail in the case study shown in Section 3.

2.4. Wind Resource Typical Year WRTY

The Typical Meteorological Year (TMY) is a set of representative meteorological data with data values for every hour in a year for a given geographical location. These data emulate the most usual weather conditions at the location and provide yearly averages consistent with their long-term averaging [103]. The values constructed from 12 months are chosen as the most typical from a dataset of at least 12 years of meteorological data. The original TMY was generated by Sandia National Laboratory [104]. This data set was generated for 248 locations in the United States using weather data from 1952 to 1975. In 1994, the TMY was updated by the National Renewable Energy Laboratory (NREL) as TMY2. The TMY employed more stations and used data from 1961 to 1990 [105]. TMY3 is the latest TMY collection created in 2005 using data from 1976 to 2005 [106]. TMY3 covers 1020 locations in the US and has become the standard set of weather databases for computer simulations of solar energy such as TRNSYS [107] and PVsyst [108]. The most recent TMY database collection is called TMYx, published by the creators of EnergyPlus 23.2.0 software [109]. TMYx is available for more than 13550 locations globally and uses data from 2006 to 2021 from the European ReAnalysis (ERA5) [110] data for climate change analysis framed within the Copernicus Climate Change Service (C3S) of the European Centre for Medium-Range Weather Forecasts (ECMWF) [111]. ERA5 climate reanalysis is based on satellite data from the METEOrological SATellites (METEOSAT), Geostationary Operational Environmental Satellites (GOES), and National Oceanic and Atmospheric Administration (NOAA) satellites.

There are many methods to obtain the TYM depending on the purpose and the availability of the data. The Sandia method uses a data set of n years of data. Then, for each month, the n data are examined; the one considered most typical is selected for each month

and combined to form a complete TMY. Depending on the method, different weather variables are considered when choosing the typical month, such as

- Global horizontal radiation, direct normal radiation, dry bulb temperature, dew point temperature, and wind speed.
- Maximum, minimum, and mean dry bulb and dew point temperatures; the maximum and mean wind velocity; and the total global horizontal solar radiation.
- Monthly mean and median and the persistence of weather patterns.

Moreover, given the probabilistic nature of the wind resource, a Gaussian Distribution (GD) is also used to generate TMY files for P50 and P90, and TMY P50 and TMY P90, representing the average climate conditions that are more probable to occur in the best and worst scenarios.

In this work, a statistical methodology is proposed to build the Wind Resource Typical Year (WRTY) by considering the statistical nature of the wind resource. The wind resource is characterized by using the Weibull Probability Distribution Function (WPDF), as described in the previous subsection. Therefore, a WPDF can be fitted to the hourly wind speed data for each month m for each year y , $WPDF_{y,m}$, where $m = 1, \dots, 12$, and $y = 1, \dots, n$. Moreover, one can characterize each month by considering all data from the n years and obtain a characteristic WPDF for month m , $WPDF_{cum,m}$. The assembling of all months forms the Characteristic Year (CY). Then, the representative month m , which will be part of the WRTY, corresponds to the month m of the year y_1 with a WPDF closest to the $WPDF_{cum,m}$ of the month m , such that $WPDF_{cum,m} \approx WPDF_{y_1,m} \equiv WPDF_{WRTY,m}$, as described in Figure 3.

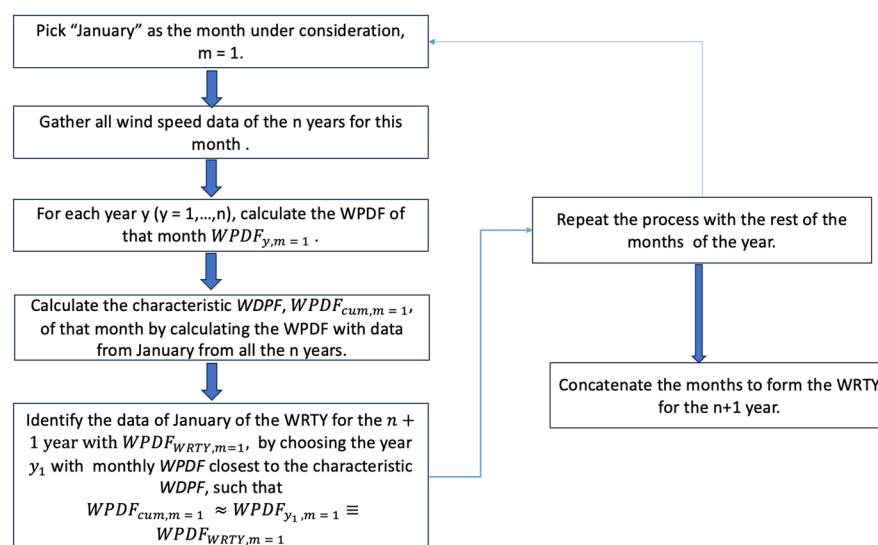


Figure 3. Methodology to construct the Typical Meteorological Year for the year $n + 1$ for the site of interest.

2.5. Estimating Wind Energy

The power output of a wind turbine is a function of the wind speed, air density, and the area swept by the blades of the wind turbine; in particular, the wind turbine's power curve describes the generated power as a function of the wind speed [79]. This curve specifies the cut-in speed at which the wind turbine generates power, the rated power output, and the cut-out speed at which the wind turbine stops working due to security reasons, as shown in Figure 4.

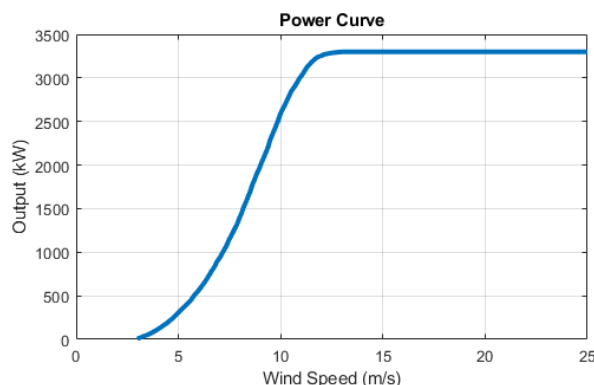


Figure 4. Power curve of wind turbine Vestas V112 3.3 MW from the Technical Sheet [112].

Thus, the produced average power $P_{average}$ is given by integrating the power curve over the wind speed PDF [113,114]

$$P_{average} = \int_{v_i}^{v_f} P_{turb}(v) f_{PDF}(v)dv W, \tag{8}$$

Additionally, the maximum annual generated energy E_{max} by a wind turbine is given as follows

$$E_{max} = 8760 \int_0^{v_f} P_{turb}(v) f_{PDF}(v)dv Wh, \tag{9}$$

where W stands for Watts and h for hours, respectively. However, in practice, it depends on the capacity factor (CF), which is the ratio of annual energy output, E_{real} to the theoretical maximum output, E_{max} , considering that the turbine was operating at its rated power during 8760 h of the year.

$$CF = \frac{E_{real}}{E_{max}}. \tag{10}$$

The CF depends on many factors such as the maintenance of the wind farm.

2.6. Forecasting Error

- The accuracy of the forecasting can be determined by using different metrics; for instance, the forecast bias, the Mean Average Deviation (MAD), and the Mean Average Percentage Error (MAPE). The MAPE is one of the most popularly used and is given as [115]:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|, \tag{11}$$

where A_t and F_t are the actual and the forecasted values of the quantity to be forecasted. The forecasting performance depends on the context of the forecastability of the data. In many cases, the MAPE score indicates the following accuracy.

$$MAPE = \begin{cases} < 5\% & \text{Excellent accuracy} \\ (5\% , 10\%) & \text{Acceptable accuracy} \\ (10\% , 25\%) & \text{Low but acceptable accuracy} \\ > 25\% & \text{Very low accuracy not acceptable} \end{cases} \tag{12}$$

- The prediction horizon refers to how far ahead the model predicts the future. The longer the forecast horizon, the less accurate the prediction is. The intermittency of the wind resource and its intrinsic relation to the local conditions of the site of interest made the long-term forecasting of wind power a complex task.
- The next section considers a case study to apply the proposed methodology to estimate wind power production at different prediction horizons. The wind power produced by

a wind farm in northern Mexico is analyzed and a detailed calculation of the previous methodology is described.

3. Results

In this section, a case study is carried out to demonstrate the proposed procedure. The Energía Sierra Juárez onshore wind farm is selected since its information is available on the internet [116]. It is located in Baja California, Mexico, close to the La Rumorosa municipality, at (32.56, −116.07). The wind farm is made of 47 Vestas V112 3.3 W wind turbines and yields a capacity generation of 155 MW.

Wind speed databases are freely available in the POWER Data Access Viewer of NASA [117]. This data repository counts more than 30 meteorological parameters corresponding to solar radiation, temperatures, humidity, precipitation, wind, and pressure, with an hourly, daily, monthly, and annual frequency for a site located using its latitude and longitude in any region of the globe. The data are available from 1981 to the present day. The output file format can be ASCII, CSV, GeoJSON, or NetCDF. The wind speed data are available at 2, 10, and 50 m above the surface. The wind speed data are obtained from the project Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) [118].

The following calculations were performed in Python 3.9.17, and the following libraries were used: (a) Matplotlib 3.7.2 to create plots and graph data, (b) Numpy 1.24.3 for array manipulation, (c) Pandas 2.0.3 for data manipulation in xls or csv forma, (d) Scipy 1.11.1 for manipulation and tuning of *PDFs*, and (e) Sklearn 1.2.2 for clustering analysis. The code can be found at the end of the article in the Supplementary Materials section.

3.1. Case Study

Wind speed data at 50 m at La Rumorosa were collected from 2001 to 2022. The data were used to analyze the wind resource at the site as described in Sections 3.1.1–3.1.3. The technical information for Vestas V112 3.3W wind turbine [112] is used to forecast wind energy production in Section 3.1.4.

As mentioned in Section 2.1, Weibull *PDF* was chosen to develop long-term wind power forecasting in the presented methodology. However, wind speed data from a random year were used to fit it to different *PDFs*. The data were fitted with the Least Squares Methods, and they are shown in Figure 5. From a visual inspection, it is clear that the Weibull *PDF* is the one that best fits the wind speed data. However, to test the goodness of fit, the RSME was calculated, and the results are shown in the upper right corner of Figure 5. The best fit corresponds to the Weibull *PDF*, followed by the Gamma and the Lognormal. Thus, the Weibull *PDF* is a good characterization of the wind resource at this site.

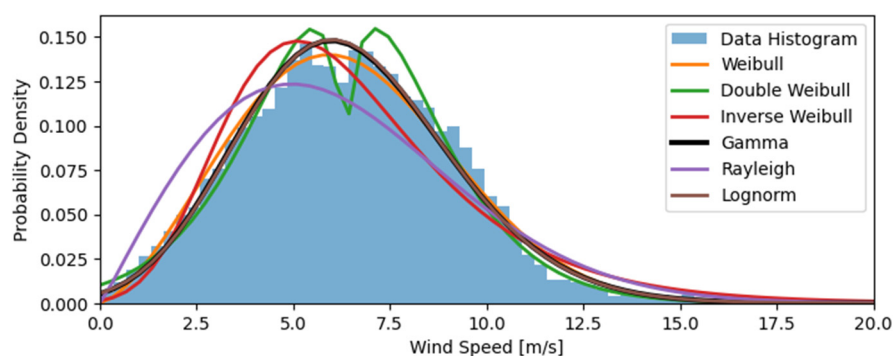


Figure 5. Characterization of wind resource at the location of interest by fitting data to a *PDF*. Histogram of annual wind speed data and 6 probability distribution functions fitted to the data. The values of the RMSE for the different probability distributions are shown in the upper right corner.

3.1.1. Cluster Analysis

Cluster analysis is performed to find the wind seasonality at the site of interest. In this section, we perform the cluster analysis for the whole set of available wind speed data from 2001 to 2022. Of course, following the methodologies shown in Figures 1 and 2, the cluster analysis should be carried out for the n available years, but our scope is to show the results for all forecasted years, 2015 to 2022, in the same figure. First, the Weibull PDFs are fitted to the monthly data from n years. In this case, we consider $n = 22$, considering the period 2001–2022. The resulting data set consists of $12 \times n = 12 \times 22 = 264$ scale and shape parameters, (λ, k) , corresponding to the Weibull PDFs for the 12 months of the 22 years, and it is enough data to perform the clustering analysis, as explained in Section 2.2. A grouping analysis for these parameters is performed to find the seasonality at the site, as depicted in Figure 6.

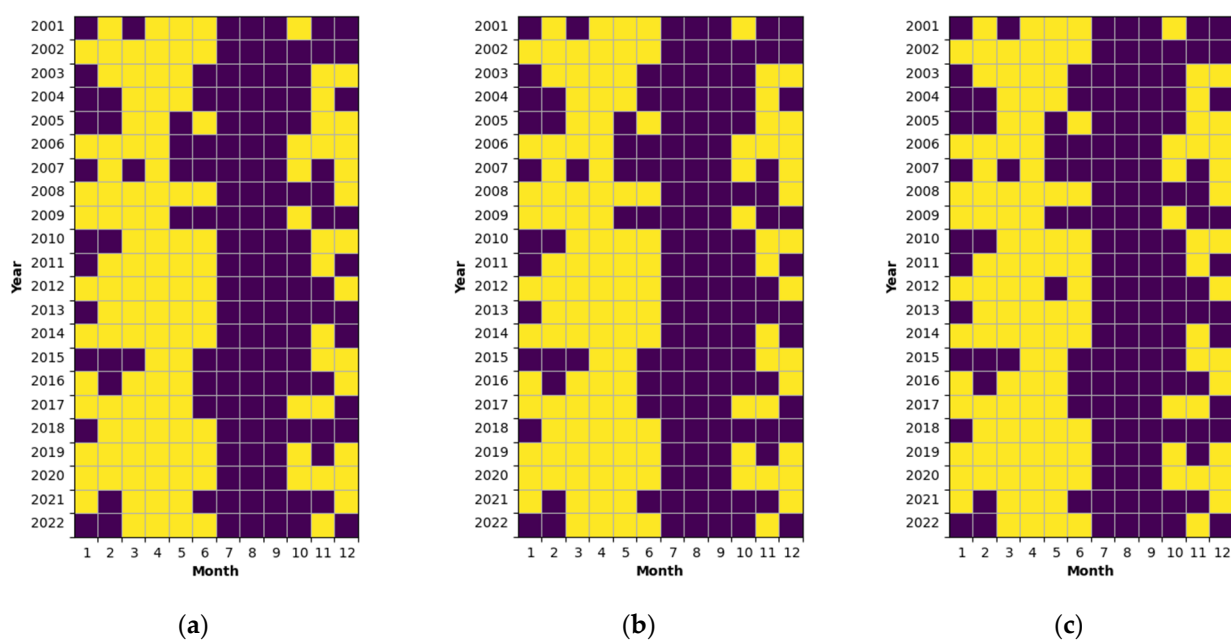


Figure 6. Cluster analysis for the wind speed data corresponding to the 12 months from the period 2001 to 2022; (a) 1D clustering for the scale parameter λ ; (b) 2D clustering for the scale and shape factor, λ and k ; (c) 3D clustering for λ , k and average wind speed. The purple and yellow colors correspond to the High Statistical Seasonality (HSS) and Low Statistical Seasonality (LSS) months, respectively.

The year and month labels correspond to the y and x coordinates of each graph, respectively. Figure 6a shows the one-dimensional grouping for the scale parameter λ . Figure 6b shows the two-dimensional grouping for the scale and shape parameters λ and k . Figure 6c shows the three-dimensional grouping for the scale and shape λ and k parameters and average wind speed. The reason that the first two figures are the same and the third almost the same is that the scale parameter is the dominant parameter in this analysis. For a smaller number of years, the next subsections are elaborated for $n = 14, \dots, 21$, and the clustering results are the same. Indeed, the corresponding graphs are the ones in Figure 6 but truncated at the n year. The reason we used all the data generated over 22 years is to show that the statistical seasonality remains throughout the years.

In this analysis, the Silhouette method is used to determine the best number of clusters. The highest SC is obtained for two clusters for each of the 1D, 2D, and 3D cluster analyses. As seen in Figure 6, there are two main statistical seasons (SSs) at the site; the purple color shows the first statistical season corresponding to the July–August–September group, and the yellow color represents the second statistical season corresponding to the rest of the months. The fact that the clusters are not 100% determined for most of the months is due to

the statistical nature of the resource. However, the July–August–September cluster, which, from now on, we refer to as the High Statistical Seasonality, HSS, is well-defined for the three cluster analyses. The Low Statistical Seasonality, LSS, corresponds to the rest of the months. Thus, the wind resource at the site is characterized by these two statistical seasons. Figure 7 shows the 3D cluster analysis. As shown, the month grouping and the SSs are well-defined.

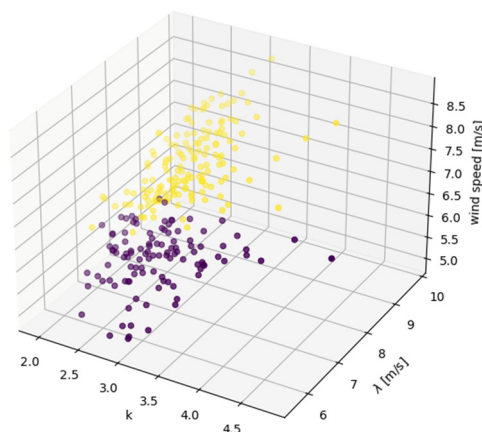


Figure 7. Three-dimensional cluster analysis for the scale factor, λ , shape parameter, k , and average wind speed. Purple and yellow dots correspond to the HSS and LSS, respectively.

Once the statistical seasons for the site are obtained, the Weibull *PDF* for each SS must be computed to estimate the produced electrical energy using the months of the WRTY for the HSS and the LSS. The WRTY to represent the monthly wind resource at the site is constructed in the next subsection.

3.1.2. Wind Resource Typical Year

In this section, we exemplify the construction of the WRTY for $n = 17$ years, from 2001 to 2017. The WRTY is constructed for La Rumorosa, initially by building monthly Weibull *PDFs*. Each month of the year data from the 17-year database ($n = 17$) is gathered and a $WPDF_{cum, m}$ is fitted; hence, 12 $WPDF_{cum, m}$ are obtained that depict each of the monthly wind resources at the site, as outlined in Section 2.4. The red curves in Figure 8 show the representative monthly Weibull *PDFs*. Then, a Weibull *PDF* is fitted for each of the 12 months $WPDF_{y, m}$ of the 17-year sample data, as shown in the color curves in Figure 8. Next, the forecasted WRTY for the year 2018 ($n + 1 = 17 + 1 = 18$), is built by selecting the year data where their *PDF* is closest to the representative *PDF* for each month $WPDF_{cum, m} \approx WPDF_{y, m} \equiv WPDF_{WRTY, m}$. In Figure 8, the month of the WRTY is indicated in red color and its corresponding characteristic *PDF* is in blue.

Table 4 shows the monthly Weibull *PDF* parameters for the characteristic year and the WRTY for 2018. As shown, the months of the WRTY are chosen such that their scale and shape factors are the closest to the ones of the characteristic *WPDF* compared to the rest of the ones of the same year.

Figure 9 shows the normalized wind speed histograms for the characteristic year, formed by all the data from 2001 to 2017, and the WRTY for 2018 built from that year's interval.

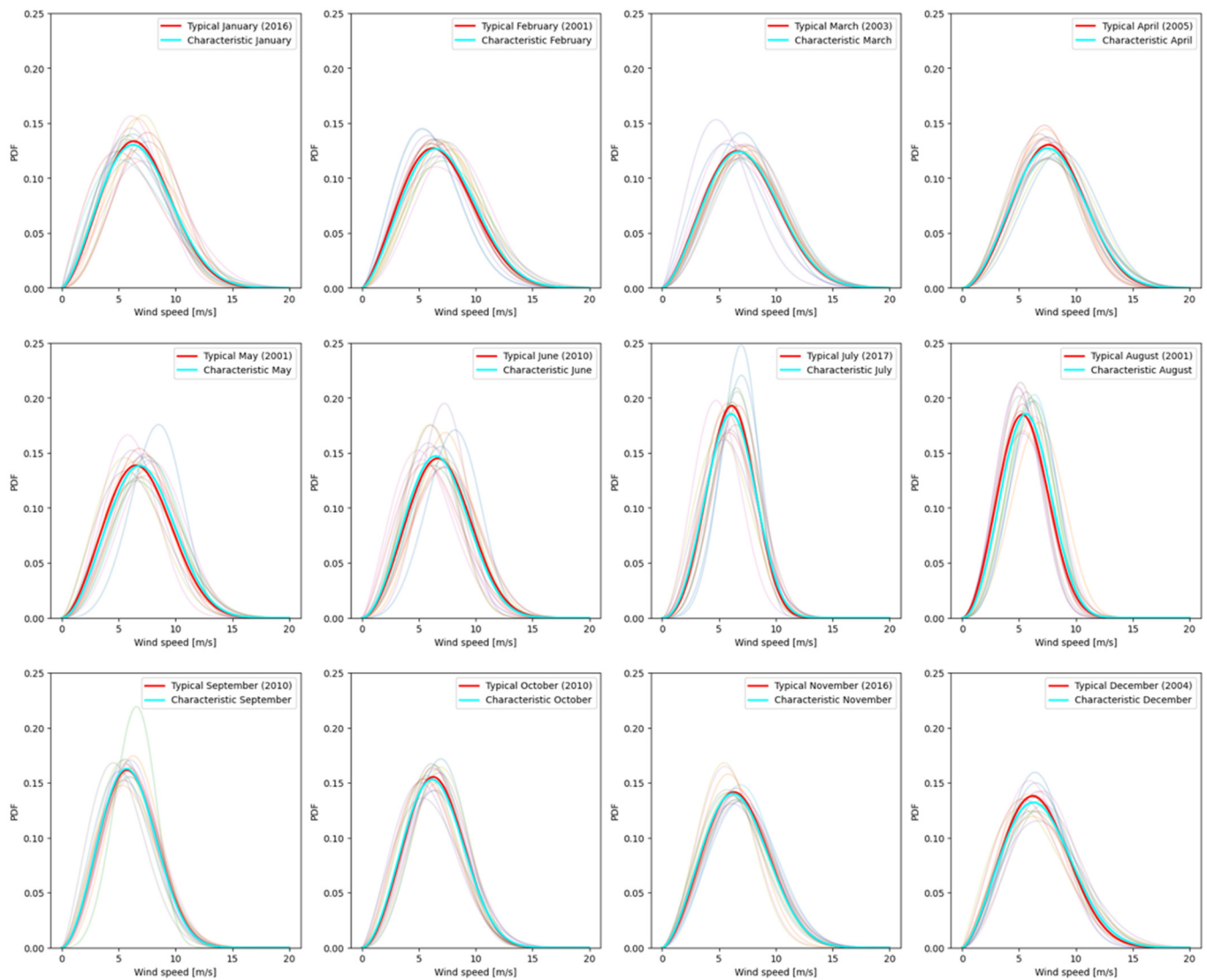


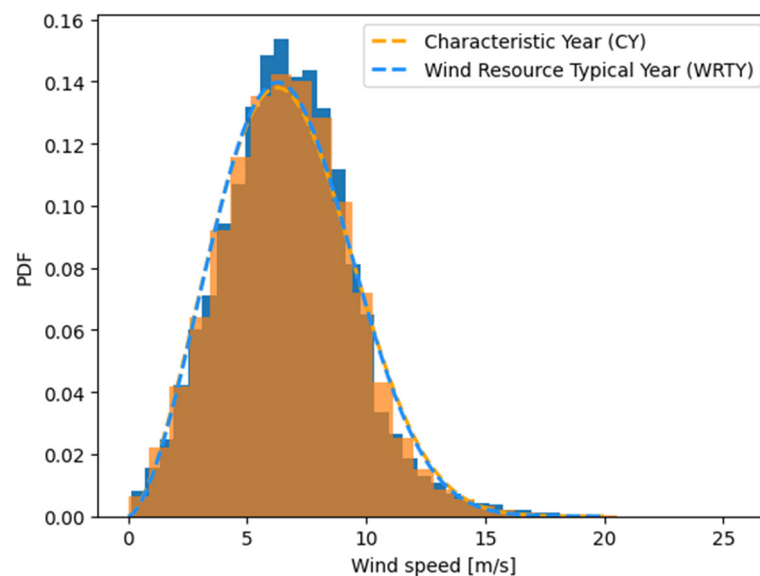
Figure 8. Construction of the WRTY for the site of interest for the period 2001–2017. Each graph corresponds to Weibull PDFs for the wind speed at the site for the twelve months of a year. $WPDF_{cum, m}$ for each month are shown using blue curves in each of the graphs. The color curves correspond to the $WPDF_{y, m}$ for each month of the 17-year database. For each month, the year in which the Weibull PDF $WPDF_{y_1, m}$ is the closest to the $WPDF_{cum, m}$ (blue curve) is chosen as the month of the WRTY for 2018 with a monthly $WPDF_{WRTY, m}$ (red curve).

Table 4. Monthly Weibull PDFs parameters for the Characteristic Year for the 2001–2017 year interval and for the Wind Resource Typical Year for 2018. In the first column, the corresponding month is indicated. The number in parenthesis indicates the year in which the data of that month were chosen to assemble the WRTY for 2018. The Weibull PDFs parameters for the characteristic year and the WRTY for 2018 are in the 2nd–3rd and 4th–5th columns, respectively.

Month	Characteristic Year		WRTY	
	k	λ [m/s]	k	λ [m/s]
January (2016)	2.45	7.68	2.54	7.67
February (2001)	2.51	7.94	2.42	7.75
March (2003)	2.51	8.27	2.50	8.12
April (2005)	2.77	8.69	2.90	8.77

Table 4. *Cont.*

Month	Characteristic Year		WRTY	
	k	λ [m/s]	k	λ [m/s]
May (2001)	2.78	8.05	2.70	7.75
June (2010)	2.87	7.61	2.85	7.75
July (2017)	3.19	6.76	3.38	6.77
August (2001)	3.01	6.41	2.87	6.12
September (2010)	2.73	6.67	2.74	6.74
October (2010)	2.73	7.23	2.86	7.26
November (2016)	2.61	7.47	2.66	7.52
December (2004)	2.50	7.69	2.57	7.49

**Figure 9.** Wind speed frequency histograms from all data from the period 2000–2017 for the characteristic year (orange dotted line) and WRTY (blue dotted line) for 2018.

The Weibull *PDF* parameters for both cases are shown in Table 5. Moreover, the parameters for the rest of the year intervals are also presented in the same table. As expected, both *PDFs* and their corresponding parameters are very similar.

Table 5. Annual Weibull *PDF* parameters for the Characteristic Year and for the Wind Resource Typical Year constructed using the wind speed data of years in the indicated interval.

Years	CY		Forecasted	WRTY	
Interval	k	λ [m/s]	Year	k	λ [m/s]
2001–2014	2.617	7.542	2015	2.636	7.555
2001–2015	2.618	7.537	2016	2.630	7.539
2001–2016	2.610	7.546	2017	2.612	7.541
2001–2017	2.601	7.541	2018	2.616	7.552
2001–2018	2.596	7.549	2019	2.603	7.560
2001–2019	2.605	7.554	2020	2.630	7.600
2001–2020	2.602	7.551	2021	2.627	7.566
2001–2021	2.600	7.553	2022	2.625	7.585

In the next subsection, the statistical seasonality is analyzed using the WRTY.

3.1.3. Statistical Seasonality

From Section 3.1.1, the wind resource at La Rumorosa is characterized by two statistical seasonalities, corresponding to the HSS formed from July, August, and September and the LSS by the rest of the months of the year. To estimate a more accurate electrical energy production, the wind speed data from the WRTY, for the $n + 1$ year, is used to calculate the Weibull PDFs for both SSs. Figure 10 shows the resulting Weibull PDFs for the HSS, LSS, and the WRTY for the forecasted year 2018.

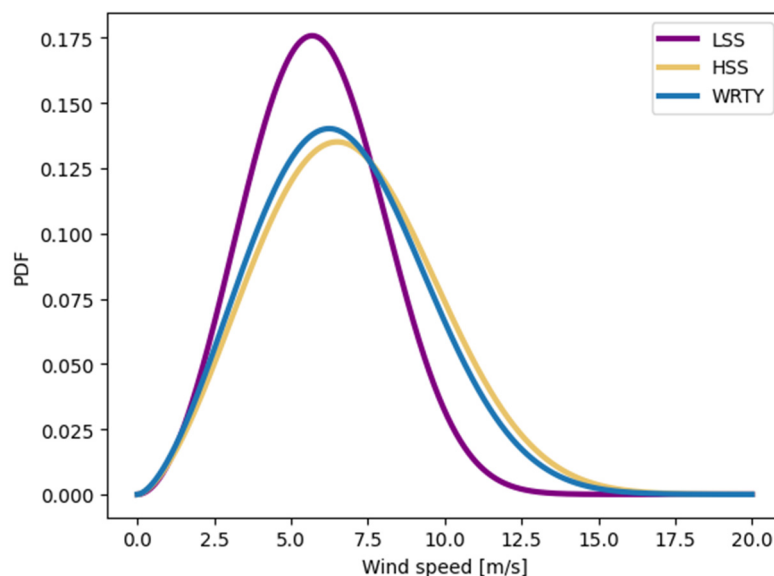


Figure 10. Wind speed Weibull PDFs calculated from the dataset corresponding to the year interval (2001–2017). The purple and yellow curves correspond to the High Statistical Season and Low Statistical Season, respectively, for 2018. The blue curve corresponds to the WRTY for 2018.

The shape and scale parameters of the distributions calculated from the year intervals (2001–2017), corresponding to Figure 10, are shown in Table 6. As expected from Figure 10, the distribution of the HSS is higher and narrower than that of the LSS. The distribution of the WRTY for the $n + 1$ is closest to that of the LSS; it is weighted by that of the HSS. Additionally, the WPDF parameters for the rest of the year periods $[1, n]$ and for the WRTY for $n + 1$ are also presented in Table 6.

Table 6. Scale and shape parameters, k and λ , for the Weibull PDFs for the two statistical seasons, HSS and LSS, and for the WRTY for the year $n + 1$ constructed using n years in the interval indicated in the first column.

Years	HSS		LSS		WRTY		
	Interval	k	λ [m/s]	k	λ [m/s]	k	λ [m/s]
2001–2014		2.942	6.568	2.614	7.766	2.613	7.471
2001–2015		2.942	6.568	2.623	7.790	2.619	7.489
2001–2016		2.930	6.553	2.634	7.796	2.624	7.490
2001–2017		2.930	6.553	2.590	7.815	2.585	7.505
2001–2018		2.972	6.684	2.623	7.790	2.631	7.517
2001–2019		2.972	6.684	2.625	7.814	2.632	7.536
2001–2020		2.972	6.684	2.625	7.814	2.632	7.536
2001–2021		3.017	6.659	2.633	7.844	2.641	7.553

The estimation of the electric power produced by the wind resource is performed with the distributions calculated previously and the power curve of the wind turbine, as shown in the next subsection.

3.1.4. Estimating the Electrical Energy

As described in Section 2.5, the electric power produced by a wind turbine is calculated using its power curve and the Weibull *PDF* of the wind resource at the site. This calculation is carried out by integrating Equation (8) in a continuous, numerical, or discrete way. We estimate the generated electric power discretely by considering the corresponding Weibull *PDFs* bin by bin.

The power curve of the wind turbines of the farm, Vestas V112 3.3 MW, is used to estimate the amount of electrical energy produced. This curve is shown in Figure 4 and can be approximated using the following equation:

$$P_{turb}(v) \approx \begin{cases} 2190 - 1170 v + 194.6 v^2 - 7.48 v^3, & 3.5 < v < 12 \\ 3300, & 12 < v < 25 \end{cases} \quad (13)$$

where P is the electric power and v is the wind speed.

Using the results of the previous subsection, the energy produced in the HSS and the LSS for each year interval is calculated by building the *WPDF* for each SS and integrating it with the power curve. Then, the annual produced energy corresponds to the sum of the generated energy at each SS. Table 7 shows the results. The first column shows the year interval over which the WRTY was calculated. Then, the second and third columns correspond to the energy produced in the HSS and LSS, respectively. These energies are calculated by computing the *WPDF* for each SS and integrating it over the wind turbine power curve. Finally, the fourth column corresponds to the annual energy for the WRTY for the corresponding year interval. This energy is computed by adding up the produced energy in both SSs. It is important to remark that the calculation is carried out by considering the wind turbine's full operation, i.e., 365 days, 24 h per day.

Table 7. Calculated wind energy for the High Statistical Season, Low Statistical Season, and the forecasted year using the *WPDF* of the WRTY corresponding to the HSS and the LSS.

Year Interval	Forecasted Year	E_{HSS} [GWh]	E_{LSS} [GWh]	E_{WRTY} [GWh]
2001–2014	2015	1.622	7.503	9.125
2001–2015	2016	1.622	7.469	9.091
2001–2016	2017	1.622	7.517	9.139
2001–2017	2018	1.612	7.527	9.139
2001–2018	2019	1.612	7.574	9.186
2001–2019	2020	1.699	7.517	9.216
2001–2020	2021	1.699	7.566	9.265
2001–2021	2022	1.676	7.627	9.303

According to the implemented methodology, the energy values in the year interval $[1, n]$ correspond to the forecasted energy for the year $n + 1$. These results were validated by comparing these energies with the energies explicitly computed from the dataset corresponding to the years 2018 to 2022.

Even though the results reported in Table 7 estimate the produced energies, there is an uncertainty associated with them. This uncertainty is mainly related to the natural fluctuations of the resource and manufacturing tolerances since the power output of a wind turbine differs from the value given by its power curve, model simplifications, and losses. The typical uncertainty associated with the wind energy yield prediction ranges between 8% and 20%. The uncertainty, U , in this case study, is assumed to be $U = 11\%$.

Due to the statistical nature of the wind resource, the results are usually reported within a Probability of Exceedance (PoE) of the Annual Energy Production (AEP). The AEP estimates the annual energy that will be reached with a 50% probability and corresponds to the base case. It is assumed that the values for the produced energy fall into a normal distribution centered around the expected forecast, i.e., the AEP, and the variance corresponding to the uncertainty on the energy yield prediction. The normal distribution provides the

cumulative probability that the forecast average level exceeds PoE. The AEP predicted via the wind data analysis has a 50% PoE, which is denoted as P50. In the case of P75, the AEP has a 25% probability of not reaching the AEP [119].

AEP at Probability of Exceedance XX%, AEP@PXX, is calculated as follows:

$$\text{AEP@PXX} = \text{AEP@P50} \times (1 - \sigma \times \text{NORMINV}(\text{XX}\%, 0, 1)), \quad (14)$$

where σ is the standard deviation of the normal distribution with mean $\mu = \text{AEP} = \text{AEP@P50}$, and it is calculated as follows:

$$\sigma = \mu \sigma_p \quad (15)$$

where σ_p is the percentage relative standard deviation of the distribution and is equal to the uncertainty $\sigma_p = U$. Finally, $\text{NORMINV}(x, 0, 1)$ is the inverse of the normal cumulative distribution with a mean equal to 0 and a standard deviation equal to 1.

The PoE level is considered when determining the sensitivity of project financing. Usually, banks apply the P90 or P95 level for their revenue forecast to determine if the interest cover is enough. On the other hand, equity investors use the P75 or even P50 levels. Table 8 shows the most used PoE cases of sensitivity.

Table 8. Probability of Exceedance (PoE) sensitivity cases.

Case of Sensitivity	Financial Situation
P50	Base case 1
P75	Base case 2
P90	Worst case 1
P95	Worst case 2

Figure 11 shows the PoE for the predicted Energy Productions of the High Statistical Seasons (EP_HSS) for the years 2015 to 2022. An uncertainty of 11% is considered for this case study. The forecasted EP_HSS@P50 lies between 1.6 and 1.7 GWh, as expected from the information shown in Table 7. In the worst cases, the forecast for the EP_HSS lies between 1.3 and 1.4 GWh, which are the scenarios with the highest probability. In the best cases, the forecast for the EP_HSS lies between 2.0 and 2.2 GWh, where the scenarios are less probable.

Figure 12 shows the PoE for the predicted energy production in the Low Statistical Seasons (EP_LSS) for the years 2015 to 2022. An uncertainty of 11% is considered for this case study. The forecasted EP_LSS@P50 lies between 7.4 and 7.7 GWh, as expected from the information shown in Table 7. In the most probable scenarios, the forecasted EP_LSS lies between 6.1 and 6.3 GWh. In the less probable scenarios, the forecasted EP_LSS lies between 9.3 GWh and 9.6 GWh.

The AEP is given by the sum of the EP_HSS and the EP_LSS. Figure 13 presents four different estimated AEPs for all years, considering the EP_HSS and the EP_LSS with different PoEs over the last eight years. A higher PoE level results in a lower AEP. In the most likely scenario, the annual forecasted energy ranges from 7.3 to 7.7 GWh. On the other hand, in the less likely scenario, the annual forecasted energy ranges from 9.0 to 9.4 MWh.

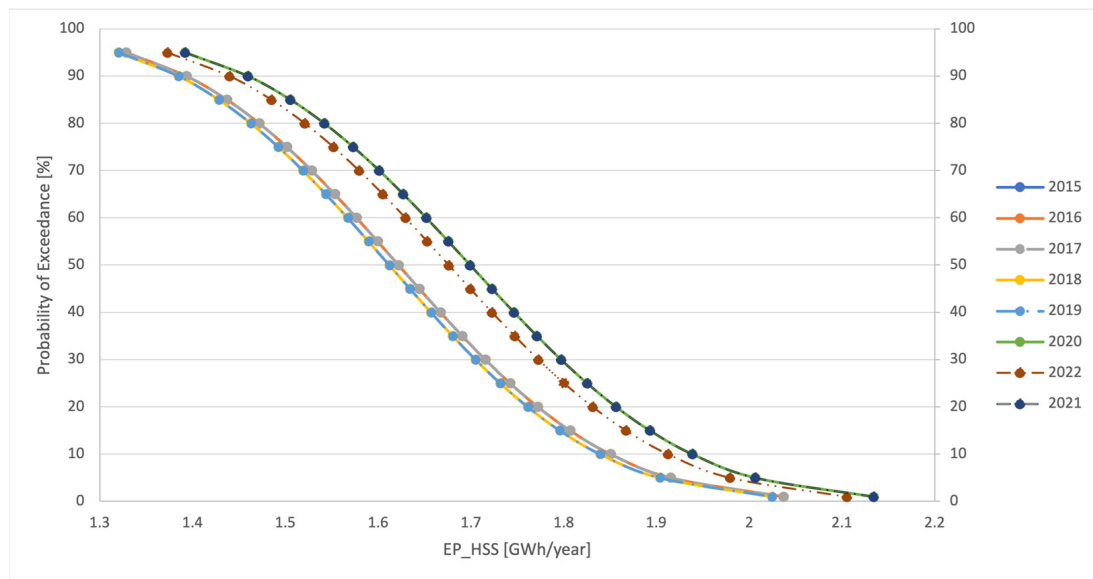


Figure 11. Forecasted energy production for the HSS with Probabilities of Exceedance and 11% uncertainty from 2015 to 2022. The EP_HSS per year and the PoE are represented on the horizontal and vertical axes, respectively.

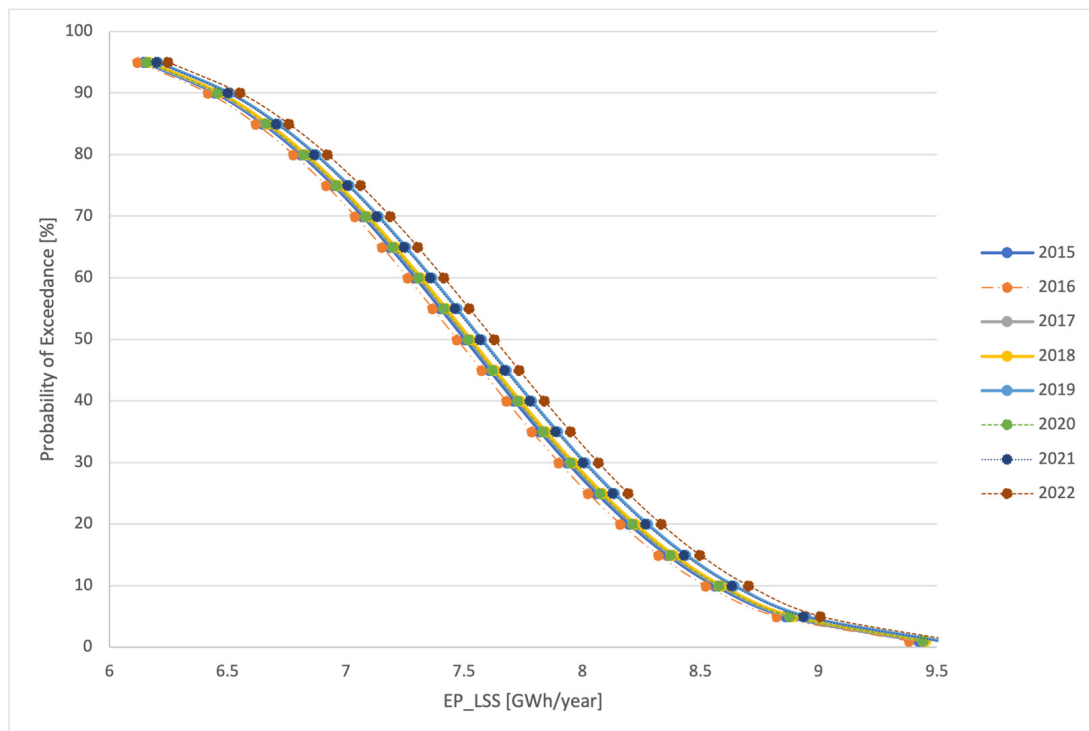


Figure 12. Forecasted energy production for the LSS with Probabilities of Exceedance and 11% uncertainty for 2015 to 2022. The EP_LSS per year and the PoE are represented on the horizontal and vertical axes, respectively.

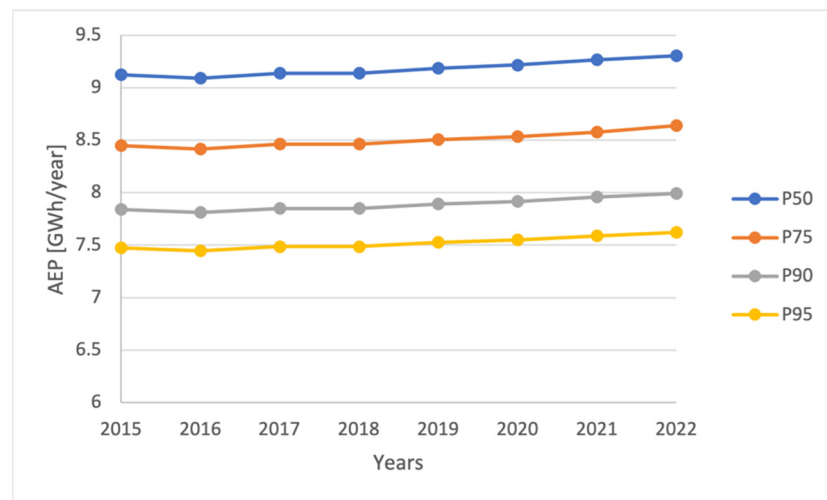


Figure 13. Annual energy production prediction for 2015 till 2022 at Probabilities of Exceedance P50, P75, P95 and P50.

3.1.5. Forecasting Errors

To test the accuracy of these results, the annual produced energy, using real data, is computed, and the hourly values of the wind speed from the database of the year under consideration are considered. The hourly produced energy by the wind turbine is calculated according to Equation (9) and adding them all. In the discrete case, this equation becomes

$$E_{annual} = \sum_{i=1}^{8760} P(v_i), \quad (16)$$

where v_i corresponds to the hourly wind speed of the year, the index i runs from 1 to 8760, and $P(v_i)$ is the wind turbine power calculated using Equation (13). The computed annual energy production, considering that the wind turbine operates 365 days, 24 h per day, is shown in Table 9.

Table 9. Calculated annual wind energy for the year shown in the first column using raw data from the wind speed database, considering that the wind turbine operates 365 days, 24 h per day. These results correspond to the reference values.

Year	E_{annual} [GWh]
2015	8.589
2016	9.053
2017	9.686
2018	9.134
2019	9.560
2020	9.700
2021	9.129
2022	9.601

The forecasted and calculated AEPs@P50, corresponding to the values in the fourth column of Table 7, are graphed in Figure 14. As seen, there are some years where the forecast is accurate and others where it is not. A more analytical way to know the accuracy of the prediction is by calculating the MAPE given via Equation (11). Table 10 shows the MAPE results calculated using the data shown in Table 7 and the reference values in Table 9.

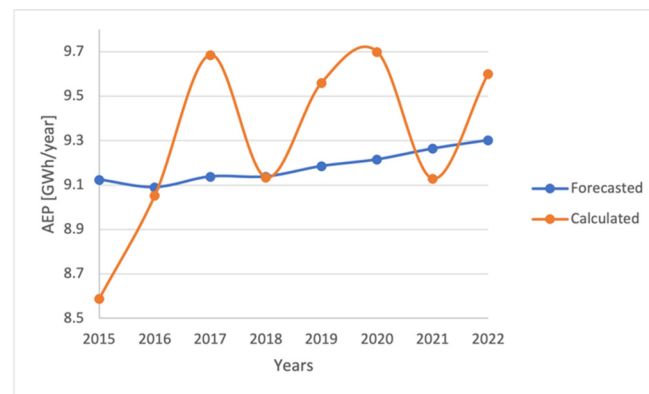


Figure 14. Forecasted annual energy production, $E_{WRTY} = AEP@P50$ and calculated annual energy production, E_{annual} .

Table 10. MAPE results for the forecasted E_{WRTY} using as reference values the results in Table 9 E_{annual} .

Year	MAPE [%] (Forecasted E_{annual})
2015	6.24
2016	0.42
2017	5.65
2018	0.05
2019	3.90
2020	5.00
2021	1.49
2022	3.10

MAPE results range between 6% for the worst case and 0.05% for the best case, with an average of 3.23%. According to Equation (5), this results in excellent accuracy since we are dealing with long-term forecasting [120].

Another way to estimate the AEP is by considering the statistical nature of the wind resource using the Weibull PDF, $E_{Weibull}$. The produced electrical energy is calculated through the use of Equation (9), using the wind turbine power curve Equation (13). The results are reported in Table 11, considering that the wind turbine operates 365 days, 24 h per day. As expected, these results are very close to those reported in Table 9, E_{annual} , and thus validate this statistical approach. The MAPE values have also been calculated, and these are shown in Table 12. As anticipated, the MAPE values are very small, with the highest value being 2.45% and the lowest value being 0.33%. However, it is important to emphasize that these calculations were carried out once the wind speed values at the site were known. Hence, they are an estimation of the energy produced.

Table 11. Estimated annual wind energy using the Weibull PDF and the average wind speed from the wind speed database.

Year	$E_{Weibull}$ [GWh]	V_{aver} [m/s]	E_{v_aver} [GWh]
2015	8.561	6.435	7.725
2016	9.114	6.637	8.300
2017	9.815	6.842	9.111
2018	9.172	6.618	8.430
2019	9.794	6.842	8.969
2020	9.617	6.826	8.923
2021	9.204	6.644	8.342
2022	9.563	6.764	8.838

Table 12. MAPE values for AEPs calculation using (a) Weibull PDF and (b) average wind speed.

Year	MAPE [%] ($E_{Weibull}$)	MAPE [%] ($E_{v_{aver}}$)
2015	0.33	10.06
2016	0.67	8.32
2017	1.33	5.93
2018	0.41	7.70
2019	2.45	6.18
2020	0.86	8.01
2021	0.81	8.62
2022	0.40	7.95

Finally, to compare the results reported in Table 7, an approach commonly used to estimate the produced energy at a given site was employed. This is a deterministic approach that leaves aside the intermittent nature of the wind resource, which considers an annual average wind speed, v_{aver} . Accordingly, this averaging is computed from the historical hourly wind speed data of the years under consideration, for example [121]. Nevertheless, wind speeds below and above the cut-in and cut-out speeds should not be contemplated. Moreover, for the computation, full wind turbine operation was considered necessary. Table 11 shows the average annual wind speed, v_{aver} , and the respective annual energy production, $E_{v_{aver}}$. It was assumed the wind turbine operates 365 days, 24 h per day.

The annual energies, $E_{v_{aver}}$, differ from E_{WRTY} . Indeed, to quantify the deviation from the actual data, the corresponding MAPEs were calculated, and these are shown in Table 12, taken as the reference values, E_{annual} , from Table 9. It is observed that the estimation of the annual energies using the annual average wind speed is less accurate than with the methodology proposed in this work, with MAPEs ranging from 5.93% to 10.06% and an average MAPE of 7.84%. Therefore, the forecasted E_{WRTY} values proposed in this work are better than those obtained with the commonly used methodology.

In the next subsection, a comparison of the estimated produced energy with that reported from the wind farm is performed.

3.2. Comparing Results

The goal of this work, as mentioned at the beginning of this section, is to compare these results with those reported in relation to the Energía Sierra Juárez wind farm. This farm is composed of 47 Vestas V112 3.3 MW wind turbines. Hence, the maximum theoretically generated energy of the farm is

$$E_{max} = 3.3 \text{ MW} \times 8760 \text{ h} \times 47 = 1,358,676 \text{ MWh.} \quad (17)$$

The reported energy can be found in the Energy Information System (SIE) of the Secretary of Energy of Mexico (SENER) [122]. For 2015 and 2016, the reported energy is 247,516 and 376,628 MWh, which derived from using Equation (10) with CFs of 0.18 and 0.27, respectively.

4. Discussion

This paper introduces the concept of wind statistical seasons, in contrast to the calendar seasons, with the intention of characterizing wind behavior with greater precision at the site of interest for a year ahead. Intuitively, a wind statistical season is composed of those months that exhibit a close statistical wind speed behavior through the years. That is, a statistical season includes all months with similar wind speed PDFs, and close statistics are determined through the use of Machine Learning data clustering analysis over the defining parameters of the 12 PDFs for a n -year database, which, for the Weibull wind speed distributions in this paper, are the shape, κ , and scale, λ , parameters. The clustering was carried out via the K-means method optimized with the Silhouette method. The annual wind behavior was characterized by using two statistical seasons for the case study in this

paper. One season comprehends the months of July, August, and September and is named HSS (High Statistical Season), while the other season includes the rest of the months and is called LSS (Low Statistical Season). This approach to finding out the annual seasonality of the wind speed produced intuitive and sound results that can be applied to other sites of interest.

The definition of meteorological typical years is common practice when characterizing the weather behavior throughout a whole year for a given location or land region based on the available weather data. For the aims and scope of this paper, we introduce the concept of the Wind Resource Typical Year (WRTY) to refer solely to the wind speed behavior from a statistical point of view based on monthly *PDFs*. The WRTY for the $n + 1$ year, composed of the wind speed *PDFs* of calendar months, was selected using the large number of monthly *PDFs* from previous n years of available wind speed data. With this aim, the 12 batches of monthly *PDFs* for the calendar months are calculated in batches throughout all years being considered. Then, the chosen monthly *PDF* for the WRTY was the closest to the batch monthly *PDF*, calculated using the MAE criteria. In this way, each month in the WRTY for the $n + 1$ year is represented with a *PDF*, that is, a *PDF* calculated from the accessible wind speed data from the previous years. Therefore, the WRTY gathers the most representative *PDFs*. A further approach could build the WRTY for the $n + 1$ year using the 12-batch monthly *PDFs* without going back to select the closet monthly *PDF*. In this case, it could be called the Wind Speed Statistical Year (WSSY), since it will contain the monthly statistical summaries of the wind speed data for all the years available.

This paper is mainly concerned with the long-term forecasting of wind power, where long-term refers to one year ahead. Ordinarily, the annual wind power is calculated using wind speed averaging; however, this method can produce inappropriate errors and should be considered only a rough estimate. On the other hand, wind power can be calculated directly from the wind speed data through the power curve of the wind turbine generators; the shorter the sampling period, the better the estimation of wind energy. The implemented MAPE criteria permit an appreciation of the magnitude of the wind power calculation error. Furthermore, the annual wind speed *PDF* allowed us to attain the wind power. The MAPE calculations enabled us to demonstrate that the forecasted AEPs calculated with the wind speed *PDFs* are more accurate than the AEPs calculated using wind speed averages, which proves the convenience of using this approach. Consequently, one-year-ahead forecasts using wind speed *PDFs* provide better results than the annual wind speed average. It was shown that the forecasting approach using the *PDFs* of the statistical seasons further improves the accuracy since the annual wind behavior is dissected with more detail.

The one-year-ahead forecasting of the wind speed or power is commonly required when carrying out operation management, dispatch planning, operation optimization, resource assessment, site selection, cost estimation, feasibility analysis, system expansion planning, bankable documentation, and financial investments. Most of the time, however, it is not enough to provide wind speed or power forecasts in these applications. Hence, the calculation of the Annual Energy Production (AEP) at different probabilities of exceedance (PoE) is required to make sound decisions. Since the proposed forecast methodology is carried out statistically and the results are mainly *PDFs*, the calculation of the AEP and PoE comes naturally. The forecast of the annual wind power obtained with the proposed methodology corresponds to the AEP with a 50% PoE or P50. The P75 indicator denotes an AEP with a 25% PoE. P50 and P75 are indicators of merit mainly used for project financing, whereas P90 and P95 are for equity investing. To gain further insight into the PoEs of wind power production forecasts, the contribution to AEP by each wind statistical season (HSS and LSS) were calculated from 2015 to 2022 for up to P95, assuming 11% uncertainty, as shown in Figures 11 and 12 for the HSS and LSS, respectively. These families of curves detail the best and worst performance per season, which can be used to time finance and investment along a calendar year. Similarly, the annual forecasts of the AEP for P50 through P95 were calculated and plotted for the same period, which can be advantageously used to estimate the annual maximum investments and minimum guaranteed profits.

5. Conclusions

Developing forecasting systems is easier or more difficult depending on the nature of the variable being forecast. Moreover, the longer the forecast horizon, the less accurate the prediction. In particular, this work deals with a big challenge, namely one-year-ahead wind power forecasting, due to three main reasons: (a) the wind resource is highly intermittent, and thus wind speed forecasting is difficult to achieve and highly depends on various conditions such as meteorological and geographical conditions; (b) the wind power depends on the wind speed cubed, which means that the accurate error in a forecasted wind speed translates into an accurate error cubed when forecasting wind power; and (c) one-year-ahead forecasting involves a huge forecast horizon and it is highly difficult to obtain predictions with good results. The main contributions of this work can be summarized as follows:

- This work presents a novel, intelligent, statistical methodology for long-term wind power forecasting.
- The forecast horizon can be in a seasonal or annual term.
- The concept of statistical seasonality is introduced and computed using a clustering analysis.
- By using the n -year wind speed database, the methodology forecasts the wind power generation in the $n + 1$ year.
- It can be applied to any region of the world since the data repository used contains data from any location, both onshore and offshore.
- It can be applied to any location under any operating conditions since it can be used with any wind speed probability distribution.
- It introduces the concept and the construction procedure of the Wind Resource Typical Year to characterize the wind resource at the location analogously to the Typical Meteorological Year that is used to characterize the meteorological conditions of a site.
- The results for the forecasted annual wind energy beat the ones obtained from the traditional and most used deterministic method using the average annual wind speed.
- This is a simple yet powerful method that, for this case study, provided forecasted annual wind energies with MAPE values, which can be as high as almost 7% and as low as less than 1%, which are excellent when compared with those obtained from the traditional method that range from 10% to 6%.
- This methodology also applies to small-scale wind turbines since the data repository also considers the wind speed data at 2 and 10 m above ground level.

Nevertheless, there are some drawbacks concerning the introduced methodology, such as the following:

- The cluster analysis is not 100% reproducible since the results depend on the initial conditions of the position of the centroids, even though the results from different simulations do not heavily differ.
- Even though this method extends to any *PDF*, it may be more complex when dealing with *PDFs* with more than two parameters.
- It has a low spatial resolution of around 0.5° in latitude and longitude due to the available data from the data repository, which corresponds approximately to a 50 km squared area.
- One could use another wind speed data repository, for instance the National Solar Radiation Database, but there is no information on the height of the wind speed data.

In future work, the proposed methodology can be explicitly extended to other *PDFs*. For instance, the Rayleigh *PDF* is an interesting case since it only has one parameter. Therefore, a site whose wind resources are characterized by this *PDF* can be used as a case study. Additionally, the Gamma *PDF* is very frequently used *PDF* with two parameters, and, as mentioned in Table 3, it is an alternative to the Weibull *PDF*. Then, other *PDFs*, such as the three-parameter Gamma, can be explored to enquire about the complexity of the calculation, including the four-parameter Generalized Gamma, which is suitable when

characterizing the wind resources in some parts of Europe. Finally, for future, long-term work, an intelligent system incorporating any *PDF* can be developed for use anywhere, and a patent could be obtained. Finally, yet importantly, it is relevant to increase the spatial resolution, and thus, another direction for future work is the search for a reliable database with higher resolution or the development of a method to increase the resolution using other means.

Supplementary Materials: The following supporting information can be downloaded at <https://github.com/FelosRG/Wind-Stat-Forecast> (accessed on 29 November 2023).

Author Contributions: Conceptualization, M.B. and R.G.; methodology, M.B. and R.G.; software, A.R.; validation, M.B. and A.R.; formal analysis, M.B.; investigation, M.B.; resources, M.B., A.R. and C.G.-B.; data curation, A.R.; writing—original draft preparation, M.B.; writing—review and editing, M.B., R.G. and R.M.; visualization, M.B. and R.G.; supervision, M.B.; project administration, M.B.; funding acquisition, C.G.-B. and R.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The meteorological datasets are freely available at the NSRDB, (<https://nsrdb.nrel.gov/>, accessed on 15 April 2023).

Acknowledgments: M.B. thanks CONACYT for her Catedra Research Position with ID 71557 and CENIDET for its hospitality and support.

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

References

- Intergovernmental Panel on Climate Change. Renewable Energy Sources and Climate Change Mitigation, Summary for Policymakers and Technical Summary. Available online: https://www.ipcc.ch/site/assets/uploads/2018/03/SRREN_FD_SPM_final-1.pdf (accessed on 24 December 2019).
- International Renewable Energy Agency. Renewable Capacity Statistics 2019. Available online: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2019/Mar/IRENA_RE_Capacity_Statistics_2019.pdf (accessed on 24 December 2019).
- Global Wind Energy Council. Global Wind Report 2022. Available online: <https://gwec.net/wp-content/uploads/2022/03/GWEC-GLOBAL-WIND-REPORT-2022.pdf> (accessed on 11 September 2023).
- International Energy Agency. Renewable Energy Market Update Outlook for 2023 and 2024. Available online: <https://build-up.ec.europa.eu/en/resources-and-tools/publications/iea-renewable-energy-market-update-outlook-2023-and-2024-published> (accessed on 11 September 2023).
- Manwell, J.F.; McGowan, J.G.; Rogers, A.L. *Wind Energy Explained, Theory, Design and Application*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2010.
- Wan, Y.H. *Long-Term Wind Power Variability*; Technical Report NREL/TP-5500-53637; National Renewable Energy Laboratory: Golden, CO, USA, 2012.
- Goater, A.; Intermittent Electricity Generation. Parliamentary Office of Science and Technology. Available online: <https://researchbriefings.files.parliament.uk/documents/POST-PN-464/POST-PN-464.pdf> (accessed on 11 September 2023).
- Jain, P.; Wijayatunga, P. Grid Integration of Wind Power: Best Practices for Emerging Wind Markets. Asian Development Bank: Mandaluyong. *Philippines* **2016**, *43*, 2–36.
- Denholm, P.; Mai, T.; Kenyon, R.W.; Kroposki, B.; O'Malley, M. *Inertia and the Power Grid: A Guide without the Spin*; Technical Report NREL/TP-6A20-73856; National Renewable Energy Laboratory: Golden, CO, USA, 2020.
- Gowrisankaran, G.; Reynolds, S.S.; Samano, M. Intermittency and the value of renewable energy. *J. Politi-Econ.* **2016**, *124*, 1187–1234. [[CrossRef](#)]
- Bandyopadhyay, R.; Ferrero, V.; Tan, X. *Coordinated Operations of Flexible Coal and Renewable Energy Power Plants: Challenges and Opportunities*; UNECE Energy Series 2017 No. 52; Economic Commission for Europe: Geneva, Switzerland, 2017.
- Ye, H.; Yang, B.; Han, Y.; Li, Q.; Deng, J.; Tian, S. Wind Speed and Power Prediction Approaches: Classifications, Methodologies, and Comments. *Front. Energy Res.* **2022**, *10*, 901767. [[CrossRef](#)]
- Soman, S.S.; Zareipour, H.; Malik, O.; Mandal, P. A review of wind power and wind speed forecasting methods with different timehorizons. In Proceedings of the North-American Power Symposium (NAPS) 2010, Arlington, TX, USA, 26–28 September 2010; pp. 1–8. [[CrossRef](#)]
- Wang, J.; Song, Y.; Liu, F.; Hou, R. Analysis and application of forecasting models in wind power integration: A review of multi-step-ahead wind speed forecasting models. *Renew. Sustain. Energy Rev.* **2016**, *60*, 960–981. [[CrossRef](#)]
- Lerner, J.; Grundmeyer, M.; Garvert, M. The importance of wind forecasting. *Renew. Energy Focus* **2009**, *10*, 64–66. [[CrossRef](#)]

16. Zheng, Z.W.; Chen, Y.Y.; Huo, M.M.; Zhao, B. An Overview: The Development of Prediction Technology of Wind and Photovoltaic Power Generation. *Energy Procedia* **2011**, *12*, 601–608. [[CrossRef](#)]
17. Iseh, A.J.; Woma, T.Y. Weather forecasting models, methods and applications. *Int. J. Eng. Res. Technol.* **2013**, *2*, 1945–1957.
18. Azad, H.B.; Mekhilef, S.; Ganapathy, V.G. Long-Term Wind Speed Forecasting and General Pattern Recognition Using Neural Networks. *IEEE Trans. Sustain. Energy* **2014**, *5*, 546–553. [[CrossRef](#)]
19. Hamilton, J. Chapter 13, “The Kalman Filter”. In *Time Series Analysis*; Princeton University Press: Princeton, NJ, USA, 1994.
20. Louka, P.; Galanis, G.; Siebert, N.; Kariniotakis, G.; Katsafados, P.; Pytharoulis, I.; Kallos, G. Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. *J. Wind. Eng. Ind. Aerodyn.* **2008**, *96*, 2348–2362. [[CrossRef](#)]
21. Langley, P. The Changing Science of Machine Learning. *Mach. Learn.* **2011**, *82*, 275–279. [[CrossRef](#)]
22. Shouman, E.R. *Wind Power Forecasting Models*; IntechOpen eBooks: London, UK, 2022.
23. Karaman, O.A. Prediction of wind power with machine learning models. *Appl. Sci.* **2023**, *13*, 11455. [[CrossRef](#)]
24. Dupré, A.; Drobinski, P.; Alonzo, B.; Badosa, J.; Briard, C.; Plougonven, R. Sub-hourly forecasting of wind speed and wind energy. *Renew. Energy* **2020**, *145*, 2373–2379. [[CrossRef](#)]
25. Du, P. Ensemble machine learning-based wind forecasting to combine NWP output with data from weather station. *IEEE Trans. Sustain. Energy* **2019**, *10*, 2133–2141. [[CrossRef](#)]
26. Yang, Y.; Zhou, H.; Wu, J.; Ding, A.; Wang, Y.G. Robustified extreme learning machine regression with applications in outlier-blended wind speed forecasting. *Appl. Soft Comput.* **2022**, *122*, 108814. [[CrossRef](#)]
27. Yang, Y.; Zhou, H.; Gao, Y.; Wu, J.; Wang, Y.G.; Fu, L. Robust penalized extreme learning machine regression with applications in wind speed forecasting. *Neural Comput. Applic.* **2022**, *34*, 391–407. [[CrossRef](#)]
28. Petković, D.; Shamshirband, S.; Anuar, N.B.; Saboohi, H.; Wahab, A.W.A.; Protić, M.; Zalnezhad, E.; Mirhashemi, S.M.A. An appraisal of wind speed distribution prediction by soft computing methodologies: A comparative study. *Energy Convers. Manag.* **2014**, *84*, 133–139. [[CrossRef](#)]
29. Cui, M.; Zhang, J.; Wang, Q.; Krishnan, V.; Hodge, B.M. A data-driven methodology for probabilistic wind power ramp forecasting. *IEEE Trans. Smart Grid* **2019**, *10*, 1326–1338. [[CrossRef](#)]
30. Al-Yahyai, S.; Charabi, Y.; Gastil, A. Review of the use of Numerical Weather Prediction (NWP) Models for wind energy assessment. *Renew. Sustain. Energy* **2010**, *14*, 3192–3198. [[CrossRef](#)]
31. Foley, A.M.; Leahy, P.; McKeogh, E. Wind power forecasting & prediction methods. In Proceedings of the 2010 9th International Conference on Environment and Electrical Engineering, Prague, Czech Republic, 16–19 May 2010.
32. Foley, A.M.; Leahy, P.G.; Marvuglia, A.; McKeogh, E.J. Current methods and advances in forecasting of wind power generation. *Renew. Energy* **2012**, *37*, 1–8. [[CrossRef](#)]
33. Wang, X.; Guo, P.; Huang, X. A Review of wind power forecasting models. *Energy Procedia* **2011**, *12*, 770–778. [[CrossRef](#)]
34. Chandra, D.R.; Kumari, M.S.; Sydulu, M. A detailed literature review on wind forecasting. In Proceedings of the International Conference on Power, Energy and Control (ICPEC), Dindigul, India, 6–8 February 2013; pp. 630–634.
35. Aggarwal, S.K.; Gupta, M. Wind power forecasting: A review of statistical models. *Int. J. Energy Sci.* **2013**, *3*, 1.
36. Chang, W.-Y. A Literature review of wind forecasting methods. *J. Power Energy Eng.* **2014**, *2*, 161–168. [[CrossRef](#)]
37. Saroha, S.; Aggarwal, S.K. A review and evaluation of current wind power prediction technologies. *WSEAS Trans. Power Syst.* **2015**, *10*, 1–12.
38. Ren, Y.; Suganthan, P.N.; Srikanth, N. Ensemble methods for wind and solar power forecasting—A state of the art review *Renew. Sustain. Energy* **2015**, *50*, 82–91. [[CrossRef](#)]
39. Varanasi, J.; Tripathi, M.M. A comparative study of wind power forecasting techniques—A review article. In Proceedings of the 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 16–18 March 2016.
40. Giebel, G.; Kariniotakis, G. Wind power forecasting—A review of the state of the art. In *Woodhead Publishing Series in Energy, Renewable Energy Forecasting*; Woodhead Publishing: Cambridge, UK, 2017; pp. 59–109.
41. Liu, H.; Chen, C.; Lv, X.; Wu, X.; Liu, M. Deterministic wind energy forecasting: A review of intelligent predictors and auxiliary methods. *Energy Convers. Manag.* **2019**, *195*, 328–345. [[CrossRef](#)]
42. Hanifi, S.; Liu, X.; Lin, Z.; Lotfian, S. A critical review of wind power forecasting methods—Past, present and future. *Energies* **2020**, *13*, 3764. [[CrossRef](#)]
43. Dhiman, H.S.; Dipankar, D. A review of wind speed and wind power forecasting techniques. *arXiv* **2020**, arXiv:2009.02279v1.
44. Jorgensen, K.L.; Shaker, H.R. Wind power forecasting using machine learning: State of the art, trends and challenges. In Proceedings of the 2020 IEEE 8th International Conference on Smart Energy Grid Engineering (SEGE), Oshawa, ON, Canada, 12–14 August 2020; pp. 44–50.
45. Wang, Y.; Zou, R.; Liu, F.; Zhang, L.; Liu, Q. A review of wind speed and wind power forecasting with deep neural networks. *Appl. Energy* **2021**, *304*, 117766. [[CrossRef](#)]
46. Saroha, S.; Rana, P. Wind power forecasting. In *Forecasting in Mathematics-Recent Advances, New Perspectives and Applications*; IntechOpen: London, UK, 2021.
47. Zhao, E.; Sun, S.; Wang, S. New developments in wind energy forecasting with artificial intelligence and big data: A scientific insight. *Data Sci. Manag.* **2022**, *5*, 84–95. [[CrossRef](#)]

48. Valdivia-Bautista, S.M.; Domínguez-Navarro, J.A.; Pérez-Cisneros, M.; Vega-Gómez, C.J.; Castillo-Téllez, B. Artificial Intelligence in Wind Speed Forecasting: A Review. *Energies* **2023**, *16*, 2457. [\[CrossRef\]](#)
49. Liu, Y.; Wang, Y.; Wang, Q.; Zhang, K.; Qiang, W.; Wen, Q.H. Recent advances in data-driven prediction for wind power. *Front. Energy Res.* **2023**, *11*, 1204343. [\[CrossRef\]](#)
50. Tsai, W.-C.; Hong, C.-M.; Tu, C.-S.; Lin, W.-M.; Chen, C.-H. A Review of modern wind power generation forecasting technologies. *Sustainability* **2023**, *15*, 10757. [\[CrossRef\]](#)
51. Shobana Devi, A.; Maragatham, G.; Boopathi, K.; Lavanya, M.C.; Saranya, R. Long-Term Wind Speed Forecasting—A Review. In *Artificial Intelligence Techniques for Advanced Computing Applications*; Springer: Singapore, 2020; pp. 79–99.
52. Wu, Y.-K.; Hong, J.-S. A literature review of wind forecasting technology in the world. In Proceedings of the 2007 IEEE Lausanne Power Tech, Lausanne, Switzerland, 1–5 July 2007; pp. 504–509. [\[CrossRef\]](#)
53. Fan, S.; Liao, J.R.; Yokoyama, R.; Chen, L.; Lee, W.-J. Forecasting the Wind Generation Using a Two-Stage Network Based on Meteorological Information. *IEEE Trans. Energy Convers.* **2009**, *24*, 474–482. [\[CrossRef\]](#)
54. Erdem, E.; Shi, J. ARMA based approaches for forecasting the tuple of wind speed and direction. *Appl. Energy* **2011**, *88*, 1405–1414. [\[CrossRef\]](#)
55. Guo, Z.-H.; Wu, J.; Lu, H.-Y.; Wang, J.-Z. A Case Study on a Hybrid Wind Speed Forecasting Method Using BP Neural Network. *Knowl. Based Syst.* **2011**, *24*, 1048–1056. [\[CrossRef\]](#)
56. Troccoli, A.; Muller, K.; Coppin, P.; Davy, R.; Russell, C.; Hirsch, A.L. Long-term wind speed trends over Australia. *J. Clim.* **2012**, *25*, 170–183. [\[CrossRef\]](#)
57. Doblas-Reyes, F.J.; García-Serrano, J.; Lienert, F.; Biescas, A.P.; Rodrigues, L.R.L. Seasonal Climate Predictability and Forecasting: Status and Prospects. *Clim. Chang.* **2013**, *4*, 245–268.
58. Azimi, R.; Ghofrani, M.; Ghayekhloo, M. A hybrid wind power forecasting model based on data mining and wavelets analysis. *Energy Convers. Manag.* **2016**, *127*, 208–225. [\[CrossRef\]](#)
59. Grigonyte, E.; Butkeviciute, E. Short-term wind speed forecasting using ARIMA model. *Energetika* **2016**, *62*, 45–55. [\[CrossRef\]](#)
60. Yatiyana, E.; Rajakaruna, S.; Ghosh, A. Wind Speed and Direction Forecasting for Wind Power Generation Using ARIMA Model. In Proceedings of the Australasian Universities Power Engineering Conference (AUPEC), Melbourne, Australia, 19–22 November 2017; pp. 1–6.
61. Lledó, L.; Bellprat, O.; Doblas-Reyes, F.J.; Soret, A. Investigating the Effects of Pacific Sea Surface Temperatures on the Wind Drought of 2015 Over the United States. *J. Geophys. Res. Atmos.* **2018**, *123*, 4837–4849. [\[CrossRef\]](#)
62. Lledó, L.; Torralba, V.; Soret, A.; Ramon, J.; Doblas-Reyes, F. Seasonal Forecasts of Wind Power Generation. *Renew. Energy* **2019**, *143*, 91–100. [\[CrossRef\]](#)
63. Molteni, F.; Stockdale, T.; Alonso-Balmaseda, M.; Balsamo, G.; Buizza, R.; Ferranti, L.; Magnusson, L.; Mogensen, K.; Palmer, T.; Vitart, F. *The New ECMWF Seasonal Forecast System (System 4)*; Technical Report 656; ECMWF: Reading, UK, 2011.
64. Mariotti, A.; Ruti, P.M.; Rixen, M. Progress in subseasonal to seasonal prediction through a joint weather and climate community effort. *NPJ Clim. Atmos. Sci.* **2018**, *1*, 4. [\[CrossRef\]](#)
65. Tyass, I.; Bellat, A.; Raihani, A.; Mansouri, K.; Khalili, T. Wind Speed Prediction Based on Seasonal ARIMA model. *E3S Web Conf.* **2022**, *336*, 00034. [\[CrossRef\]](#)
66. Tawn, R.; Browell, J.; McMillan, D. Subseasonal-to-seasonal forecasting for wind turbine maintenance scheduling. *Wind* **2022**, *2*, 260–287. [\[CrossRef\]](#)
67. Sulagna, M.; Harsh, P.; Shekher, V.; Rai, P. A statistical analysis model of wind power generation forecasting for the Western Region of India. *TechRxiv* **2023**. [\[CrossRef\]](#)
68. Mesa-Jiménez, J.; Tzianoumis, A.; Stokes, L.; Yang, Q.; Livina, V. Long-term wind and solar energy generation forecasts, and optimisation of Power Purchase Agreements. *Energy Rep.* **2023**, *9*, 292–302. [\[CrossRef\]](#)
69. Magaña-González, R.; Rodríguez-Hernández, O.; Canul-Reyes, D. Analysis of seasonal variability and complementarity of wind and solar resources in Mexico. *Sustain. Energy Technol. Assess.* **2023**, *60*, 103456. [\[CrossRef\]](#)
70. Dayton, G.H. *Seasonality. Encyclopedia of Ecology*; Academic Press: Cambridge, MA, USA, 2008; pp. 3168–3171.
71. Khavrus, V.; Shelevytsky, I. Geometry and the physics of seasons. *Phys. Educ.* **2012**, *47*, 680–692. [\[CrossRef\]](#)
72. Emmert-Streib, F.; Moutari, S.; Dehmer, M. *Clustering. Elements of Data Science, Machine Learning, and Artificial Intelligence Using R*; Springer: Berlin/Heidelberg, Germany, 2023.
73. Bruhn, J.A.; Fry, W.E.; Fick, G.W. Simulation of daily weather data using theoretical probability distributions. *J. Appl. Meteorol.* **1980**, *19*, 1029–1036. [\[CrossRef\]](#)
74. Dubrovský, M. Creating daily weather series with use of the weather generator. *Environmetrics* **1997**, *8*, 409–424. [\[CrossRef\]](#)
75. Wilks, D.S.; Wilby, R.L. The weather generation game: A review of stochastic weather models. *Prog. Phys. Geogr. Earth Environ.* **1999**, *23*, 329–357. [\[CrossRef\]](#)
76. Weibull, W. A statistical distribution function of wide applicability. *J. Appl. Mech.* **1951**, *18*, 293–297. [\[CrossRef\]](#)
77. Carta, J.; Ramírez, P.; Velázquez, 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*, 933–955. [\[CrossRef\]](#)
78. Shi, H.; Dong, A.; Xiao, N.; Huan, Q. Wind speed distributions used in wind energy assessment: A review. *Front. Energy Res.* **2021**, *9*, 769920. [\[CrossRef\]](#)

79. Shokrzadeh, S.; Jozani, M.J.; Bibeau, E. Wind turbine power curve modeling using advanced parametric and nonparametric methods. *IEEE Trans. Sustain. Energy* **2014**, *5*, 1262–1269. [[CrossRef](#)]
80. Qin, Z.; Li, W.; Xiong, X. Estimating wind speed probability distribution using kernel density method. *Electr. Power Syst. Res.* **2011**, *81*, 2139–2146. [[CrossRef](#)]
81. Montoya, J.A.; Díaz-Francés, E.; Figueroa, G. Estimation of the reliability parameter for three-parameter Weibull models. *Appl. Math. Model.* **2018**, *67*, 621–633. [[CrossRef](#)]
82. Saleh, H.; Abou El-Azm Aly, A.; Abdel-Hady, S. Assessment of different methods used to estimate Weibull distribution parameters for wind speed in Zafarana wind farms, Suez Gulf, Egypt. *Energy* **2012**, *44*, 710–719. [[CrossRef](#)]
83. Torrielli, A.; Repetto, M.P.; Solari, G. Extreme wind speeds from long-term synthetic records. *J. Wind. Eng. Ind. Aerodyn.* **2013**, *115*, 22–38. [[CrossRef](#)]
84. Akgül, F.G.; Şenoğlu, B.; Arslan, T. An alternative distribution to Weibull for modeling the wind speed data: Inverse Weibull distribution. *Energy Convers. Manag.* **2016**, *114*, 234–240. [[CrossRef](#)]
85. Sarkar, A.; Deep, S.; Datta, D.; Vijaywargiya, A.; Roy, R.; Phanikanth, V.S. Weibull and generalized extreme value distributions for wind speed data analysis of some locations in India. *KSCE J. Civ. Eng.* **2019**, *23*, 3476–3492. [[CrossRef](#)]
86. Aries, N.; Boudia, S.M.; Ounis, H. Deep assessment of wind speed distribution models: A case study of four sites in Algeria. *Energy Convers. Manag.* **2018**, *155*, 78–90. [[CrossRef](#)]
87. Guedes, K.S.; de Andrade, C.F.; Rocha, P.A.C.; Mangueria, R.D.S.; Moura, E.P. Performance analysis of metaheuristic optimization algorithms in estimating the parameters of several wind speed distributions. *Appl. Energy* **2020**, *268*, 114952. [[CrossRef](#)]
88. Alavi, O.; Sedaghat, A.; Mostafaeipour, A. Sensitivity analysis of different wind speed distribution models with actual and truncated wind data: A case study for Kerman, Iran. *Energy Convers. Manag.* **2016**, *120*, 51–61. [[CrossRef](#)]
89. Brano, V.L.; Orioli, A.; Ciulla, G.; Culotta, S. Quality of wind speed fitting distributions for the urban area of Palermo, Italy. *Renew. Energy* **2011**, *36*, 1026–1039. [[CrossRef](#)]
90. Soukissian, T. Use of multi-parameter distributions for offshore wind speed modeling: The Johnson SB distribution. *Appl. Energy* **2013**, *111*, 982–1000. [[CrossRef](#)]
91. Jung, C.; Schindler, D. Global comparison of the goodness-of-fit of wind speed distributions. *Energy Convers. Manag.* **2017**, *133*, 216–234. [[CrossRef](#)]
92. Wais, P. A review of Weibull functions in wind sector. *Renew. Sustain. Energy Rev.* **2017**, *70*, 1099–1107. [[CrossRef](#)]
93. Rinne, H. *The Weibull Distribution, a Handbook*; CRC Press: Boca Raton, FL, USA, 2009.
94. Sumair, M.; Aized, T.; Gardezi, S.A.; Rehman, S.U.U.; Rehman, S.M.S. Wind potential estimation and proposed energy production in Southern Punjab using Weibull probability density function and surface measured data. *Energy Explor. Exploit.* **2021**, *39*, 2150–2168. [[CrossRef](#)]
95. Unnikrishna, P.S.; Papoulis, A.P. *Probability, Random Variables and Stochastic Processes*; McGraw-Hill: Boston, MA, USA, 2002.
96. Jaramillo, O.A.; Borja, M.A. Bimodal versus weibull wind speed distributions: An analysis of wind energy potential in La Venta, Mexico. *Wind Eng.* **2004**, *28*, 225–234. [[CrossRef](#)]
97. Pelleg, D.; Moore, A. Accelerating exact k-means algorithms with geometric reasoning. In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD'99, California, CA, USA, 15–18 August 1999.
98. Anil, K.J. Data clustering: 50 years beyond k-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666.
99. Qiu, W.; Joe, H. Generation of random clusters with specified degree of separation. *J. Classif.* **2006**, *23*, 315–334. [[CrossRef](#)]
100. Qiu, W.; Joe, H. clusterGeneration: Random Cluster Generation (with specified degree of separation). *R Package* **2009**, *1*, 75275-0122.
101. Azhar, A.; Hashim, H. A review of wind clustering methods based on the wind speed and trend in Malaysia. *Energies* **2023**, *16*, 3388. [[CrossRef](#)]
102. Rousseeuw, P.J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [[CrossRef](#)]
103. Vignola, F.E.; McMahan, A.C.; Grover, C.N. Chapter 5—Bankable Solar Radiation Datasets. In *Solar Energy Forecasting and Re-source Assessment*; Academic Press: Cambridge, MA, USA, 2013; pp. 97–131.
104. Hall, I.; Prairie, R.; Anderson, H.; Boes, E. *Generation of Typical Meteorological Years from 26 SOLMET Stations*; Technical Report SAND78-1601; Sandia National Laboratories: Albuquerque, NM, USA, 1978.
105. Marion, W.; Urban, K. *Users Manual for TMY2s—Typical Meteorological Years Derived from the 1961–1990 National Solar Radiation Data Base*; Technical Report NREL/TP-463-7668; National Renewable Energy Laboratory: Golden, CO, USA, 1995.
106. Wilcox, W. *Marion. User's Manual for TMY3 Data Sets*; Technical Report NREL/TP-581-43156; National Renewable Energy Laboratory: Golden, CO, USA, 2008.
107. Available online: <http://www.trnsys.com> (accessed on 27 May 2023).
108. Available online: <https://www.pvsyst.com/> (accessed on 27 May 2023).
109. Available online: <https://energyplus.net> (accessed on 27 May 2023).
110. Available online: <https://climate.onebuilding.org/> (accessed on 27 May 2023).
111. Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. [[CrossRef](#)]

112. Available online: https://www.nhsec.nh.gov/projects/2013-02/documents/131212appendix_15.pdf (accessed on 20 September 2023).
113. Rau, V.G.; Jangamshetti, S.H. Normalized power curves as a tool for identification of optimum wind turbine generator parameters. *IEEE Trans. Energy Convers.* **2001**, *16*, 283–288.
114. El-Sharkawi, M.A. *Wind Energy, an Introduction*; CRC Press: Boca Raton, FL, USA, 2015.
115. St-Aubin, P.; Agard, B. Precision and Reliability of Forecasts Performance Metrics. *Forecasting* **2022**, *4*, 882–903. [[CrossRef](#)]
116. Available online: <https://www.energiasj.com/> (accessed on 17 July 2023).
117. National Aeronautics and Space Administration, NASA. Available online: <https://power.larc.nasa.gov/data-access-viewer/> (accessed on 13 October 2023).
118. Available online: <https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/> (accessed on 13 October 2023).
119. Klug, H. *What Does Exceedance Probabilities P90, P75, P50 Mean?* DEWI Magazin: Jakarta, Indonesia, 2006; p. 28.
120. Borunda, M.; Rodriguez, K.; Garduno, R.; De la Cruz, J.; Antunez-Estrada, J.; Jaramillo, O.A. Long-term Estimation of Wind Power by Probabilistic Forecast Using Genetic Programming. *Energies* **2020**, *13*, 1885. [[CrossRef](#)]
121. Klyuev, R.; Bosikov, I.; Gavrina, O. Use of wind power stations for energy supply to consumers in mountain territories. In Proceedings of the International Ural Conference on Electrical Power Engineering (UralCon), Chelyabinsk, Russia, 1–3 October 2019; pp. 116–121.
122. Available online: <https://sie.energia.gob.mx/> (accessed on 8 June 2023).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.