

Article

Comparison of Optimum Spline-Based Probability Density Functions to Parametric Distributions for the Wind Speed Data in Terms of Annual Energy Production

Munir Ali Elfarra * and Mustafa Kaya

Department of Aeronautical Engineering, Ankara Yildirim Beyazit University, Ankara 06050, Turkey; mukaya@ybu.edu.tr

* Correspondence: melfarra@ybu.edu.tr; Tel.: +90-541-205-1192

Received: 26 October 2018; Accepted: 14 November 2018; Published: 17 November 2018



Abstract: The common approach to wind energy feasibility studies is to use Weibull distribution for wind speed data to estimate the annual energy production (AEP). However, if the wind speed data has more than one mode in the probability density, the conventional distributions including Weibull fail to fit the wind speed data. This highly affects the technical and economic assessment of a wind energy project by causing crucial errors. This paper presents a novel way to define the probability density for wind speed data using splines. The splines are determined as a solution of constrained optimization problems. The constraints are the characteristics of probability density functions. The proposed method is implemented for different wind speed distributions including multimodal data and compared with Weibull, Weibull and Weibull and Beta Exponentiated Power Lindley (BEPL) distributions. It is also compared with two other nonparametric distributions. The results show that the spline-based probability density functions produce a minimum fitting error for all the analyzed cases. The AEP calculated based on this method is considered to have high fidelity, which will decrease the investment risk.

Keywords: nonparametric probability density function; Weibull distribution; Weibull and Weibull distribution; spline-based probability density function; annual energy production; multimodal wind speed data

1. Introduction

Population and industry increase in the world which means that the demand for energy increases too. Fossil-fuel based energy sources have negative effects on the environment and they are running out. The greenhouse gases, such as carbon dioxide (CO_2), Methane (CH_4) and Nitrous Oxide (N_2O), are the main factors of increasing global warming [1,2]. Those gases arise from burning fossil fuels. Alternative renewable sources must be developed to decrease the emission of greenhouse gases. Wind energy is one of the renewable energy sources which is rapidly growing in most of the countries due to its relatively low cost [3].

Before the installation of a wind farm, detailed analyses of the wind energy potential of the site including the estimation of the annual energy production (AEP), capacity factor calculation and cost analysis should be performed to minimize the investment risk [4]. The available wind power in a certain site is evaluated using wind speed data. Finding accurate and reliable methods to represent the wind speed data is a crucial issue in wind energy assessment. Most of the methods used in the literature are based on the probability distribution functions (PDF). Due to its simplicity, Weibull function [5] is the most commonly used method to fit the wind speed data [6–15]. The Weibull based



However, in many regions, when the seasonal, climatic, regional and diurnal effects are considered, the frequency distribution of wind data behaves as a bimodal or even sometimes a multimodal distribution. Examples for bimodal wind data distribution include; the observed wind data in Lesvos and Mykonos sites in Greece [16], the wind data distribution in La Ventosa, Mexico [17], the wind data in Santa Elena, Colombia [18] and the data in Butler Grade site in USA [19].

The conventional two-parameter Weibull based methods are not sufficient to fit bimodal or multimodal wind data and, if used, will lead to incorrect results regarding the calculations of AEP, capacity factor and cost analysis.

There are many trials to use other methods to overcome the fitting problem for bimodal and multimodal data. Akdağ et al. [16], have shown that the annual energy production estimation for wind farms to be installed in the Eastern Mediterranean must be calculated using two-component mixture WW-PDF. They have stated that using a typical two-parameter Weibull method will underestimate the AEP.

Jaramillo and Borja [17] have used a bimodal probability density function (PDF) based on a Weibull and Weibull distribution to estimate a bimodal wind speed frequency distribution. They have mixed two component distributions and the method is called two-component mixture WW-PDF. Their results have shown better, but not very sufficient, fitting compared to the conventional Weibull method.

Yürüşen and Melero [20] have developed a procedure to improve the fitting problem for bimodal and multimodal wind data. First, the wind data are inspected to determine the distribution type. They carried out the inspection using the Silverman test [21]. If the distribution is unimodal, then the conventional two-parameter Weibull PDF is implemented. If the data have a multimodal behavior, then (based on whether the origin of modality is known or not) either several unimodal Weibull PDF's are used or Beta Exponentiated Power Lindley [22] with Weibull mixture distributions are performed. This method gives better fitting compared to the classical Weibull method. However, information about the modality type and origin of the modality are necessary to proceed with the method.

Using splines [23,24] or polynoms [25] is a nonparametric approach to model the probability density functions. Morrissey and Greene [26] have examined different expansions of polynomials to approximate the wind speed PDF's which have poor fitting problems. Their results have shown that proper selection of orthogonal polynomials can give excellent fit much better than the fit of the Weibull method.

Wijnands et al. [27] have developed a spline-based mathematical model to describe the temporal-spatial wind speeds near-surface distribution in a Tropical Cyclone (TC). Their spline model is able to capture the asymmetries of wind profile and can be used successfully to simulate the TC wind speeds over waters.

In the present study, a nonparametric probability density function model is proposed to fit the observed wind data distribution. For this purpose, splines based on cubic polynomials are used. The nodes of the splines are determined by solving an optimization problem. The constraints of the optimization problem are the general characteristics (explained in the next section) of a typical probability density function for wind speed data. Moreover, the mean and the standard deviation values are also considered as constraints. The results are compared to those of some parametric (Weibull, Weibull and BEPL distributions) and nonparametric (Empirical and 3rd order polynomial based kernel distributions) probability functions.

2. Methodology

2.1. Spline-Based Probability Density Function

The wind speed data is originated from meteorological observations. It provides the frequency of the observed wind speed along a certain time, which is usually one year. A typical wind speed data is shown in Figure 1 as a histogram plot. This frequency data of the wind speed gives natural information about the probability density distribution of the wind speed and shown by $OD(w_j)$, j = 1, 2, ..., M where M is the number of the distinct values of the observed wind speed. A common approach is to fit this data using a parametric probability density function such as Weibull or Rayleigh distributions.



Figure 1. An example for the wind speed data in histogram format.

In this study, the wind speed data is fitted using splines to construct piecewise polynomial approximations on a number of intervals out of the observed wind speed range. In statistics, a probability density function obtained using such an approach is known as nonparametric distribution [28].

The procedure of implementing the proposed method is as follow. The range of the wind speed from 0 m/s (w_0) to the maximum observed wind speed (w_M) is first evenly divided into N intervals that lead to N polynoms which pass through N + 1 nodes. It should be noted that N is an input by the user and its value is less than M, which means that the spline to be fitted is not required to pass through all the observed wind speed data. This condition is needed to satisfy some constraints on the probability density function such as providing the unity probability for the whole range. Also, the fitted probability density function will be smooth similar to the conventional distributions. In the present paper, cubic polynomials are used for constructing the spline. The piecewise cubic polynomial $f_i(w)$, between each successive pair of nodes has the following form:

$$f_i(w) = a_i w^3 + b_i w^2 + c_i w + d_i, \ i = 1, 2, \dots, N$$
⁽¹⁾

 $f_i(w)$ is defined between nodes i - 1 and i. The spline, f(w), is the union of all $f_i(w)$ and is C^2 continuous through each node. f(w) is zero when $w < w_0$ and $w > w_N$. The unknown values of the function, $f_i(w)$, at each node and its first derivative values in the first and last nodes are determined

together with the spline coefficients, a_i , b_i , c_i and d_i by solving the constrained optimization problem given below:

Minimize
$$f(w_i)$$
, $f'(w_0)$, $f'(w_M)$
 $RMSE = \sqrt{\frac{1}{M} \sum_{j=1}^{M} [OD(w_j) - f(w_j)]^2}$
subject to
Constraint 1: $\int_{w_0}^{w_M} f(w) dw = 1$ (2)
Constraint 2: $f(w) \ge 0$
Constraint 3: $f(w) \le f_{max}$
Constraint 4: $\int_{w_0}^{w_M} wf(w) dw = \overline{w}$
Constraint 5: $\int_{w_0}^{w_M} w^2 f(w) dw = \overline{w^2}$

where, \overline{w} is the mean wind speed of the observed data while $\overline{w^2}$ is the summation of variance of the observed wind speed and square of the mean wind speed.

The optimization variables are $f(w_i)$, (i = 0, 1, ..., N), $f'(w_0)$ and $f'(w_M)$. Prime sign (*i*) denotes the first derivative with respect to the wind speed, *w*. Therefore, the total number of the unknowns is N + 3. Note that, a_i , b_i , c_i and d_i , (i = 1, 2, ..., N) are not optimization variables since they are determined while constructing the spline f(w) based on the node values $f(w_i)$. For example, we have 9 variables in the 7-node optimum spline, 7 variables in 5-node optimum spline and 5 variables in the 3-node optimum spline. However, since there are 5 constraints, three of which are equality constraints, the effective number of variables in an optimum spline decreases at least by 3. Therefore, it is quite fair to compare a 7-node optimum spline with Weibull and Weibull and BEPL methods which have 5 parameters each.

RMSE is the root mean square error which is a measure of the residuals between the fitted distribution and the observed distribution. Constraint 1 guarantees that the cumulative probability over the observed wind speed range is unity because the first and the third terms in the right hand side of Equation 3 vanish since f(w) is defined as zero beyond the observed wind speed range:

$$\int_{0}^{\infty} f(w)dw = \int_{0}^{w_{0}} f(w)dw + \int_{w_{0}}^{w_{M}} f(w)dw + \int_{M}^{\infty} f(w)dw = 1.0$$
(3)

Constraint 2 is due to the nature of the probability density functions which are always positive in value. Constraint 3 is required to avoid statistically possible but physically meaningless solutions. f_{max} is selected to be 1.0 because there is no site (to be considered for wind energy project) that has a wind speed PDF which is greater than 1.0. It should be noted that inequalities (Constraints 2 and 3) in the optimization problem are given in a discrete way while solving the problem numerically.

In Constraints 4 and 5, the first and second moments of the probability density function are used. This is proposed in order to force the same mean and standard deviation values as the observed data. This approach is known as the method of moments in fitting the conventional distributions. However, the conventional distributions obtained using methods of moments do not minimize the *RMSE* and might result in large errors. On the other hand, the proposed method in this study does not only satisfy this empirical condition but also provides minimum *RMSE*. Therefore, it is considered that the fitting spline will provide the optimum probability density function for the related wind speed data.

In order to solve the problem given in Equation 2, the constrained optimization is transformed into the unconstrained optimization by adding the inequality constraints as exterior penalty functions and the equality constraints with Lagrange multipliers to the objective function [29]. The final form

of the objective problem is given in Equation 3. It is solved using a gradient-based optimization algorithm, the conjugate gradient method [30].

$$\begin{aligned} \text{Minimize } f(w_i), \ f'(w_0), \ f'(w_M), \ \lambda_1, \ \lambda_4, \ \lambda_5 \\ RMSE &= \sqrt{\frac{1}{M} \sum_{j=1}^{M} \left[OD(w_j) - f(w_j) \right]^2} \\ &+ \lambda_1 \left(\int_{w_0}^{w_M} f(w) dw - 1 \right) \\ &+ r_{p2} \{ min[0, f(w)] \}^2 + r_{p3} \{ max[0, f(w) - 1] \}^2 \\ &+ \lambda_4 \left(\int_{w_0}^{w_M} wf(w) dw - \overline{w} \right) + \lambda_5 \left(\int_{w_0}^{w_M} w^2 f(w) dw - \overline{w^2} \right) \end{aligned}$$
(4)

The penalty multipliers, r_{p2} and r_{p3} , in Equation (4) are chosen as a small value at the beginning of the optimization process. Along the optimization iterations, the penalty multipliers are increased by a factor. In this study the initial value of the multipliers is selected as 1, and the factor as 3. It should be noted that the number of the optimization variables is increased in this unconstrained form. The new added optimization variables are the Lagrange multipliers λ_1 , λ_4 and λ_5 [29].

At each optimization iteration, *l*, the optimization variables are updated as follow [30]:

$$\begin{aligned}
f(w_i)^{l+1} &= f(w_i)^l + \beta g_1 \\
f'(w_0)^{l+1} &= f'(w_0)^l + \beta g_2 \\
f'(w_M)^{l+1} &= f'(w_N)^l + \beta g_3 \\
\lambda_1^{l+1} &= \lambda_1^l + \beta g_4 \\
\lambda_4^{l+1} &= \lambda_4^l + \beta g_5 \\
\lambda_5^{l+1} &= \lambda_5^l + \beta g_6
\end{aligned} \tag{5}$$

where g_1, g_2, g_3, g_4, g_5 and g_6 are the components of the conjugate gradient direction for the six optimization variables $f(w_i)$, $f'(w_0)$, $f'(w_M)$, λ_1 , λ_4 and λ_5 , respectively. β is the step size along the conjugate gradient direction and is calculated using a simple line search based on trial and error method.

2.2. Weibull Distribution Based Probability Density Function

The Weibull distribution is a two-parametric probability density function. These two parameters are the shape factor, *k*, and the scale factor, *c*, as shown in Equation (6) [31]:

$$WF(w) = \left(\frac{k}{c}\right) \left(\frac{w}{c}\right)^{k-1} \exp\left[-\left(\frac{w}{c}\right)^k\right]$$
(6)

There are different ways to estimate the Weibull parameters, *k* and *c*. The method used in this study is based on the minimization of the root mean square error, RMSE.

2.3. Weibull and Weibull Distribution Based Probability Density Function

The Weibull and Weibull (W and W) distribution is defined in the convex hull of two Weibull distributions containing 5 parameters as given in Equation (7) as [17]:

$$WWF(w) = p\left(\frac{k_1}{c_1}\right)\left(\frac{w}{c_1}\right)^{k_1-1} \exp\left[-\left(\frac{w}{c_1}\right)^{k_1}\right] + (1-p)\left(\frac{k_2}{c_2}\right)\left(\frac{w}{c_2}\right)^{k_2-1} \exp\left[-\left(\frac{w}{c_2}\right)^{k_2}\right]$$
(7)

These five parameters, p, k_1 , k_2 , c_1 and c_2 are estimated based on the minimization of RMSE.

2.4. Beta Exponentiated Power Lindley Distribution Based Probability Density Function

The recently proposed BEPL distribution is defined in Equation (8).

$$BEPL(w) = \frac{k c^2 \omega}{B(a,b)(c+1)} \left(1 + w^k\right) w^{k-1} \exp\left(-c w^k\right) \left[1 - \left(1 + \frac{c w^k}{1+c}\right) \exp\left(-c w^k\right)\right]^{\omega a-1} \times \left\{1 - \left[1 - \left(1 + \frac{c w^k}{1+c}\right) \exp\left(-c w^k\right)\right]^{\omega}\right\}^{b-1}$$
(8)

where, B(a, b) is the Beta function. The five parameters (k, c, ω , a and b) used in the distribution are again estimated by minimizing the RMSE.

2.5. Empirical Probability Density Function

Assigning the normalized frequency as the probability for each observation is known as the empirical probability density function. The normalization has been done by the total number of observations throughout the recorded time period. It should be noted that the empirical method is a nonparametric distribution.

2.6. Kernel Probability Density Function Based on 3rd Order Polynomial

A kernel distribution produces a nonparametric probability density fit which adapts itself to the data. In this study, we defined this distribution by a kernel density estimator as a 3rd order polynomial. Moreover, we implemented two more conditions while defining the distribution: first and second moments of the distribution are set same as the observation data.

2.7. Annual Energy Production

The annual energy production calculations are very vital in the evaluation of any wind energy project. The wind speed distribution is combined with the power curve of a specific wind turbine to give the energy generated at each wind speed and hence the expected total energy generated overall the year. The annual energy production (AEP) can be expressed mathematically as follow:

$$AEP = 8760 \int_{w_0}^{w_M} f(w)P(w)dw$$
(9)

where, P(w), is the power output of a wind turbine at wind speed, w, and, 8760, is the number of hours in the year. f(w) is the distribution function which has a closed form.

3. Analyzed Cases

The wind speed data cases analyzed in this study are shown in Figure 2. Those cases are selected from Yürüşen and Melero [20] as illustrative examples to show the effectiveness of the proposed method over the parametric distributions. It should be noted that those data are synthetic wind speed data. Visual inspection of the histograms hints that Case 1, 2 and Case 4 correspond to multimodal distributions. As for Case 3, the distribution is a simple unimodal one. The statistical properties of the observation data are summarized in Table 1. It should be noted that the mean and the standard deviation values of the distributions are calculated within the range of the observation data.



Figure 2. The analyzed wind speed distribution cases. (**a**) Multimodal distribution 1; (**b**) Multimodal distribution 2; (**c**) Unimodal distribution; (**d**) Multimodal distribution 3.

Table 1. Statistical properties of the observed data.

Case	Minimum Speed (m/s)	Maximum Speed (m/s)	Mean Speed (m/s)	Standard Deviation (m/s)
Case 1	0.0	32.0	19.06	6.87
Case 2	0.0	25.0	8.52	5.37
Case 3	0.0	33.0	13.28	6.58
Case 4	0.0	32.0	10.45	7.21

4. Results

Three different optimum splines are fitted for the cases analyzed. The splines are obtained using 3, 5 and 7 nodes. The mean speed, standard deviation and the RMS error of the different fitted distributions for all the analyzed cases are summarized in Table 2.

The mean speeds and standard deviations of the optimum splines are the same as the observation data thanks to the constraints specified in the optimization problem given in Section 2.1. As seen from Table 2, the RMS errors of Weibull distribution are much higher than the errors of the optimum splines in Cases 1, 2 and 4 (around 4 times more). In Case 3, the RMS error of the Weibull distribution is about twice of the errors of the optimum spline method. Although the number of parameters in the BEPL distribution is 5 and therefore a good performance was expected, the results for all the analyzed cases were observed to be poor in terms of the RMS error. As for the Weibull and Weibull distribution, the RMS errors for all cases were as low as the error of the distribution based on the 5-node optimum spline. However, the distribution based on the 7-node optimum spline is superior to the Weibull and Weibull distribution as seen from the RMS error results in the table. The kernel distribution shows almost the same RMS error performance as the 3-node optimum spline for all the cases.

Case	Method	Mean Speed (m/s)	Standard Deviation (m/s)	RMS Error
	Observed data	19.06	6.87	-
	3-node optimum spline	19.06	6.87	0.00170
	5-node optimum spline	19.06	6.87	0.00054
C 1	7-node optimum spline	19.06	6.87	0.00040
Case 1	Kernel distribution	19.06	6.87	0.00170
	Weibull distribution	20.72	5.11	0.00151
	W and W distribution	18.98	6.75	0.00054
	BEPL distribution	20.68	5.05	0.00161
	Observed data	8.52	5.37	-
	3-node optimum spline	8.52	5.37	0.00140
	5-node optimum spline	8.52	5.37	0.00059
Case 2	7-node optimum spline	8.52	5.37	0.00042
Case 2	Kernel distribution	8.52	5.37	0.00150
	Weibull distribution	8.49	5.64	0.00141
	W and W distribution	8.59	5.36	0.00080
	BEPL distribution	7.54	5.61	0.00226
	Observed data	13.28	6.58	-
	3-node optimum spline	13.28	6.58	0.00046
	5-node optimum spline	13.28	6.58	0.00041
Case 2	7-node optimum spline	13.28	6.58	0.00029
Case 5	Kernel distribution	13.28	6.58	0.00047
	Weibull distribution	13.45	6.74	0.00054
	W and W distribution	13.45	6.64	0.00042
	BEPL distribution	13.16	6.79	0.00131
	Observed data	10.45	7.21	-
	3-node optimum spline	10.45	7.21	0.00140
	5-node optimum spline	10.45	7.21	0.00090
Casa 4	7-node optimum spline	10.45	7.21	0.00023
Case 4	Kernel distribution	10.45	7.21	0.00150
	Weibull distribution	9.95	7.41	0.00117
	W and W distribution	10.57	7.29	0.00063
	BEPL distribution	8.70	7.22	0.00153

Table 2. Fitting results for the analyzed cases.

A comparison of the 7-node optimum spline based distribution fit with the Weibull distribution for the analyzed cases is given in Figure 3. One may notice that the optimum spline based distributions fits very well against the observation data for all the analyzed cases. On the other hand, Weibull based distribution fits well only for Case 3 in which the observation data has a simple unimodal distribution. However, for the other more complicated cases (multimodal distributions), the distributions given by the Weibull method significantly deviate from the observation data. This will obviously affect the accuracy of the expected value of the annual energy production.

Fitting the observation data using the Weibull and Weibull method is compared to the 7-node optimum spline fit in Figure 4. In Cases 2 and 4, the distributions based on the optimum spline and the Weibull and Weibull method are observed to be very close to each other. An eye inspection of Case 1 states that there are more than two modes. Therefore, the proposed method based on the 7-node optimum spline captures the observed data better than the Weibull and Weibull method which is suitable for dual modes. Since there is only one mode in Case 3, there is no superiority of the Weibull and Weibull distribution over the original Weibull distribution as expected. It is also noticed that the optimum spline captures the peak point more accurately for this case.



Figure 3. Comparison of the spline-based and Weibull fits with the observed data. (**a**) Multimodal distribution 1; (**b**) Multimodal distribution 2; (**c**) Unimodal distribution; (**d**) Multimodal distribution 3.



Figure 4. Cont.



Figure 4. Comparison of the splin- based and Weibull and Weibull fits with the observed data. (a) Multimodal distribution 1; (b) Multimodal distribution 2; (c) Unimodal distribution; (d) Multimodal distribution 3.

The BEPL fitting is plotted and compared in Figure 5. As seen from the figure, the BEPL distribution does not show a satisfactory performance in capturing the observation data.



Figure 5. Comparison of the spline based and the Beta Exponentiated Power Lindley fits with the observed data. (**a**) Multimodal distribution 1; (**b**) Multimodal distribution 2; (**c**) Unimodal distribution; (**d**) Multimodal distribution 3.

The fitting results of the optimum splines with different node numbers are compared in Figure 6. It is noticed that the optimum splines based on 5 nodes and 7 nodes are very close to each other and

to the observation data. This result was already seen in Table 2 with very small RMS error values. The optimum spline based on 3 nodes fit only well for Case 3. However, it is still better than the Weibull distribution in terms of RMS error values as seen in Table 2.



Figure 6. Comparison of the optimum splines based on different nodes. (**a**) Multimodal distribution 1; (**b**) Multimodal distribution 2; (**c**) Unimodal distribution; (**d**) Multimodal distribution 3.

In this study, the selected wind turbine is Enercon E53-800 kW. This turbine, like the majority of large size wind turbines, is pitch regulated. This leads to a fixed power output at speeds higher than the rated wind speed. The technical specifications and the power curve of this wind turbine are given in Table 3 and Figure 7, respectively.

Item	Description
Turbine Model	Enercon E53-800 kW
Configuration	Three blade, horizontal axis, upwind
Rated Power	800 kW
Cut-in wind speed	2 m/s
Rated wind speed	12 m/s
Cut-out wind speed	28–34 m/s
Rotor Speed	12–28.3 RPM
Rotor diameter	52.9 m



Figure 7. Power curve of the selected wind turbine.

Using Equation (9), the annual energy production is calculated in all the cases for spline and parametric methods. The results are shown in Table 4. The baseline for comparison with all the methods is selected as the empirical probability density function, that is, the observation data itself. The AEP results based on 7-node optimum spline are almost the same as the baseline with a difference of less than 0.4%. 5-node optimum spline gives also similar AEP values. The Weibull and Weibull method gives AEP estimations with less than 1% deviation. The Weibull distribution resulted in about 5 to 8% differences from the baseline AEP for Cases 2 and 4. The AEP results based on the BEPL method differ within the range of 4 to 22%. The 4% difference belongs to Case 3, which is a unimodal case. The kernel distribution deviates between 1 to 2% from the baseline AEP.

Inaccurate estimations in the AEP will obviously mislead the assessment project and highly increase the investment risk. For example, a difference of 4% leads to about 0.2 MWh in AEP which should not be neglected.

Case	Method	AEP (kWh)	Difference (%)
Case 1	Observation data (empirical distribution)	6.3906×10^{6}	-
	3-node optimum spline	$6.5133 imes 10^6$	1.92
	5-node optimum spline	$6.3616 imes 10^6$	-0.45
	7-node optimum spline	$6.3898 imes10^6$	-0.01
	Kernel distribution	6.5099×10^{6}	1.87
	Weibull distribution	6.9003×10^{6}	7.98
	W and W distribution	$6.3298 imes10^6$	-0.95
	BEPL distribution	$6.8607 imes 10^6$	7.36
	Observation data (empirical distribution)	$3.4685 imes 10^6$	-
	3-node optimum spline	$3.4025 imes 10^6$	-1.90
	5-node optimum spline	$3.4468 imes10^6$	-0.63
Case ?	7-node optimum spline	$3.4560 imes10^6$	-0.36
Cuse 2	Kernel distribution	$3.4374 imes10^6$	-0.90
	Weibull distribution	3.3020×10^{6}	-4.80
	W and W distribution	$3.4925 imes10^6$	0.69
	BEPL distribution	2.7944×10^{6}	-19.43
	Observation data (empirical distribution)	$5.2854 imes 10^6$	-
	3-node optimum spline	$5.2552 imes 10^6$	-0.57
	5-node optimum spline	$5.2636 imes 10^6$	-0.41
Case 3	7-node optimum spline	$5.2857 imes10^6$	0.01
Case 5	Kernel distribution	$5.2525 imes 10^6$	-0.62
	Weibull distribution	5.2831×10^{6}	-0.04
	W and W distribution	$5.2974 imes10^6$	0.23
	BEPL distribution	5.0835×10^{6}	-3.82

Table 4. Comparison of annual energy production calculations based on different distribution fits.

Case	Method	AEP (kWh)	Difference (%)	
	Observation data (empirical distribution)	3.9435×10^{6}	-	
	3-node optimum spline	$3.8694 imes 10^6$	-1.88	
	5-node optimum spline	3.9352×10^6	-0.21	
Casa 4	7-node optimum spline	$3.9335 imes 10^6$	-0.25	
Case 4	Kernel distribution	3.9643×10^{6}	0.53	
	Weibull distribution	3.6287×10^6	-7.98	
	W and W distribution	3.9596×10^{6}	0.41	
	BEPL distribution	$3.0431 imes 10^6$	-22.83	

Table 4. Cont.

5. Conclusions

In this work, a new method is proposed for the probability distribution function of the observed wind speed data. The method is based on fitting splines for a given wind speed versus probability data. The splines obtained are optimized for minimum root mean square error subjected to a number of constraints such as satisfying given mean and standard deviations. The method is compared to the Weibull, Weibull and Weibull and BEPL methods. Among those three distributions, the Weibull method is the most widely used in wind assessment applications. The method is also compared with two nonparametric probability distribution functions: empirical and 3rd order polynomial based kernel distributions.

Four different wind speed distributions, three of which involve multi-modality, are analyzed. The results show that the Weibull and BEPL methods fail in fitting the multi-modal distributions with high RMS errors. The disagreement in fitting causes miscalculation in annual energy production. On the other hand, the optimum spline and the Weibull and Weibull methods produce very small RMS errors with a slight superiority of the distribution based on the 7-node optimum spline. Accurate prediction of AEP is vital in wind energy project assessment to minimize the investment risk. Since the investment cost of wind energy projects is relatively high, even slight prediction error in AEP calculations leads to a great difference in the payback period of the project. Therefore, one may conclude that the AEP calculated based on the optimum splines distributions are considered to be more accurate and reliable.

However, since obtaining the optimum splines requires the solution of a constrained optimization problem, a computation involving a lot of mathematical operations is necessary. This method is recommended when multimodality is present in the wind speed data.

Author Contributions: Investigation, M.A.E. and M.K.; Methodology, M.A.E. and M.K.; Software, M.K.; Validation, M.A.E.; Writing—original draft, M.A.E. and M.K.

Funding: This research received no external funding.

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

References

- 1. Fazelpour, F.; Markarian, E.; Soltani, N. Wind energy potential and economic assessment of four locations in Sistan and Balouchestan province in Iran. *Renew. Energy* **2017**, *109*, 646–667. [CrossRef]
- Fazelpour, F.; Soltani, N. Feasibility study of renewable energy resources and optimization of electrical hybrid energy systems case study for Islamic Azad University-South Tehran branch, Iran. *Therm. Sci.* 2017, 21, 335–351. [CrossRef]
- 3. Dabbaghiyan, A.; Fazelpour, F.; Abnavi, M.D.; Rosen, M.A. Evaluation of wind energy potential in province of Bushehr, Iran. *Renew.Sustain. Energy Rev.* **2016**, *55*, 455–466. [CrossRef]
- Jiang, H.; Wang, J.; Dong, Y.; Lu, H. Comprehensive assessment of wind resources and the low-carbon economy: An empirical study in the Alxa and Xilin Gol Leagues of inner Mongolia, China. *Renew. Sustain. Energy Rev.* 2015, *50*, 1304–1319. [CrossRef]
- 5. Weibull, W. A statistical distribution function of wide applicability. J. Appl. Mech. 1951, 18, 293–297.

- Costa Rocha, P.A.; de Sousa, R.C.; de Andrade, C.F.; da Silva, M.E.V. Comparison of seven numerical methods for determining Weibull parameters for wind energy generation in the northeast region of Brazil. *Appl. Energy* 2011, *89*, 395–400. [CrossRef]
- 7. Oyedepo, S.O.; Adaramola, M.S.; Paul, S.S. Analysis of wind speed data and wind energy potential in three selected locations in south-east Nigeria. *Int. J. Energy Environ. Eng.* **2012**, *3*, 2–11. [CrossRef]
- 8. Mostafaeipour, A.; Jadidi, M.; Mohammadi, K.; Sedaghat, A. An analysis of wind energy potential and economic evaluation in Zahedan, Iran. *Renew. Sustain. Energy Rev.* **2014**, *30*, 641–650. [CrossRef]
- 9. Carneiro, T.C.; Melo, S.P.; Carvalho, P.C.M.; de Braga S., A.P. Particle swarm optimization method for estimation of Weibull parameters: A case study for the Brazilian northeast region. *Renew. Energy* **2016**, *86*, 751–759. [CrossRef]
- 10. Wais, P. Two and three-parameter Weibull distribution in available wind power analysis. *Renew. Energy* **2017**, *103*, 15–29. [CrossRef]
- 11. Shu, Z.R.; Li, Q.S.; Chan, P.W. Investigation of offshore wind energy potential in Hong Kong based on Weibull distribution function. *Appl. Energy* **2015**, *156*, 362–373. [CrossRef]
- 12. Celik, A.N. Energy output estimation for small-scale wind power generators using Weibull-representative wind data. *J. Wind Eng. Ind.Aerod.* **2003**, *91*, 693–707. [CrossRef]
- 13. Fyrippis, I.; Axaopoulos, P.J.; Panayiotou, G. Wind energy potential assessment in Naxos Island, Greece. *Appl. Energy* **2010**, *87*, 577–586. [CrossRef]
- 14. Altunkaynak, A.; Erdik, T.; Dabanlı, İ.; Şen, Z. Theoretical derivation of wind power probability distribution function and applications. *Appl. Energy* **2012**, *92*, 809–814. [CrossRef]
- Mohammadi, K.; Alavi, O.; Mostafaeipour, A.; Goudarzi, N.; Jalilvand, M. Assessing different parameters estimation methods of Weibull distribution to compute wind power density. *Energy Convers. Manag.* 2016, 108, 322–335. [CrossRef]
- 16. Akdağ, S.A.; Bagiorgas, H.S.; Mihalakakou, G. Use of two-component Weibull mixtures in the analysis of wind speed in the Eastern Mediterranean. *Appl. Energy* **2010**, *87*, 2566–2573. [CrossRef]
- 17. Jaramillo, O.A.; Borja, M.A. Wind speed analysis in La Ventosa, Mexico: A bimodal probability distribution case. *Renew. Energy* **2004**, *29*, 1613–1630. [CrossRef]
- 18. Martinez, P.A.A.; Montes, P.J.S.; Zuluaga, E.I.A. A statistical analysis of wind speed distribution models in the Aburrá Valley, Colombia. *J. Oil Gas Altern. Energy Sources* **2014**, *5*, 121–136.
- 19. Draxl, C.; Clifton, A.; Hodge, B.M.; McCaa, J. The wind integration national dataset (WIND) toolkit. *Appl. Energy* **2015**, *151*, 355–366. [CrossRef]
- 20. Yürüşen, N.Y.; Melero, J.J. Probability density function selection based on the characteristics of wind speed data. *J. Phys.: Conf. Ser.* **2016**, 753, 032067. [CrossRef]
- 21. Silverman, B.W. Using kernel density estimates to investigate multimodality. J. R. Stat. Soc. Ser. B (Methodol.) 1981, 43, 97–99.
- 22. Pararai, M.; Liyanage, G.W.; Oluyede, B.O. A new class of generalized power Lindley distribution with applications to lifetime data. *Theor. Math. Appl.* **2015**, *5*, 53–96.
- 23. Docenko, D.; Berzins, K. Spline histogram method for reconstruction of probability density function of clusters of galaxies. *Galaxy J.* 2003, 626, 294–301. [CrossRef]
- 24. Chen, C. Spline Estimators of the Distribution Function of a Variable Measured with Error. Ph.D. Thesis, IOWA State University, Ames, IA, USA, 1999.
- 25. Munkhammar, J.; Mattsson, L.; Ryden, J. Polynomial probability distribution estimation using the method of moments. *PLoS ONE* **2017**, *12*. [CrossRef] [PubMed]
- 26. Morrissey, M.L.; Greene, J.S. Tractable analytic expressions for the wind speed probability density functions using expansions of orthogonal polynomials. *J. Appl. Meteorol. Clim.* **2012**, *51*, 1310–1320. [CrossRef]
- 27. Wijnands, J.S.; Qian, G.; Kuleshov, Y. Spline-based modelling of near-surface wind speeds in tropical cyclones. *Appl. Math. Model.* **2016**, *40*, 8685–8707. [CrossRef]
- Izenman, A.J. Recent developments in nonparametric density estimation. J. Am. Stat. Assoc. 1991, 86, 205–224. [CrossRef]
- 29. Garret, N.V. *Numerical Optimization Techniques for Engineering Design*, 3rd ed.; Vanderplaats Research and Development, Inc.: Colorado Springs, CO, USA, 2001; Chapter 5.

- 30. Douglass, J.W. Optimum Seeking Methods., 1st ed.; Prentice Hall: Englewood Cliffs, NJ, USA, 1964.
- 31. Manwell, J.F.; McGowan, J.G.; Rogers, A.L. *Wind Energy Explained: Theory, Design and Application*; John Wiley & Sons Ltd.: West Sussex, UK, 2002; pp. 55–60, ISBN 0-470-84612-7.



 \odot 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).