TEM Journal. Volume 14, Issue 1, pages 107-116, ISSN 2217-8309, DOI: 10.18421/TEM141-10, February 2025.

Modeling of Wind Speed Distribution in Urban Environment for the Application of Wind Energy Potential Estimation: Case Study

Oleksii Mykhailenko [1], Nadezhda Karabut [1], Volodymyr Doskoch [2], Olena Burtseva [3], Vitaliy Kuznetsov [2], Sergij Tsvirkun [4], Hanna Kolomits [1]

[1] *Kryvyi Rih National University, Kryvyi Rih, Ukraine*
[2] *Ukrainian State University of Science and Technologies, Dnipro, Ukraine*
[3] *Bogdan Khmelnitsky Melitopol State Pedagogical University, Zaporizhzhya, Ukraine*
[4] *SSS "Kryvyi Rih Professional College of National Aviation University", Kryvyi Rih, Ukraine*

***Abstract* – When designing wind farms, the first stage is always an assessment of the target area wind energy potential. It is necessary to have a mathematical description of the wind speeds occurrence probability at the wind turbines potential location to do this. An analysis of relevant studies shows that the most effective approach to obtaining such dependencies is when the wind speed is taken as a random variable. In this case, wind speed distribution in the target area can be modeled using continuous probability distributions. This article is devoted to determining the typical probability distribution models for representing wind conditions in certain areas of the Dnipropetrovsk oblast (Ukraine), which can be used to estimate expected level of power generation by wind power plants. To obtain the data, a series of wind speed measurements were taken at three locations throughout the year. After that, frequency wind speed distributions with ranges of 0.2, 0.5, and 0.8 m/s were created from the obtained dataset and then approximated by continuous probability distributions. Frequency distributions were modeled by Weibull, Rayleigh, Nakagami, gamma, normal, log-normal, generalized extreme value, Birnbaum-Saunders, Wald and Rice continuous distributions.**

To determine the parameters of each type probability distribution, which is the most relevant to the frequency distribution, the maximum likelihood estimation method was used. To assess the accuracy of the models, the Pearson test, coefficient of determination and normalized root mean square deviation are used. The probability distributions quality is also evaluated graphically using Q-Q plots. The best fit to wind speed frequency distributions demonstrated by the Weibull probability distributions. A slightly lower accuracy was provided by the normal, Rice and Nakagami distributions than Weibull distribution. But in some cases, these distributions have even smaller error than the last one. Therefore, after detailed analysis and validation, they can also be used. The Rayleigh distribution had the worst accuracy, the Pearson test for it rejected the null hypothesis that the probability distributions correspond to the frequency distributions at all three locations. Additionally, the effect of the frequency distribution wind speed grouping range on the quality of maximum likelihood estimation of continuous distribution parameters was analyzed. It showed that the approximation accuracy decreases with increasing range.

***Keywords* – Power system, wind energy, wind speed, modeling, probability distributions.**

DOI: 10.18421/TEM141-10
https://doi.org/10.18421/TEM141-10

Corresponding author: Oleksii Mykhailenko,
Kryvyi Rih National University, Kryvyi Rih, Ukraine.
Email: mykhailenko@knu.edu.ua

Received: 17 July 2024.
Revised: 30 December 2024.
Accepted: 31 January 2025.
Published: 27 February 2025.

1. Introduction

Provision of uninterrupted power supply to settlements is a mandatory condition for the development of territorial communities. Taking into account large capacity, significant number of consumers in the urban environment, and significant distances between settlements and substations; the reliability of electricity supply does not always meet the requirements. Also, the time to restore electricity supply after emergency situations can be quite long. The solution to this problem is the use of backup power sources located near one or more separate settlements.

Currently, among the back-up power stations – solar, pumped storage, and wind power plants have become widespread [1]. Their effectiveness depends significantly on the meteorological and ecological conditions of the area in which they are located. Thus, the level of electricity production by photovoltaic converters is low when they are placed in areas with a high level of cloudiness throughout the year. Also, a major factor affecting efficiency is significant dustiness, which is characteristic of industrial cities and metropolises. Wind power stations are more suitable for use in areas with high wind speeds [2], [3], [4]. However, given the presence of rotating parts and the high level of ultrasound generated by the wind turbine, wind power plants are placed outside cities and villages. This leads to the need to build an appropriate electrical network for the integration of such a station into the structure of existing power supply systems. Also, since generating equipment is located quite far away, transmission losses occur. A pumped storage power plant requires special conditions for its implementation [5], [6].

An increase in the number of wind power plants on the territory of the EU countries [1] indicates that the most promising alternative source of electricity right now is wind energy.

In order to eliminate the mentioned shortcomings, it is expedient to use wind turbines with a capacity of up to 100 kW, which can be placed through the deep inlet system in the vicinity of the power supply facility. However, determining the effectiveness of the implementation of a wind power plant is possible only with a preliminary assessment of the wind conditions of the area. At the same time, it is impractical to rely on the data of individual existing weather stations due to their location outside the city limits, because the nature of development significantly affects the wind speed even on the small areas.

This work is devoted to the issue of assessing the wind conditions of an area suitable for the location of a power plant with low-power wind energy installations, intended for the supply of household consumers of a settlement or its separate district, by measuring and determining the statistical models of the wind speed probability distribution.

Traditionally, statistical models of the distribution of probabilities of turning off the wind speed of a certain value are used to estimate the wind potential of the area [7]. This is due to the possibility of modeling on the basis of organized data on the results of meteorological conditions measurements.

Currently, in wind power engineering, the two- [8], [9], [10] or three-parameter [11], [12] Weibull distribution is widely used to estimate the expected level of power generation.

This distribution demonstrates high accuracy of approximation of experimentally obtained wind speed histograms [13]. At the same time, a sufficient quality of approximation is achieved when using a distribution with two parameters – shape and scale, which greatly simplifies the process of model estimation.

Despite the generally accepted use of the Weibull distribution, results of research on wind conditions in different areas demonstrate that in some cases it is possible to use gamma [14], [15], [16] distribution. It is also common to use Gaussian-type distributions, such as normal [17], [18] and log-normal [14], [15]. [19]. Recently, the number of studies on modeling wind conditions using generalized extreme value [15], [17] and Nakagami [18], [19] distributions has increased. It is also common to use Rician [19], Rayleigh–Rician [19], Rayleigh [20], [21], [22] distributions. The wind speed distribution is also modeled by a combination of the distributions presented earlier in [17], [23], [24]. Other methods that are not based on probability distributions are used in [25], [26], [27].

The quality of the built model is significantly affected by the method of determining its parameters. Most of the studies are devoted to evaluating the effectiveness of the parameter's estimation of the Weibull model [28], [29], [30]. In [30], a comparative analysis of the application of the main empirical and computational methods for determining distribution parameters was carried out. The authors used the different criteria to assess the accuracy of the approximation. As a result, it was established that when the measured data do not match the theoretical distribution, which is characteristic for real conditions, the model whose parameters are determined using the maximum likelihood estimation has better quality.

2. Methodology Section

An article is aimed at to determine the statistical model, which allows to approximate with the best accuracy the empirical distribution of the probability of the occurrence of wind speed of a certain value in the conditions of urban areas, using the example of the Dnipropetrovsk oblast. To achieve this goal, the following tasks must be solved;

Conduct an analysis of the geographical location of the Dnipropetrovsk oblast and determine the most suitable areas for the location of a wind power plant;

Determine the points of probable installation of wind turbines in the specified territories and measure the speed and direction of the wind in a time and with a frequency sufficient to obtain a representative sample.

Determine the structures of statistical models based on theoretical probability distributions, which will be used to approximate empirical data, and establish a method for estimating their parameters.

To perform a comparative analysis of the received models and to determine the most suitable for practical application for assessing the wind energy potential of the area.

3. Results

The analysis of data, carried out using the Google Maps resource, showed that in the immediate vicinity of one of the settlements there are three locations that are most suitable for placing a wind power plant. They are located in the Dnipropetrovsk oblast.

To evaluate the wind energy potential of a separate territory, three points were chosen, which are located in the geometric center of each area. In these points a series of wind speed measurements were conducted. Measurements were made by a digital anemometer placed on a ten-meter mast during 2021 up to 5 times a day. As a result, samples of 1825 elements each for a wind speed were obtained for each location. The total volume of received wind speed data was 5475 values. The mast height for the measuring device was chosen according to the standard installation height of a low power wind turbine.

The obtained data was divided into three parts according to the wind speed measurement points, which are related to the northern, eastern and western sites. Hereinafter referred to as the first, second and third sites, respectively. A sample corresponding to the entire territory of possible placement of wind turbines, which combines all data set, was studied separately.

Preparation for the parametric modeling of wind speed distribution was carried out by grouping the data separately in four sets using the method of normalization suitable for approximation by the theoretical probability distribution density.

To study the possibility of empirical distributions approximation by theoretical ones, it was proposed to consider Weibull, Rayleigh, Nakagami, gamma, normal, log-normal, generalized extreme value, Birnbaum-Saunders, Wald and Rice distributions. The Rayleigh distribution is one-parametric, all others are two-parametric (Table 1).

Parameter estimation of relevant statistical models was carried out using the method of maximum likelihood estimation (MLE).

The adequacy of the obtained models was first established using the Pearson's chi-squared test, which allows to check the statistical hypothesis of the correspondence of the wind speed as a random variable to the theoretical law of probability distribution.

For this, the criterion was calculated:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(y_i^O - y_i^E\right)^2}{y_i^E}, \tag{1}$$

where y_i^O is the probability value of the empirical distribution; y_i^E is the probability value of the theoretical distribution.

Table 1. Equations of the probability density functions

Probability distributions	PDF
Weibull (WE)	$p(\upsilon\|\alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{\upsilon}{\alpha}\right)^{\beta-1} e^{-\left(\frac{\upsilon}{\alpha}\right)^{\beta}}$
Rayleigh (RA)	$p(\upsilon\|\alpha) = \frac{\upsilon}{\alpha^2} e^{-\frac{1}{2}\left(\frac{\upsilon}{\alpha}\right)^2}$
Gamma (GA)	$p(\upsilon\|\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \upsilon^{\alpha-1} e^{-\frac{\upsilon}{\beta}}$, where $\Gamma(\,\cdot\,)$ is Gamma-function
Normal (NO)	$p(\upsilon\|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\frac{1}{2}\left(\frac{\upsilon-\mu}{\sigma}\right)^2}$
Log-normal (LN)	$p(\upsilon\|\mu,\sigma) = \frac{1}{\upsilon\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{\ln(\upsilon)-\mu}{\sigma}\right)^2}$
Birnbaum–Saunders (BS)	$p(\upsilon\|\beta,\gamma) = \frac{1}{\sqrt{2\pi}} \times \left(\frac{\left[(x/\beta)^{0.5} + (\beta/x)^{0.5}\right]}{2\gamma x}\right) e^{-\frac{\left[\left(\frac{x}{\beta}\right)^{0.5} - \left(\frac{\beta}{x}\right)^{0.5}\right]}{2\gamma^2}}$,
Wald (IG)	$p(\upsilon\|\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi\upsilon^3}} e^{-\frac{1}{2\upsilon}\lambda\left(\frac{\upsilon-\mu}{\mu}\right)^2}$
Nakagami (NA)	$p(\upsilon\|\mu,\omega) = \frac{2}{\Gamma(\mu)}\left(\frac{\mu}{\omega}\right)^{\mu} \upsilon^{(2\mu-1)} e^{-\frac{\mu}{\omega}\upsilon^2}$, where $\Gamma(\,\cdot\,)$ is Gamma-function
Rice (RI)	$p(\upsilon\|\sigma,\xi) = \frac{1}{\sigma^2} I_0\left(\frac{\upsilon\xi}{\sigma^2}\right) \upsilon e^{-\left(\frac{\upsilon^2+\xi^2}{2\sigma^2}\right)}$ where $I_0(\,\cdot\,)$ is a modified Bessel function
Generalized extreme value (GEV)	$p(\upsilon \mid k,\mu,\sigma) = \left(\frac{1}{\sigma}\right) e^{\left(-\left(1+k\frac{x-\mu}{\sigma}\right)^{-\frac{1}{k}}\right)} \left(1+k\frac{x-\mu}{\sigma}\right)^{-1-\frac{1}{k}}$

Calculations were performed for the points corresponding to the starting points of the grouping ranges of the empirical distribution. The evaluation was carried out for the significance levels $\alpha = 0.05$ and $\alpha = 0.95$.

Taking into account the criticism of statistical hypotheses tests methods that use p-values, the degree of accuracy of the measurement description of results by statistical models was additionally checked by calculating the coefficient of determination R2 and the normalized root mean square deviation NRMSD or $\overline{\overline{RMSD}}$. The last indicator is used instead of the more commonly used root mean square deviation RMSD. The expressions for calculating the above-mentioned indicators are as follows:

$$R^2 = 1 - \left(\frac{\sum_{i=1}^{k} \left(y_i^O - y_i^E \right)^2}{\sum_{i=1}^{k} \left(y_i^O - \overline{y} \right)^2} \right), \qquad (2)$$

where $\overline{y}$ is the average value of the probability of wind calculated according to the data of the empirical distribution;

$$\overline{\overline{RMSD}} = \frac{1}{y_{\max}^O - y_{\min}^O} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left(y_i^O - y_i^E \right)^2}. \qquad (3)$$

Computational experiments were conducted for three different grouping ranges of empirical data at different wind speeds: 0.2 m/s; 0.5 m/s; 0.8 m/s.

Table 2 demonstrates the parameters of the models that provide the best simulation accuracy.

Table 2. Optimal parameters of theoretical distributions at bins width of 0.2 m/s, 0.5 m/s and 0.8 m/s

Distribution	The first location	The second location	The third location
WE	$\alpha = 4.54084$; $\beta = 3.60672$	$\alpha = 4.05054$; $\beta = 3.5194$	$\alpha = 4.11749$; $\beta = 2.95887$
RA	$\alpha = 3.02748$	$\alpha = 2.70047$	$\alpha = 2.7648$
GA	$\alpha = 9.11005$; $\beta = 0.448968$	$\alpha = 8.93411$; $\beta = 0.407451$	$\alpha = 6.74598$; $\beta = 0.543503$
NO	$\mu = 4.09012$; $\sigma = 1.26642$	$\mu = 3.64022$; $\sigma = 1.15653$	$\mu = 3.66646$; $\sigma = 1.3595$
LN	$\mu = 1.35269$; $\sigma = 0.352284$	$\mu = 1.23504$; $\sigma = 0.352156$	$\mu = 1.22328$; $\sigma = 0.404665$
BS	$\beta = 3.8413$; $\gamma = 0.359067$	$\beta = 3.42133$; $\gamma = 0.357071$	$\beta = 3.37853$; $\gamma = 0.412019$
IG	$\mu = 4.09012$; $\lambda = 30.7333$	$\mu = 3.64022$; $\lambda = 27.6688$	$\mu = 3.66646$; $\lambda = 20.7186$
NA	$\mu = 2.61484$; $\omega = 18.3313$	$\mu = 2.536$; $\omega = 14.5851$	$\mu = 1.9324$; $\omega = 15.2883$
RI	$\sigma = 3.85884$; $\xi = 1.31161$	$\sigma = 3.42137$; $\xi = 1.19986$	$\sigma = 3.32617$; $\xi = 1.45342$
GEV	$k = -0.240268$; $\sigma = 1.2536$; $\mu = 3.61383$	$k = -0.252431$; $\sigma = 1.133$; $\mu = 3.21404$	$k = -0.195025$; $\sigma = 1.26367$; $\mu = 3.13328$

Figure 1, Figure 2, and Figure 3 show the Q-Q graphs of the quality of modeling of wind conditions by the proposed theoretical distribution densities for all considered plots of potential wind turbine installation.

The Q-Q graphs demonstrate a high level of agreement between the measurement results and the considered probabilistic models for the normal, Weibull and Rice distributions.

The results of the chi-squared test, calculation of the coefficient of determination and the normalized root mean square deviation for different locations and three speed grouping ranges are summarized in the Table 3.

Table 3. Results of the wind speeds distribution modeling

Bin width, m/s	Distribution type	The first location			The second location			The third location		
		χ^2	$\overline{\overline{RMSD}}$	R^2	χ^2	$\overline{\overline{RMSD}}$	R^2	χ^2	$\overline{\overline{RMSD}}$	R^2
0.2	WE	2.6106	0.0224	0.7288	5.157	0.0382	0.5360	7.2851	0.0356	0.4562
	RA	134.51	0.0341	0.3702	104.67	0.0475	0.2813	63.731	0.0389	0.3515
	GA	13.128	0.0264	0.6219	6.9492	0.0408	0.4704	4.3686	0.0355	0.4592
	NO	4.1465	0.023	0.7128	6.6787	0.0388	0.5199	11.063	0.0367	0.4205
	LN	30.213	0.0288	0.5499	11.88	0.0428	0.4163	7.7999	0.0364	0.4325
	BS	36.396	0.0292	0.5383	12.593	0.043	0.4115	8.5428	0.0364	0.4296
	IG	39.282	0.0294	0.5332	13.181	0.0432	0.4070	9.1836	0.0366	0.4252
	NK	5.9947	0.0245	0.6745	4.9488	0.0384	0.5061	4.8483	0.0354	0.4619
	RI	3.7881	0.0231	0.7102	6.0442	0.0388	0.5208	9.0255	0.0361	0.4409
	GEV	3.707	0.0235	0.70118	5.1843	0.0386	0.52655	6.1404	0.0357	0.45284
0.5	WE	1.4837	0.0350	0.8354	2.1878	0.0489	0.7868	2.2872	0.0445	0.6762
	RA	55.8170	0.0685	0.3705	38.2670	0.0811	0.4126	23.1250	0.0563	0.4806
	GA	5.9529	0.0494	0.6720	4.0749	0.0640	0.6348	1.7939	0.0474	0.6321
	NO	2.2206	0.0375	0.8111	2.7869	0.0523	0.7562	3.4962	0.0472	0.6356
	LN	12.3960	0.0565	0.5709	6.2438	0.0724	0.5319	3.2213	0.0522	0.5533
	BS	15.1920	0.0577	0.5538	6.3418	0.0731	0.5226	3.4472	0.0530	0.5408
	IG	16.4380	0.0581	0.5466	6.5958	0.0738	0.5142	3.6876	0.0535	0.5311
	NK	3.0745	0.0433	0.7486	2.8404	0.0570	0.7100	1.6944	0.0452	0.6653
	RI	2.0696	0.0380	0.8059	2.6294	0.0525	0.7544	2.8201	0.0457	0.6583
	GEV	2.1113	0.0391	0.7943	2.4673	0.0517	0.76106	2.0946	0.0455	0.6609
0.8	WE	0.5896	0.0538	0.8003	1.7587	0.077	0.6724	1.4393	0.0673	0.7430
	RA	32.682	0.097	0.352	26.257	0.1167	0.2466	14.2490	0.0939	0.499
	GA	2.4352	0.0745	0.6176	1.964	0.0943	0.5081	1.2199	0.0787	0.6485
	NO	0.905	0.0577	0.7704	2.1974	0.0807	0.6401	2.2703	0.0696	0.7251
	LN	4.9618	0.0844	0.5086	2.8518	0.1042	0.3996	2.0889	0.0891	0.5491
	BS	5.4473	0.0861	0.4896	2.9843	0.1051	0.3887	2.12	0.0904	0.5361
	IG	5.7227	0.0867	0.4816	3.0908	0.1059	0.3797	2.2753	0.0914	0.5257
	NK	1.1893	0.0659	0.7009	1.6579	0.0863	0.5882	1.1381	0.0716	0.7089
	RI	0.8405	0.0584	0.7647	2.0396	0.081	0.6375	1.7925	0.0686	0.7328
	GEV	0.7998	0.0592	0.75812	1.7705	0.0794	0.6512	1.4406	0.0712	0.7121

At the range of 0.2 m/s (Table 3) and $\alpha = 0.05$, the null hypothesis that the theoretical distribution corresponds to the empirical one was rejected for the Rayleigh distribution on all four plots. For the significance level $\alpha = 0.95$, in addition to the already mentioned Rayleigh distribution, the Pearson test also failed the normal distribution for the first location and the Birnbaum-Sanders distribution and the Wald distribution for both the first location and the total area. In all other cases, the null hypothesis was confirmed.

The analysis of the approximation accuracy indicators at the experimental data grouping ranges of 0.5 m/s and 0.8 m/s (Table 3) for the theoretical distributions demonstrated identical results.

At the significance level $\alpha = 0.05$, the Rayleigh models did not pass the Person test of the statistical hypothesis for all considered locations, where wind turbines are expected to be installed, except for the third one.

At the significance level $\alpha = 0.95$, Pearson's test rejected the models of log-normal, Birnbaum-Sanders and Wald distributions, as in the case of the grouping range of 0.2 m/s. For the grouping range 0.5 and 0.8 m/s, the Birnbaum-Sanders, the log-normal and the Wald distributions passed the test, except case when range was 0.5 m/s or 0.8 m/s for the first location, but with values very close to the cutoff value.

Thus, the Pearson test for the Birnbaum-Sanders distribution, log-normal distribution and the Wald distribution was no more than $\chi^2 = 6.2438$, $\chi^2 = 6.3418$ and $\chi^2 = 6.5958$, respectively. At the limit value of $\chi^2_{\max} = 7.96$ in the case of using a grouping range of 0.5 m/s, and when the grouping range was 0.8 m/s, the Pearson test was equal to $\chi^2 = 2.8518$, $\chi^2 = 2.9843$ and $\chi^2 = 3.0908$, respectively, at the critical value $\chi^2_{\max} = 3.94$. This means that the usage of the above distribution models is permissible.

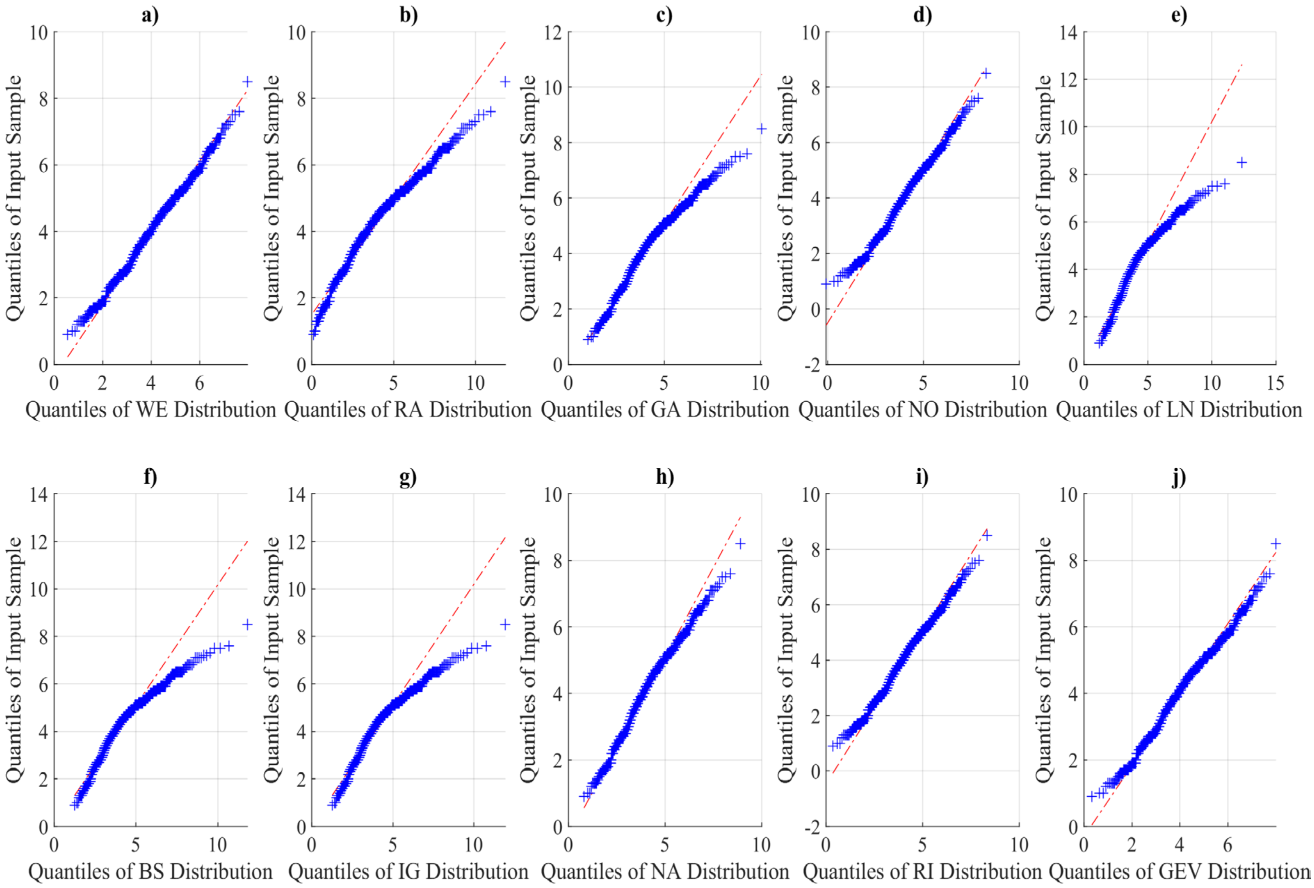

Figure 1. Q-Q plots of the modeling quality of the wind speed distribution for the first location regardless of the bin width

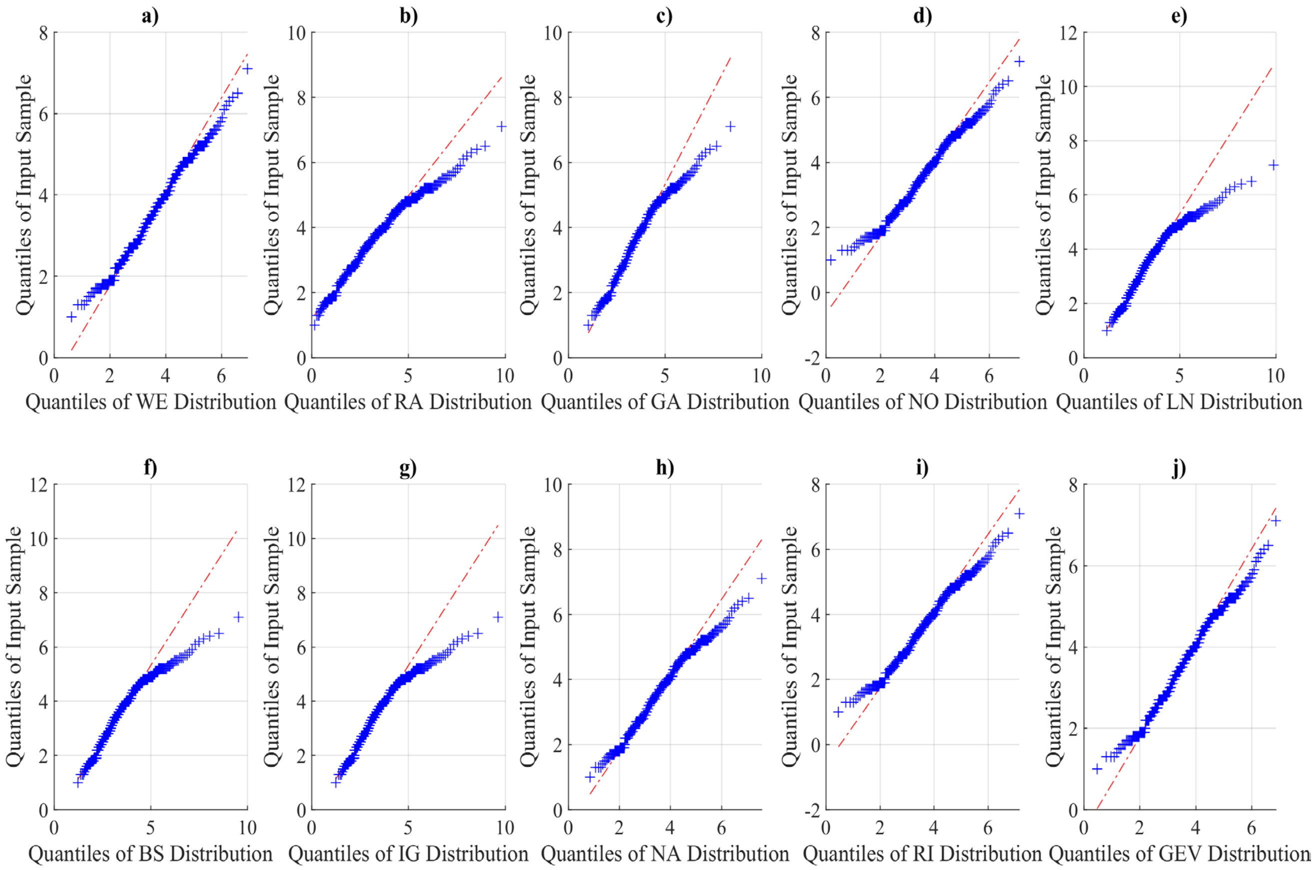

Figure 2. Q-Q plots of the modeling quality of the wind speed distribution for the second location regardless of the bin width

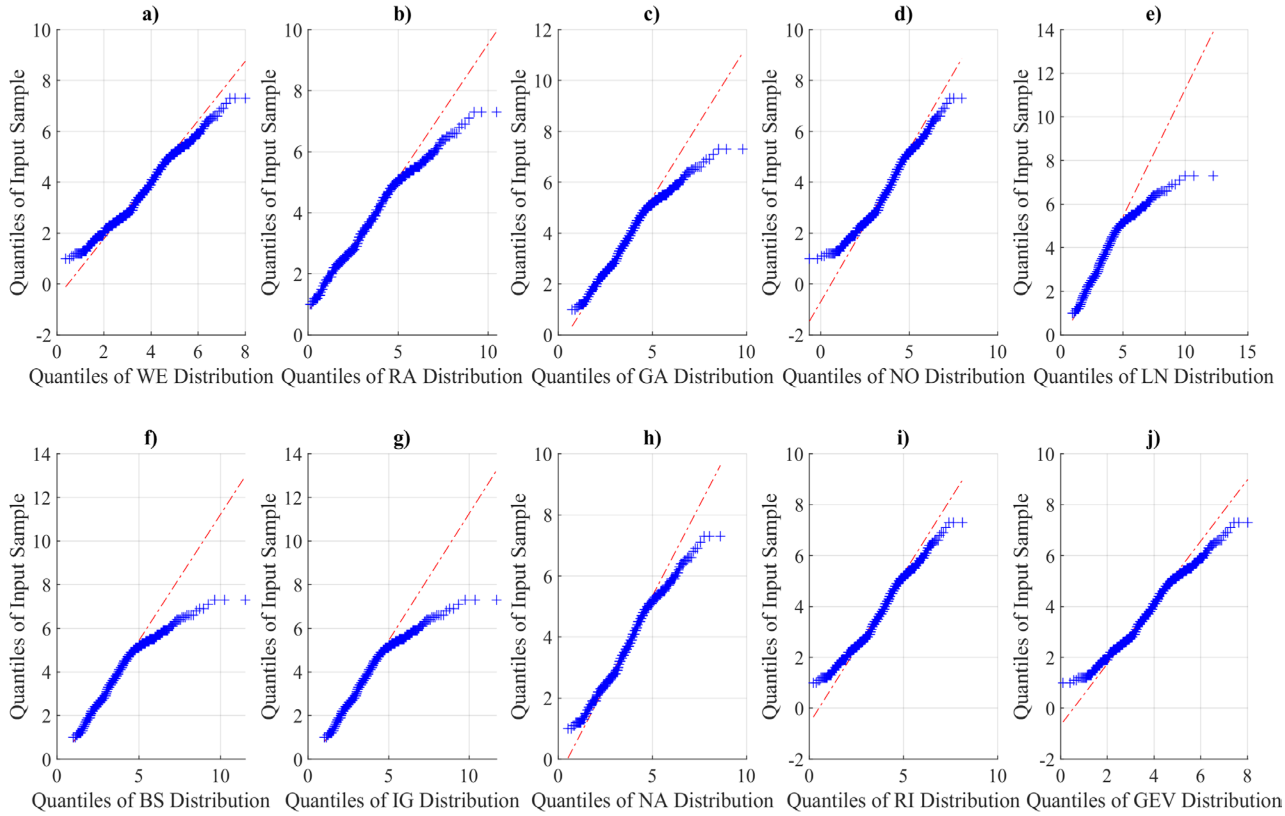

Figure 3. Q-Q plots of the modeling quality of the wind speed distribution for the third location regardless of the bin width

As a result, the Rayleigh distribution showed the worst results when passing the Pearson test for all presented grouping ranges. The Birnbaum-Sanders, log-normal and the Wald distributions are also of low quality. The Weibull distribution has the lowest values of the criterion, regardless of the size of the range and the studied plot. In addition, the Nakagami and Rice distributions, as well as the normal distribution, can be used to approximate the empirical distribution. Particularly, the stability of the Nakagami distribution should be noted. Considering the fact that the calculated values of the criterion for it in most cases exceed the values determined for the Rice and the normal distributions at the respective plots. However, they never exceeded the limit values and did not lead to the deviation of the model. Similar stability was demonstrated by the Weibull distribution.

4. Discussion

The evaluation of the quality of the obtained models using the normalized root mean square deviation and the coefficient of determination showed that the Rayleigh distribution has the worst accuracy when modeling empirical wind speed data. The Wald, lognormal and gamma distributions, as well as the Nakagami distribution, performed best.

The Weibull and Rice distributions, as well as the normal distribution, are the most suitable for practical use.

At the same time, it should be noted that the last two showed approximately the same accuracy.

However, the data analysis made it possible to determine the feature that when the grouping range increases, the quality of the description of the experimental data by the normal distribution increases. Thus, for the first location at a range of 0.2 m/s, the normalized root mean square deviation and the coefficient of determination were approximately the same: $\overline{\overline{RMSD}} = 0.0224$ and $R^2 = 0.7288$ for Weibull distribution; $\overline{\overline{RMSD}} = 0.023$ and $R^2 = 0.7128$ for normal distribution; $\overline{\overline{RMSD}} = 0.0231$ and $R^2 = 0.7102$ for Rice distribution. When changing the range to a value of 0.8 m/s, the quality indicators were $\overline{\overline{RMSD}} = 0.0538$ and $R^2 = 0.8003$; $\overline{\overline{RMSD}} = 0.0577$ and $R^2 = 0.7704$; $\overline{\overline{RMSD}} = 0.0584$ and $R^2 = 0.7647$, respectively. This means that the quality of data approximation by the normal distribution has improved. As a result, only when simulating the wind conditions of the first (northern) plot, the Rice distribution has better accuracy on all three investigated grouping ranges.

In most cases, the model based on the Weibull distribution has the best quality. Only for the first location at ranges of 0.2 m/s and 0.5 m/s Nakagami distribution demonstrated higher accuracy indicators.

Moreover, for a range of 0.5 m/s with equal values of and R2, this fact can be established by comparing the value of the Pearson test for the indicated theoretical distributions.

The results of computational experiments also made it possible to establish the feature that in some cases the application of the Pearson test demonstrates a greater consistency of the theoretical distribution with the empirical one for the model that has worse approximation quality indicators compared to others.

For example, in the mathematical description of the wind speed probability distribution of at the first location (grouping range of 0.2 m/s, Table 3) using the Rice model, the Pearson test was $\chi^2 = 2.6106$ at the values $\overline{\overline{RMSD}} = 0.0224$ and $R^2 = 0.7288$; when applying Rice distribution was $\chi^2 = 3.7881$ at the values $\overline{\overline{RMSD}} = 0.0231$ and $R^2 = 0.7102$; and when applying the normal distribution $\chi^2 = 4.1465$ at the values of $\overline{\overline{RMSD}} = 0.023$ and $R^2 = 0.7128$.

That is, according to the Pearson test and the coefficient of determination, the Rice model is more accurate, and according to the indicator of the normalized root mean square deviation, on the contrary, the normal distribution model.

Therefore, the indicators R^2 and χ^2 align with each other, but normalized root mean square deviation shows the opposite result.

A different situation is observed for the second location (Table 3). According to the Pearson test and the normalized root mean square deviation, the Nakagami distribution has better accuracy compared to the normal and Rice distributions – $\chi^2 = 4.9488$, $\overline{\overline{RMSD}} = 0.0384$ versus $\chi^2 = 6.6787$, $\overline{\overline{RMSD}} = 0.0388$ and $\chi^2 = 6.0442$, $\overline{\overline{RMSD}} = 0.0388$, respectively. According to the coefficient of determination, the results are opposite – the normal and Rice distribution are more qualitative than Nakagami. The indicators for them are $R^2 = 0.5199$ and $R^2 = 0.5208$ against $R^2 = 0.5061$ for the Nakagami distribution.

The analysis of the impact of the wind speed grouping range on the quality of the obtained statistical models showed that the approximation accuracy decreases as the range value increases.

It should be noted that the distribution of models by quality on the three considered grouping ranges has differences for individual plots.

However, the obtained results (Table 3) showed that, in most cases, the Weibull distribution has the best modeling performance.

5. Conclusion

The research of the effectiveness of the application of statistical models based on theoretical distributions in the approximation of the empirical distribution of the occurrence of wind speed of a certain value in the conditions of the Dnipropetrovsk oblast (Ukraine) was carried out. Data on wind conditions were obtained by measuring in plots where wind energy installations are likely to be located; on plots with different openness classes, located in the immediate vicinity of one of the settlements of the oblast. Such location of the wind power plant will allow to realize the advantages of deep input.

The obtained results demonstrate that the Weibull distribution has better approximation accuracy. Rice, Nakagami, and normal distributions showed lower quality, but their error, especially at small grouping ranges, is not significantly different from the Weibull distribution. In one case, with a grouping range of 0.2 m/s, the best quality was demonstrated by the Nakagami distribution. Log-normal and Gamma distributions, as well as the Wald distribution, are suitable for use in tasks that do not require high accuracy. The model based on the Rayleigh distribution has shown the least to achieve in performance. It has the highest normalized root mean square deviation value, the lowest coefficient of determination, and was rejected when tested by the Pearson test in all investigated cases.

It is advisable to calculate the normalized root mean square deviation or the coefficient of determination to estimate the statistical model parameters. The calculation of both at the same time is redundant, since they show the same result. Additionally, the model should be checked by Pearson's test. This will reduce the computational load when choosing a probability distribution function suitable for modeling.

The results of the study showed that continuous probability distributions can be used to model the wind speeds frequency distributions for this case. At the same time, the condition for building high-quality models, as for any statistical method, is the availability of numerous experimental data. And the larger the dataset, the more accurate the models will be. This makes it difficult to apply this approach to assessing wind conditions in real time applications.

Further research will be devoted to determining the level of power generation by wind turbines in a given area using the obtained probability distributions. This will allow optimizing the configuration of the wind farm that can be built. Also, the methodology for modeling the wind speeds frequency distribution presented in this article can be used to assess wind potential in any other location.

References:

[1]. Wind Europe. (2023). *Wind energy in Europe: 2023 Statistics and the outlook for 2024-2030.* WindEurope. Retrieved from: https://windeurope.org/intelligence-platform/product/wind-energy-in-europe-2023-statistics-and-the-outlook-for-2024-2030/. [accessed: 05 June 2024].

[2]. Bartolini, N., Scappaticci, L., Garinei, A., Becchetti, M., & Terzi, L. (2016). Analysing wind turbine state dynamics for fault diagnosis. *Diagnostyka*, *17*(4), 19-25.

[3]. Kachan, Yu. G., Kuznetsov, V. V., & Shram, O. A. (2023). Minimization of power fluctuations of wind power plants when constructed in exclusion zones of enterprises. *Naukovyi Scientific Bulletin of National Mining University, (1),* 115–121. Doi: 10.33271/nvngu/2023-1/115

[4]. Jąderko, A., & Kowalewski, M. K. (2015). Wyznaczanie parametrów wiatru w energetyce odnawialnej. *Przegląd Elektrotechniczny*, (1), 148-151. Doi: 10.15199/48.2015.01.31

[5]. Mykhailenko, O., & Budnikov, K. (2022). Economic aspects of introducing pumped-storage hydroelectric power plants into the mine dewatering system for distributed power generation. *IOP Conference Series: Earth and Environmental Science*, *1049*(1), 012055. Doi: 10.1088/1755-1315/1049/1/012055.

[6]. Sinchuk, I., et al. (2022). Developing the algorithm for the smart control system of distributed power generation of water drainage complexes at iron ore underground mines. *2022 IEEE 8th International Conference on Energy Smart Systems (ESS)*, 116–122. Doi: 10.1109/ESS57819.2022.9969263.

[7]. Shi, H., et al. (2021). Wind speed distributions used in wind energy assessment: a review. *Frontiers in Energy Research*, *9*, 769920. Doi: 10.3389/fenrg.2021.769920.

[8]. Aljeddani, S. M., & Mohammed, M. A. (2023). A novel approach to Weibull distribution for the assessment of wind energy speed. *Alexandria Engineering Journal*, *78*, 56-64. Doi: 10.1016/j.aej.2023.07.027.

[9]. Sedliačková, Z., et al. (2022). Wind speed modeling using Weibull distribution: A case of Liptovský Mikuláš, Slovakia. *MATEC Web of Conferences*, *357*, 08005. Doi: 10.1051/matecconf/202235708005.

[10]. Bidaoui, H., et al. (2019). Wind speed data analysis using Weibull and Rayleigh distribution functions, case study: five cities northern Morocco. *Procedia Manufacturing*, *32*, 786-793. Doi: 10.1016/j.promfg.2019.02.286.

[11]. Chiodo, E., et al. (2023). A Review on Wind Speed Extreme Values Modeling and Bayes Estimation for Wind Power Plant Design and Construction. *Energies*, *16*(14), 5456. Doi: 10.3390/en16145456.

[12]. Pobočíková, I., et al. (2023). Modelling the wind speed using exponentiated Weibull distribution: case study of Poprad-Tatry, Slovakia. *Applied Sciences*, *13*(6), 4031. Doi: 10.3390/app13064031.

[13]. Wang, Z., & Liu, W. (2021). Wind energy potential assessment based on wind speed, its direction and power data. *Scientific reports*, *11*(1), 16879. Doi: 10.1038/s41598-021-96376-7.

[14]. Tizgui, I., et al. (2019). Wind speed distribution modeling for wind power estimation: Case of Agadir in Morocco. *Wind Engineering*, *43*(2), 190-200. Doi: /10.1177/0309524X18780391.

[15]. Filom, S., Radfar, S., & Panahi, R. (2020). A comparative study of different wind speed distribution models for accurate evaluation of onshore wind energy potential: A case study on the southern coasts of Iran. Doi: 10.20944/preprints202004.0067.v1.

[16]. Khan, M. A., et al. (2021). Determination of optimal parametric distribution and technical evaluation of wind resource characteristics for wind power potential at Jhimpir, Pakistan. *IEEE Access*, *9*, 70118-70141. Doi: 10.1109/ACCESS.2021.3078511.

[17]. Kollu, R., et al. (2012). Mixture probability distribution functions to model wind speed distributions. *International Journal of energy and environmental engineering*, *3*, 1-10. Doi: 10.1186/2251-6832-3-27.

[18]. Chen, H., et al. (2021). Assessing probabilistic modelling for wind speed from numerical weather prediction model and observation in the Arctic. *Scientific reports*, *11*(1), 7613. Doi: 10.1038/s41598-021-87299-4.

[19]. Baggio, R., & Muzy, J. F. (2024). Improving probabilistic wind speed forecasting using M-Rice distribution and spatial data integration. *Applied Energy*, *360*, 122840. Doi: 10.1016/j.apenergy.2024.122840.

[20]. Paraschiv, L. S., Paraschiv, S., & Ion, I. V. (2019). Investigation of wind power density distribution using Rayleigh probability density function. *Energy Procedia*, *157*, 1546-1552. Doi: 10.1016/j.egypro.2018.11.320.

[21]. Gorla, R. S. R., Pallikonda, M. K., & Walunj, G. (2020). Use of Rayleigh distribution method for assessment of wind energy output in Cleveland–Ohio. *Renewable Energy Research and Applications*, *1*(1), 11-18. Doi: 10.22044/rera.2019.1601.

[22]. Suwarno, S., & Rohana, R. (2021). Wind speed modeling based on measurement data to predict future wind speed with modified Rayleigh model. *International Journal of Power Electronics and Drive Systems*, *12*(3), 1823. Doi: 10.11591/ijpeds.v12.i3.pp1823-1831.

[23]. Suwarno, S., & Zambak, M. F. (2021). The probability density function for wind speed using modified Weibull distribution. *International Journal of Energy Economics and Policy*, *11*(6), 544-550. Doi: 10.32479/ijeep.11625.

[24]. Syaputra, A., Yendra, R., Marizal, M., & Desvina, A. P. (2022). Wind Speed Modelling using Some Modified Gamma Distribution. *International Journal of Mathematics Trends and Technology-IJMTT*, *68*(6).

[25]. Wang, L., Liu, J., & Qian, F. (2021). Wind speed frequency distribution modeling and wind energy resource assessment based on polynomial regression model. *International Journal of Electrical Power & Energy Systems*, *130*, 106964. Doi: 10.1016/j.ijepes.2021.106964.

[26]. Koivisto, M., Seppänen, J., Mellin, I., Ekström, J., Millar, J., Mammarella, I., ... & Lehtonen, M. (2016). Wind speed modeling using a vector autoregressive process with a time-dependent intercept term. *International Journal of Electrical Power & Energy Systems*, *77*, 91-99. Doi: 10.1016/j.ijepes.2015.11.027.

[27]. Wang, L., & Xue, A. (2023). Wind speed modeling for wind farms based on deterministic broad learning system. *Atmosphere*, *14*(8), 1308. Doi: 10.3390/atmos14081308.

[28]. Shu, Z. R., & Jesson, M. (2021). Estimation of Weibull parameters for wind energy analysis across the UK. *Journal of Renewable and Sustainable Energy*, *13*(2). Doi: 10.1063/5.0038001.

[29]. Lencastre, P., Yazidi, A., & Lind, P. G. (2024). Modeling Wind-Speed Statistics beyond the Weibull Distribution. *Energies*, *17*(11), 2621. Doi: 10.3390/en17112621.

[30]. Alsamamra, H. R., et al. 92022). A comparative study of five numerical methods for the estimation of Weibull parameters for wind energy evaluation at Eastern Jerusalem, Palestine. *Energy Reports*, *8*, 4801-4810. Doi: 10.1016/j.egyr.2022.03.180.