Research Article GU J Sci 36(3): x-x (2023) DOI: 10.35378/gujs.1026834

Gazi University

Journal of Science

http://dergipark.gov.tr/gujs

A Comparative Assessment of Five Different Distributions Based on Five Different Optimization Methods for Modeling Wind Speed Distribution

Mohammed WADI[1,*], Wisam ELMASRY[2]

[1]*Istanbul Sabahattin Zaim University, Faculty of Engineering and Natural Sciences, Istanbul, Turkiye*
[2]*Istanbul Kultur University, Faculty of Engineering, Istanbul, Turkiye*

Highlights

- Modelling wind energy based on five different distributions.
- Five optimization methods are employed to select the optimal parameters per distribution.
- A good-of-fitness is computed based on five error criteria and seven statistical descriptors.

Article Info

Received: 23 Nov 2021
Accepted: 21 June 2022

Keywords

Wind energy modelling
PDF
CDF
Good of fitness
Optimization method

Abstract

Determining wind regime distribution patterns is essential for many reasons; modelling wind power potential is one of the most crucial. In that regard, Weibull, Gamma, and Rayleigh functions are the most widely used distributions for describing wind speed distribution. However, they could not be the best for describing all wind systems. Also, estimation methods play a significant role in deciding which distribution can achieve the best matching. Consequently, alternative distributions and estimation methods are required to be studied. An extensive analysis of five different distributions to describe the wind speeds distribution, namely Rayleigh, Weibull, Inverse Gaussian, Burr Type XII, and Generalized Pareto, are introduced in this study. Further, five metaheuristic optimization methods, Grasshopper Optimization Algorithm, Grey Wolf Optimization, Moth-Flame Optimization, Salp Swarm Algorithm, and Whale Optimization Algorithm, are employed to specify the optimum parameters per distribution. Five error criteria and seven statistical descriptors are utilized to compare the good-of-fitness of the introduced distributions. Therefore, this paper provides different important methods to estimate the wind potential at any site.

REVIEW
FAKE

1. INTRODUCTION

Specifying wind regime distribution patterns is essential for many reasons, like estimating wind potential and designing wind turbines and farms. Therefore, selecting appropriate distributions for determining wind speed distribution accurately is a significant task. Besides, the fineness of the results fundamentally depends on the used distribution function which successes to represent the best fitness of the real wind data. During the last decades, numerous Probability Density Functions (PDFs) have been appeared in the literature to describe wind speed distributions, in particular, Weibull [1-3], Rayleigh [4], Gamma [5], Normal [6-8], Lognormal [9], Logistic [10, 11], Beta [12], Nakagami [13], Burr Type XII [14], Extreme Value [15, 16], Inverse Gaussian [17], Log-Logistic [18], Laplace [19], Half Normal [20, 21], Generalized Extreme Value [22, 23], Generalized Pareto [24, 25], T Location-Scale [26], and others. However, Weibull distribution is the extensively used among these distributions [27-29].

Using Weibull distribution and three years of wind datasets at Catalca in Istanbul, Turkey, the statistical study for wind data was performed [28]. Energy Pattern Factor (EPF), approximation, and graphical assessment methods were examined. Two years of wind data in Pakistan were employed to model the wind profile with the Weibull function [29]. The Weibull parameters were assessed by Cuckoo Search Optimization (CSO), Grey Wolf Optimization (GWO), and Particle Swarm Optimization (PSO). Moreover, four numerical approaches, the Empirical Method of Justus (EMJ), Modified Maximum Likelihood (MML), Method of Moments (MOM), and EPF, were manipulated. Junk and Schindler [30] introduced 24 distributions to assess their Goodness-Of-Fit (GOF) based on four years of wind speed data at different

*Corresponding author, e-mail: mohammed.wadi@izu.edu.tr

sites worldwide. MOM, L-moment, Maximum Likelihood (ML), and least-squares estimation methods were used to estimate these parameters. In addition, the empirical, graphical, MML, and EBF methods were employed with the Weibull function to evaluate the wind turbines' capacity factor [31]. MML outperformed the other methods, whereas the graphical presented the most insufficient matching. Two and three parameters Weibull, two-parameter Gamma, and two-parameter Lognormal were presented to model wind speed at the airport site in Dolny-Hricov [32]. ML method was applied to determine the parameter values. The Weibull function performed the most satisfactory matching. Weibull, Elliptical, and Non-Gaussian distributions were introduced to estimate wind speed distribution at 89 locations over France [33]. The study showed that Elliptical and Non-Gaussian distributions outperformed Weibull in some locations due to their topography and anisotropy discrepancy.

Many invaluable research works appeared in the literature which covered Extreme Value distribution [15, 25, 34-36]. Xaio et al. [37] proposed a technique to describe the extreme wind speed data depends on Extreme Value distribution in Hong Kong. It was compared to two and three parameters Weibull distributions. The obtained results demonstrated that the Extreme Value and Weibull distributions matched best.

Many research works for describing wind speed distribution based on Inverse Gaussian distribution were appeared in [17, 38, 39]. Inverse Gaussian and other nine distributions were examined to evaluate the wind potential in India [38]. The parameters were optimized via Moth Flame Optimization (MFO) and ML estimation method. The results indicated that the bimodal Weibull outperformed the others.

Many invaluable studies based on both Logistic and Log-Logistic distributions to determine the wind speed pattern were presented in [11, 18, 40, 41]. Lin et al. [41] introduced 15 distributions to describe wind speed distribution in Xiamen in China. Among of these distributions both Logistic and Log-Logistic were used. The parameters are estimated by generalized unified probability plotting method. The obtained results showed that Log-Logistic distribution was better than Logistic distribution in terms of performance.

Laplace distribution was also introduced in [8, 19]. Laplace, Uniform, and Gaussian distributions based on golden search method were introduced to estimate wind speed distribution [19]. Laplace distribution achieved better performance than Uniform and Gaussian distributions. Normal distribution was also utilized in many studies [8, 18]. Ten distributions to assess the wind speed shape in Northern Cyprus were introduced by Alayat et al. [18]. The results proved that Generalized Extreme Value distribution gave the top matching for some locations while Log-Logistic, Weibull, and Gamma distributions were the best for the others. However, Normal distribution was not able to provide a good matching in all cases.

The generalized Extreme Value function has been extensively employed in wind distribution modelling [22, 23, 35, 38, 42]. Weibull and Generalized Extreme Value distributions in India were presented to evaluate the wind speed at different locations [23]. The results referred to that Weibull distribution accurately assessed the lower tail of the wind speed data, but it failed to assess the upper tail. In addition, combining Generalized Extreme Value distribution with Weibull distribution was also proposed to improve the matching. Likewise, Generalized Pareto distribution has been utilized in many studies [24, 43-45]. The Generalized Pareto distribution was introduced by Holmes and Moriarty [45] to characterize extreme wind speeds. The results confirmed that this distribution is crucial for selecting the convenient value of the shape parameter. Moreover, Generalized Pareto distribution was also exploited to model the extreme wind speeds on small time scales in Denmark [46] and the method suggested to be an alternative modeling approach for Weibull distribution.

This paper's primary incentive is to scrutinize the performance of Rayleigh, Weibull, Inverse Gaussian, Burr Type XII, and Generalized Pareto distributions in modelling the pattern of wind distribution. The reason behind selecting these distributions is that they may be feasible substitutes to the widely used Weibull and Gamma distributions which in many cases cannot accurately match the wind speed pattern because of the variations of wind speeds characteristic from one site to another. To demonstrate the effectiveness of these distributions on describing wind speed distribution, two recent wind datasets in the

EARLY VIEW

North of Turkey are selected to conduct this analysis. Besides, the performance of the employed distributions and the approximation methods are analyzed by five different accuracy measures.

Regarding the estimation methods, five metaheuristic optimization methods namely, Salp Swarm Algorithm (SSA), Grasshopper Optimization Algorithm (GOA), Whale Optimization Algorithm (WOA), GWO, and MFO are utilized. These methods vary in their robustness, computation complexity and ability to approach solutions.

The rest of this article consists of many sections: Section two introduces the statistical distribution functions, including the formulas of PDF, CDF, and ICDF for each distribution. Furthermore, Section three presents the methodology, including a concise preface of each optimization method. Section four explains the accuracy measures employed to examine the distribution's performance. Besides, Section five explores the results. Finally, Section six concludes the article.

2. STATISTICAL DISTRIBUTIONS

Wind speed profile characterization is crucial in assessing wind availability at a particular location. When the wind speed pattern is described accurately, the other studies belonging to the site can be appropriately specified. Five different distributions are introduced to characterize the distribution of wind speeds. Hereinafter, a concise explanation of these distributions is given.

2.1. Rayleigh Distribution

The British physicist Lord Rayleigh derived the Rayleigh distribution by the end of the 19th century. The Rayleigh distribution with one parameter is a special case of the Weibull distribution when k_W equals two [47]. Due to its ability to accurately describe wind regimes, many research works utilized Rayleigh distribution to assess wind potential at various locations worldwide [48-50]. One-parameter Rayleigh PDF is defined as follows [51]

$$f(v) = \frac{2v}{b_R^2} \exp\left(-\frac{v}{b_R}\right)^2 \tag{1}$$

where $f(v)$ is the probability of the predicted wind speed $(v), v = 0,1,2,\dots,N, N$ is the wind speed vector length, and $b_R > 0$.

The Rayleigh CDF $F(v)$ and ICDF $G(p)$ are defined according to Equations (2) and (3), respectively

$$F(v) = 1 - \exp\left(\frac{v^2}{2b_R^2}\right) \tag{2}$$

$$G(p) = b_R\sqrt{-2\ln(p-1)} \tag{3}$$

where $0 \leq p \leq 1$.

2.2. Weibull Distribution

Weibull distribution is one of the commonly used distributions to assess wind speed because of its high performance [31, 52]. The bi-parameter Weibull distribution depends on shape (k_W) and scale (c_W) parameters. The Weibull PDF is defined [4]:

$$f(v) = \frac{k_W}{c_W}\left(\frac{v}{c_W}\right)^{k_W-1} \exp\left(-\left(\frac{v}{c_W}\right)^{k_W}\right). \tag{4}$$

EARLY VIEW

The Weibull function has essential features, one of them that its parameters that estimated at a particular height can be extrapolated at other heights [53]. The Weibull CDF and ICDF, also known as the quantile [54, 55] are defined as Equations (5) and (6), respectively

$$F(v) = 1 - \exp\left(-\left(\frac{v}{c_W}\right)^{k_W}\right) \tag{5}$$

$$G(p) = -c_W[\ln(1-p)]^{\frac{1}{k_W}}. \tag{6}$$

2.3. Inverse Gaussian Distribution

The Inverse Gaussian function, also called the Wald distribution, is the inverse of the Gaussian distribution. The density of Inverse Gaussian is like Gamma with more significant skewness and a sharper peak. The PDF formula of two-parameter Inverse Gaussian distribution is as given [17]:

$$f(v) = \sqrt{\frac{k_I}{2\pi v^3}} \exp\left(-\frac{k_I(v-c_I)^2}{2c_I^2 v}\right) \tag{7}$$

where c_I and k_I are Inverse Gaussian scale and shape parameters, respectively.

One of the valuable features of Inverse Gaussian distribution is its ability to model the low frequency and the low-speed wind speed data. Moreover, it has a simple computation complexity of its parameters compared to other distributions [56]. The CDF of Inverse Gaussian distribution is given by [17]

$$F(v) = \Phi\left(\sqrt{\frac{k_I}{v}}\left(\frac{v}{c_I} - 1\right)\right) + \exp\left(\frac{2k_I}{c_I}\right)\Phi\left(-\sqrt{\frac{k_I}{v}}\left(\frac{v}{c_I} + 1\right)\right) \tag{8}$$

where $\Phi(v)$ is the standard Gaussian CDF. Finally, Inverse Gaussian distribution has no known closed-form for its ICDF, but it can be numerically estimated using a guess refined by Newton-Raphson iteration.

2.4. Burr Type XII Distribution

The Burr type XII function is a continuous distribution which is originally proposed by Irving W. Burr [57]. The three-parameter Burr PDF is given by the following Equation [58-60]

$$f(v) = \frac{\frac{k_{B2}k_{B1}}{c_B}\left(\frac{v}{c_B}\right)^{k_{B1}-1}}{\left(1+\left(\frac{v}{c_B}\right)^{k_{B1}}\right)^{k_{B2}+1}},\ v > 0, c_B > 0, k_{B1} > 0, k_{B2} > 0 \tag{9}$$

where c_B, k_{B1}, and k_{B2} are the scale, the first and the second shape parameters.

One of the crucial features of Burr distribution is the controllable scale and shape parameters which make it convenient for matching different distributions [14]. The Burr distribution formulas for the CDF [61] and ICDF [62] can be computed from Equations (10) and (11), respectively

$$F(v) = 1 - \frac{1}{\left(1+\left(\frac{v}{c_B}\right)^{k_{B1}}\right)^{k_{B2}}},\ v > 0, c_B > 0, k_{B1} > 0, k_{B2} > 0 \tag{10}$$

$$G(p) = \left((1-p)^{\frac{-1}{k_{B2}}} - 1\right)^{\frac{1}{k_{B1}}}. \tag{11}$$

EARLY VIEW

2.5. Generalized Pareto Distribution

The Generalized Pareto function is a continuous distribution that generally can describe the tails of different distributions [63]. Its PDF can be given [64]:

$$f(v) = \frac{1}{c_{GP}}\left(1 + \frac{k_{GP}(v-\lambda_{GP})}{c_{GP}}\right)^{-\frac{1}{k_{GP}}-1} \tag{12}$$

where c_{GP}, k_{GP}, and λ_{GP} are the scale, shape, and location parameters. One of the important features of Generalized Pareto distribution is its suitability for analyzing gust wind data [65]. The CDF and ICDF of the Generalized Pareto distribution are as presented as follows [66]

$$F(v) = \begin{cases} \exp\left(-\exp\left(\frac{v-\lambda_{GE}}{c_{GE}}\right)\right), k_{GE} = 0 \\ \exp\left(-\left(1 + k_{GE}\frac{v-\lambda_{GE}}{c_{GE}}\right)^{\frac{-1}{k_{GE}}}\right), 1 + k_{GE}\frac{v-\lambda_{GE}}{c_{GE}} > 0, k_{GE} \neq 0 \end{cases} \tag{13}$$

$$G(p) = \begin{cases} \lambda_{GE} - c_{GE}\log(-\log(p)), k_{GE} = 0 \\ \lambda_{GE} + \frac{c_{GE}}{k_{GE}}\left((-\log(p))^{-k_{GE}} - 1\right), k_{GE} \neq 0 \end{cases} \tag{14}$$

Table 1 summarizes the employed distributions with their parameters.

Table 1. *The employed distributions*

Distributions	**Number, (Name) of parameters**	**Parameters**
Rayleigh	1, (Defining parameter)	P1=b_R
Weibull	2, (Scale, Shape)	P1=c_W, P2=k_W
Inverse Gaussian	2, (Scale, Shape)	P1=c_I, P2=k_I
Burr Type XII	3, (Scale, Shape1, Shape2)	P1=c_B, P2=k_{B1}, P3=k_{B2}
Generalized Pareto	3, (Shape, Scale, Location)	P1=k_{GP}, P2=c_{GP}, P3=λ_{GP}

3. METHODOLOGY AND MATERIALS

However, Evolutionary Algorithms (EAs) are very popular in many research areas due to their simplicity, flexibility, and capability to avoid local optima. They may suffer from drawbacks such as long computation time, no guarantee to converge, and having several operating parameters to be adjusted before starting [67]. Five EAs, GOA, GWO, MFO, SSA, and WOA, are employed to overcome these limitations. In this section, these EAs are introduced briefly.

3.1. Grasshopper Optimization Algorithm

GOA is a new algorithm found in 2017 to solve various optimization problems [68]. It is an algorithm that emulates the way of food-seeking of grasshoppers. The grasshopper is a pest that can usually be seen individually in nature, but grasshoppers join swarms when seeking food. The way which a grasshopper swarm is used to move towards food sources is utilized artificially in GOA.

Indeed, the grasshopper position in a grasshopper swarm is naturally affected by three distinct powers, namely, gravity, wind, and social interaction. However, in GOA, only social interaction among grasshoppers in a swarm is considered whereas the others are neglected [68]. GOA is like its EAs counterparts has two major phases, they are exploration and exploitation. In the exploration phase, the

EARLY VIEW

algorithm tends to a highly random behavior that lets the grasshoppers to change their position in a significant and large movements. This helps the swarm to explore its promising regions fastly. On another hand, in the exploitation phase, the grasshopper changes their position in a small scale to let the swarm to search the space locally [69].

Mathematically, the GOA consists of three stages. Firstly, all grasshoppers in the swarm are randomly initialized within the lower and upper bounds [LB, UB]. Afterward, the fitness score per search agent is evaluated by the accompanying index function, and the best solution is saved (T). Then, the algorithm starts the first iteration and, per search agent, updates its position according to the following Equation

$$X_i^D = c\left[\left(\sum_{j=1,\, j\neq i}^{N} c\,\frac{UB_D - LB_D}{2}\, S\left(\left|x_j^D - x_i^D\right|\right)\frac{x_j - x_i}{d_{ij}}\right] + T_D \tag{15}$$

where X_i^D is the next position of the i^{th} search agent in the D^{th} dimension, c is a decreasing factor, N is the swarm size, LB_D and UB_D are the lower and upper boundary in the D^{th} dimension, LB_D is the lower bound in the D^{th} dimension, and S(r) is a function to specify the social relation between two grasshoppers and can be calculated by:

$$S(r) = f\exp\left(\frac{-r}{l}\right) - \exp(-r) \tag{16}$$

where l and f are constants and equal 1.5 and 0.5, respectively [68]. X_j^D and X_i^D are the current positions in the D^{th} dimension of the j^{th} and i^{th} search agents, respectively. X_j and X_i are the current positions of the j^{th} and i^{th} search agents, respectively. d_{ij} is the distance between the current positions of the j^{th} and i^{th} and can be computed based on the following Equation

$$d_{ij} = \left| x_j - x_i \right| . \tag{17}$$

T_D represents the gain of the D^{th} dimension in the optimal solution and indicates the tendency of the grasshoppers to move towards sources of food. The decreasing parameter c is applied twice. The outer c preserves the balance between the exploration and exploitation of the swarm near the target; meanwhile, the inner c reduces the attraction, comfort, and repulsion zones between grasshoppers. Each iteration, c is updated [70]

$$c = UB - t\,\frac{UB - LB}{L} \tag{18}$$

where t is the current iteration and L is the maximum number of iterations. GOA enters the next iteration and repeats the same procedure until L is attained. Figure 1(a) shows the flowchart of GOA.

3.2. Grey Wolf Optimization

GWO simulates the grey wolves hunting in nature. It mainly based on three steps of hunting, namely, searching for booty, encircling booty, then, attacking booty, which is utilized to carry out optimization [71].

The GWO mechanism includes three primary stages. Initially, the search agents (the grey wolves) in the population (collection) are arbitrarily initialized within the period [LB, UB]. Later, the optimum score index per search agent is assessed by the accompanying criterion function. The best solution is alpha, then beta, and delta. The omega includes the remaining group of wolves. Subsequently, the positions of search agents are updated according to the prey position as in Equations (19) to (21)

$$D_\alpha = \left| C_1 X_\alpha - X(t) \right|, D_\beta = \left| C_2 X_\beta - X(t) \right|, D_\delta = \left| C_3 X_\delta - X(t) \right| \tag{19}$$

$$X_\alpha = X_\alpha - A_1 D_\alpha, X_\beta = X_\beta - A_2 D_\beta, X_\delta = X_\delta - A_3 D_\delta \tag{20}$$

EARLY VIEW

$$X(t+1) = \frac{X_\alpha + X_\beta + X_\delta}{3} \quad (21)$$

where t, $X(t)$, $X(t+1)$, X_α, X_β, X_δ denote the current iteration, the current and following positions of the prey, and the alpha, beta, and delta positions, respectively. The vectors A and C are computed as follows

$$A = 2.a.r_1 - a \quad (22)$$

$$C = 2.r_2 \quad (23)$$

where the random vectors r_1 and r_2 are uniformly distributed in the period [0, 1], and the control vector a is linearly decreased within the period [2, 0] for best candidate solutions. GWO continues in the same procedure until the maximum iteration number is fulfilled. Figure 1 (b) shows the flowchart of GWO.

3.3. Moth-Flame Optimization

MFO is a bio-inspired optimization that simulates the moth behavior when flying towards moonlight or artificial light (flame) [72]. Mathematically, The MFA mechanism includes two steps. Firstly, all the search agents (moths) in the population (swarm) are arbitrarily initialized within the period [LB, UB]. Afterward, each search agent estimates the best score utilizing the accompanying criterion function. The best solution so far for each moth is saved as a separate flame ($F_i^N, i = 1, \ldots, N$, where N represents the population size). Then, the algorithm starts the first iteration by sorting the moths according to their flames. Afterwards, MFO calculates the values of $Flame_{No}$, a, and t using Equations (24), (25), and (26), respectively [73]

$$Flame\_No = round\left((N-l)\frac{N-1}{T}\right) \quad (24)$$

where l and T are the current and the maximum number of iterations

$$a = -1 + (l\ \frac{-1}{T}) \quad (25)$$

$$t = (a-1)\,r + 1 \quad (26)$$

where t is the vector of random numbers in the period [-1, 1], and r is the vector of random numbers in the period [0, 1].

Then, the algorithm determines the distance between each search agent and its corresponding flame as follows:

$$D_i = | F_j - M_i | \quad (27)$$

where D_i is the space between the i^{th} moth (M_i) and the j^{th} flame (F_j). Consequently, MFO updates the position for each search agent according to the logarithmic spiral function (S):

$$S(M_i, F_j) = D_i\ \exp(bt)\ \cos(2\pi t) + F_j \quad (28)$$

where the constant b is for representing the logarithmic spiral, SSA continues in the same procedure until T is attained. Figure 1 (c) shows the flowchart of SSA.

EARLY VIEW

3.4. Salp Swarm Algorithm

SSA is a recent metaheuristic algorithm that mimics the sea squirts (salps) behavior when navigating for food in deep seas [74]. Indeed, salps usually tend to move in a chain. Accordingly, the swarm consists of the commander and the disciples. The commander is at the front of the salp and guides the disciples.

The SSA mechanism contains three steps: initially, the search agents (salps) in the population (chain) are arbitrarily initialized within the period [LB, UB]. Then, the criterion function assesses the optimal score per search agent, and the best solution is saved (F). After that, the algorithm starts the first iteration, and the leader search agent renews its position according to Equation (29) [74]

$$x_j^1 = \begin{cases} F_j + c_1((UB_j - LB_j)c_2 + LB_j & c_3 \geq 0 \\ F_j - c_1((UB_j - LB_j)c_2 + LB_j & c_3 < 0 \end{cases} \quad (29)$$

where x_j^1 is the commander salp position in the j^{th} dimension, F_j is the optimum solution in the j^{th} dimension, UB_j and LB_j are the boundaries of upper and lower parameters in the j^{th} dimension, respectively, c_2 and c_3 are uniformly distributed values in the period [0, 1], and c_1 is the coefficient that maintains the equilibrium the exploration and exploitation processes of SSA and computed at each iteration as follows:

$$c_1 = 2\exp\left(-\left(\frac{4t}{L}\right)^2\right) \quad (30)$$

where t and L are the current and the maximum number of iterations. Moreover, the followers salps updating their positions according to the following Equation [74].

$$x_j^i = 0.5\left(x_j^i + x_j^{i-1}\right) \quad (31)$$

where $(i \geq 2)$ and x_j^i is the present placement of the i^{th} follower salp in the j^{th} dimension. Then, SSA proceeds to the next iteration and repeats the same procedure until L is fulfilled. Figure 1 (d) shows the flowchart of SSA.

3.5. Whale Optimization Algorithm

WOA is a new algorithm that appeared in 2016 for solving optimization problems [75]. It mimics the behavior of whales in hunting their prey [76]. In WOA, a population of whales (search agents) evolves to find the global optima after a defined iteration number. WOA begins with the initialization of search agents randomly upon the interval of LB and UB of the problem variables. After that, WOA evaluates the best score per search agent by using the fitness function. The best solution is saved for further processing later. Then, WOA updates the position of each search agent depending on the following cases. If a random number (z) is less than 0.5 as well as A vector is less than 1, then the particular search agent applies the Encircling method by updating its position for the next iteration using the following formulas

$$X(t+1) = X^*(t) - A.D \quad (32)$$

$$D = | C.X^*(t) - X(t) | \quad (33)$$

$$A = 2.a.r - a \quad (34)$$

$$C = 2.r \quad (35)$$

where $X^*(t)$ is the best solution position. Furthermore, A and C are factor vectors. The a vector is linearly decreased in the period [1, 0]. The r is the random vector in the period [0, 1]. Else if (random number $(z) \leq$

EARLY VIEW

0.5), but A vector $\geq$ 1, then the particular search agent applies the exploration method by updating its position for the next iteration using the following formulas:

$$D = | C.X_{rand} - X(t) | \tag{36}$$

$$X(t+1) = X_{rand} - A.D \tag{37}$$

where X_{rand} is the random whale in the present iteration, otherwise, the particular search agent applies the Spiral method by updating its position for the next iteration using the following Equation:

$$X(t+1) = \begin{cases} X^*(t) - A.D, if z < 0.5 \\ D'.\exp(bl).\cos(2\pi l) + X^*, if z \geq 0.5 \end{cases} \tag{38}$$

where $D' = |X^* - X(t)|$ represents the space between the search agent to the prey, b is constant, and l is a random value within the period [-1, 1]. Finally, WOA continues in the exact methodology until the pre-described number of iterations is attained. Figure 1 (e) depicts the flowchart of WOA.

3.6. Methodology

A distribution's parameter selection can be regarded as a nonlinear problem that reduces the MAE between the collected and expected wind vectors, as follows

$$min\{MAE(V_m, V_d)\} \tag{39}$$

where V_m and V_d are the collected and predicted speed vectors, respectively, and V_d can be calculated based on the ICDF, as expressed in Section 2.

To solve the problem of Equation (39), EAs can be used [77-80]. A population of search agents characterizes the optimal solution in GOA, GWO, MFO, SSA, and WOA. The initial population of search agents is arbitrarily produced per distribution based on the boundaries of their parameters.

Afterward, Equation (39) computes the best score values per search agent. The population then increases by exploring the optimum solution based on the defined operations of the presented EAs. It continues in the exact methodology until the pre-described number of iterations is attained. Table 2 displays the distribution parameter values within their recommended range.

Table 2. *The GWO and WOA operating parameters*

Item	**Range**	**Recommended value**
Size of population	[5,50]	50
Maximum iteration numbers	[50,300]	100 for Rayleigh, 200 for Weibull and Inverse Gaussian, 300 for Burr and Generalized Pareto
Termination criterion	[1xE^{-4},1xE^{-6}]	1xE^{-6}

Table 3 shows the optimal parameter values per the presented optimization methods and distributions. Notably, Rayleigh distribution based on all optimization methods provided very steady parameter values. Furthermore, Rayleigh was the only distribution that compatible with GOA. on the other hand, GWO was

EARLY VIEW

the only optimization method that compatible with all distributions. WOA was also compatible with all distributions except Generalized Pareto. Besides, all optimization methods except GWO failed to approach the optimal parameter values with Generalized Pareto. Regarding the running time, Table 4 illustrates the elapsed time of the introduced optimization methods in seconds. It can be perceived that Rayleigh based on all optimization methods was the faster in convergence than other distributions, but Inverse Gaussian was the worst in all cases.

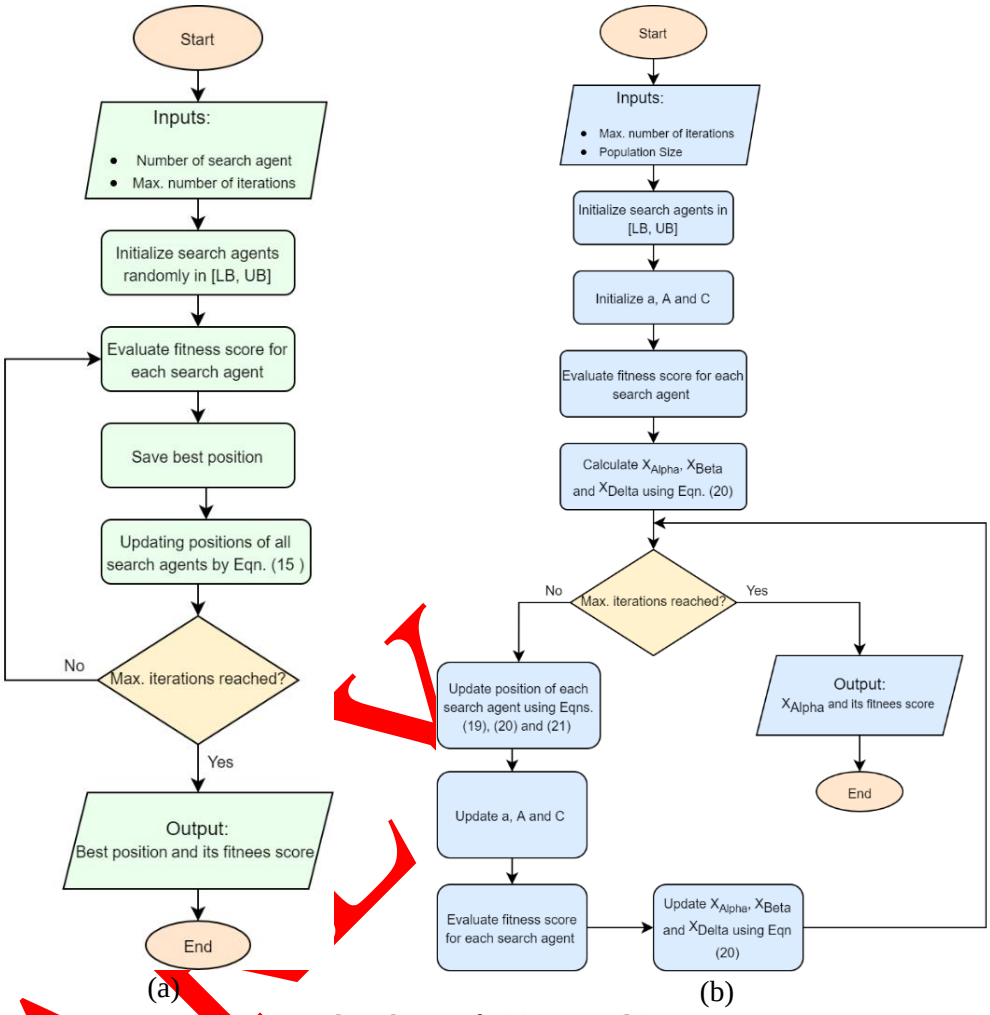


Figure 1. *Flowcharts of: (a) GOA, (b) GWO*

4. ERROR CRITERIA

Many error criteria can be used to describe the performance of any distribution. In this paper, four widely used error criteria in addition to the remarkable net fitness test are applied. These criteria measure the grade of GOF. The utilized error criteria hereinafter are briefly discussed:

- ***MAE*** represents the arithmetic average between the actual (x) and the expected (y) wind speed vectors [81].

$$MAE = \frac{\sum_{i=1}^{N} |y_i - x_i|}{N} \tag{40}$$

where *N* represents the vector size.

- ***RMSE*** is the square root of the average of the square of the differences between the expected and the actual wind speed values [82-85].

EARLY

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - x_i)^2}{N}} \tag{41}$$

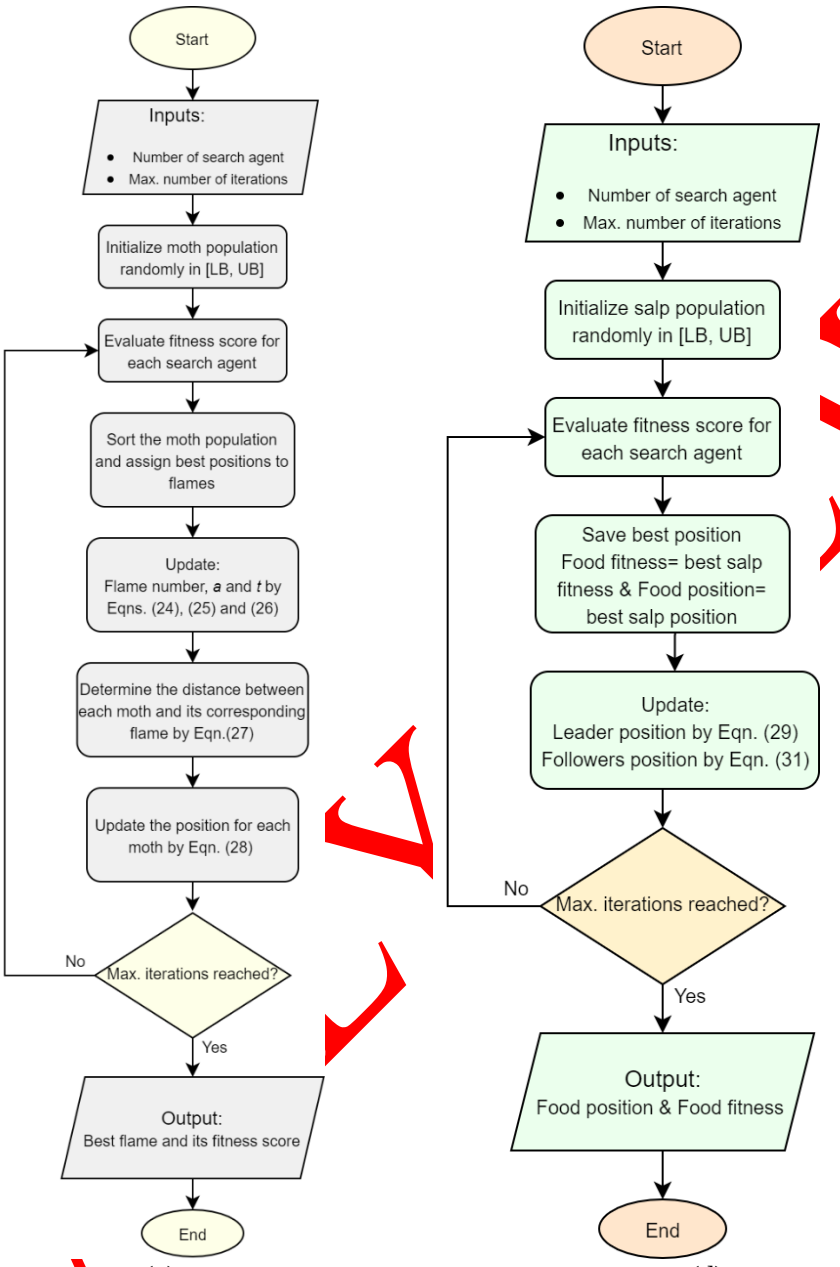


Figure 1. *Flowcharts of: (c) MFO, (d) SSA*

- ***The coefficient of regression*** shows the degree of the linearity between the expected and the actual data. It can be calculated by Equation (42)

$$R^2 = \frac{\sum_{i=1}^{N}(x_i - \mu_i)^2 - \sum_{i=1}^{N}(x_i - y_i)^2}{\sum_{i=1}^{N}(x_i - \mu_i)^2} \tag{42}$$

where μ_i is the i^{th} average of gathered wind speed data.

- ***Correlation Coefficient*** describes the correlation degree between two sets. It has three values, -1 (opposite correlation) and 1 (a perfect positive correlation). However, the zero value indicates that

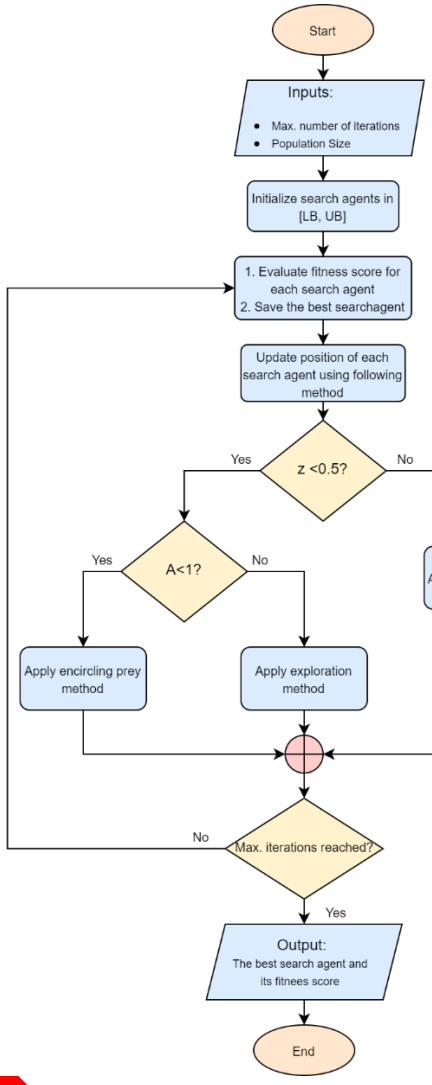


Figure 1. *Flowcharts of: (e) WOA*

Table 3. *Distributions parameter values generated by optimization methods*

Distributions	Parameter	Datasets									
		2019					2020				
		GOA	GWO	MFO	SSA	WOA	GOA	GWO	MFO	SSA	WOA
Rayleigh	P1	3.25	3.25	3.25	3.25	3.25	3.46	3.46	3.46	3.46	3.46
	P2	-	-	-	-	-	-	-	-	-	-
	P3	-	-	-	-	-	-	-	-	-	-
Weibull	P1	4.17	4.53	4.52	4.17	4.53	4.17	4.78	4.79	4.79	4.79
	P2	852.9	1.84	1.84	442.4	1.84	977.7	1.75	1.75	1.75	1.75
	P3	-	-	-	-	-	-	-	-	-	-
Inverse Gaussian	P1	4.52	4.20	4.20	4.20	4.23	4.87	4.42	4.42	4.47	4.43
	P2	223.8	15.1	15.14	15.2	14.43	78.35	13.52	13.5	16.4	13.5
	P3	-	-	-	-	-	-	-	-	-	-
Burr Type XII	P1	4.280	24.2	1000	999.3	5.77	4.20	189.8	1000	997.6	380.3
	P2	270.9	1.91	1.000	1.017	2.45	965.7	1.75	1.000	1.01	1.573
	P3	1000	24.8	290.2	293.2	2.36	734.2	627.4	263.5	262.5	1000

FAIL VIEW

Generalized Pareto	P1	-1000	-0.42	-183.9	-611.9	-115.6	-1000	-0.513	-65.35	-14.0	-120
	P2	859.62	4.1	918.2	321.4	258.9	957.4	5.06	751.74	155.2	369.3
	P3	3.310	1.212	-0.82	3.645	1.93	3.21	0.896	-7.33	-6.40	1.10

Table 4. *Elapsed running time in seconds*

Distributions	**Datasets**									
	2019					**2020**				
	GOA	GWO	MFO	SSA	WOA	GOA	GWO	MFO	SSA	WOA
Rayleigh	5.794	4.87	3.26	3.26	3.52	14.71	6.07	4.48	4.43	4.8
Weibull	30.06	14.28	12.91	11.2	11.71	21.94	19.72	16.42	14.9	16.75
Inverse Gaussian	856.05	477.7	188.9	339.1	190.13	411.5	295	267.2	467.4	243.5
Burr Type XII	55.97	99.10	29.3	38.6	42.03	67.39	68.7	39.13	53.41	56.7
Generalized Pareto	23.65	41.61	15.1	15.4	16.40	22.69	25.5	19.46	21.76	19.97

the two datasets are wholly different (no correlation). The correlation coefficient is given by Equation (43) [83]

$$R = \frac{1}{N-1}\sum_{i=1}^{N}\frac{(x_i-\bar{x})(y_i-\bar{y})}{\sigma_x\sigma_y} \tag{43}$$

where $(\bar{x}, \bar{y})$ and (σ_x, σ_y) denote the average and the standard deviation of the measured and expected wind speed vectors, respectively.

- ***Net Fitness*** is a great test used to find the average of other error measures. It is essential to rank all used distributions and estimation methods according to their accuracy. In this paper, its formula is defined in Equation (44) [29]

$$NetFitness = \frac{\sum_{i=1}^{n}|MAE_i| + \sum_{i=1}^{n}|RMSE_i| + \sum_{i=1}^{n}(1-R_i^2) + \sum_{i=1}^{n}(1-R_i)}{4n} \tag{44}$$

where *n* is the whole number of error measures.

5. RESULTS AND DISCUSSION

The collected half-hourly wind speed data from the Catalca zone in Turkey for the 2019 and 2020 years at 10m height are employed to test the performance of the introduced distributions and optimization methods. Table 5 depicts the Catalca information.

Table 5. *Data of the selected site*

Station name	**Site name**	**State**	**Country**	**Latitude (^{O}) N**	**Longitude (^{O}) E**	**Altitude (m)**	**Years**
Ataturk	Catalca	Istanbul	Turkey	40.967	28.817	37	2019-2020

Five optimization algorithms, GOA, GWO, MFO, SSA and WOA were exploited to estimate the parameters per distribution. The statistical identifiers like μ, σ, σ^2, Minimum (Min), Maximum (Max), Skewness (Skew), and Kurtosis (Kurt) in addition to average power density (P_{avg}) to characterize the pattern of wind for the two years are given in Tables 6 and 7.

EARLY VIEW

The average wind speed value is a crucial indication of wind availability at any site. The mean wind speed values of the actual data are 4.30 and 4.5 m/s for the 2019 and 2020 datasets, respectively. The standard deviation value shows how the speed values differ from the mean. The variance shows the degree of the spread between wind speed values from their mean. The minimum actual wind speed was zero, whereas the maximum varied between 14.40 and 15.60 m/s.

Table 6. *Statistical scrutiny for all distributions of the 2019 dataset*

Optimization	Distribution	μ	σ	σ^2	Min	Max	Skew	Kurt	P_{avg} (W/m2)
-	Real	4.307	2.254	5.079	0.000	14.440	0.756	0.505	94.41
GOA	Rayleigh	4.329	2.126	4.520	0.367	13.968	0.622	0.276	89.32
	Weibull	4.169	0.006	0.000	4.146	4.182	-0.955	1.421	44.37
	Inverse Gaussian	4.601	0.633	0.400	3.151	7.545	0.473	0.372	63.11
	Burr	4.166	0.018	0.000	4.095	4.207	-0.943	1.384	44.29
	Generalized Pareto	4.170	0.000	0.000	4.169	4.170	-12.43	152.42	44.41
GWO	Rayleigh	4.323	2.123	4.507	0.367	13.948	0.622	0.276	88.94
	Weibull	4.285	2.273	5.167	0.290	15.142	0.734	0.501	94.15
	Inverse Gaussian	4.440	2.278	5.188	1.138	21.144	1.583	4.095	107.4
	Burr	4.294	2.266	5.134	0.315	15.856	0.791	0.736	94.63
	Generalized Pareto	4.343	2.152	4.630	1.238	10.843	0.623	-0.425	90.92
MFO	Rayleigh	4.323	2.123	4.507	0.367	13.948	0.622	0.276	88.934
	Weibull	4.279	2.266	5.134	0.291	15.089	0.732	0.495	93.567
	Inverse Gaussian	4.440	2.276	5.180	1.139	21.126	1.581	4.090	107.27
	Burr	3.799	3.624	13.131	0.022	32.252	1.959	5.728	182.31
	Generalized Pareto	4.158	0.124	0.015	2.624	4.170	-11.94	143.84	44.13
SSA	Rayleigh	4.323	2.123	4.507	0.367	13.948	0.622	0.276	88.94
	Weibull	4.168	0.011	0.000	4.124	4.192	-0.950	1.406	44.34
	Inverse Gaussian	4.440	2.276	5.180	1.139	21.126	1.581	4.090	107.3
	Burr	4.090	3.837	14.724	0.026	33.742	1.912	5.430	218.7
	Generalized Pareto	4.170	0.001	0.000	4.159	4.170	-12.43	152.420	44.41
WOA	Rayleigh	4.323	2.123	4.507	0.367	13.95	0.622	0.276	88.93
	Weibull	4.286	2.275	5.174	0.289	15.154	0.735	0.503	94.26
	Inverse Gaussian	4.476	2.359	5.563	1.110	22.00	1.623	4.306	113.7
	Burr	4.404	2.446	5.982	0.514	28.23	1.807	7.356	116.93
	Generalized Pareto	4.160	0.090	0.008	3.098	4.170	-10.66	119.272	44.15

Skewness indicates the degree of deviation from the symmetrical normal distribution. The skewness of the actual data of 2019 and 2020 datasets shows that the actual data tend to have a moderate and positively skewed shape. Examining the skewness values of Rayleigh and Inverse Gaussian distributions based on GOA, it can be noticed that only both followed the skewness pattern of the actual data. However, other

EARLY VIEW

distributions completely failed. In addition, Generalized Pareto was the most anomaly with negative skewness values (negatively skewed).

Kurtosis measures the outliers values of a frequency distribution. Kurtosis has three classes: positive, negative, and zero. Zero kurtosis distribution generally follows the Normal distribution; the kurtosis with positive values has heavier tails and more peak-heavy than Normal, whereas the kurtosis with negative has lighter tails and a flatter peak than Normal [84]. The actual data of the 2019 and 2020 datasets tend to the positive kurtosis pattern with a flat top near the mean rather than a sharp peak. Apparently, the Generalized Pareto distribution recorded very high kurtosis values compared to actual data. Consequently, the estimated wind pattern radically differs from the actual wind pattern.

Table 7. *Statistical scrutiny for all distributions of the 2020 dataset*

Optimization	Distribution	μ	σ	σ^2	Min	Max	Skew	Kurt	P_{avg} (W/m2)
-	Real	4.470	2.491	6.205	0.000	15.56	0.632	0.218	111.65
GOA	Rayleigh	4.576	2.271	5.158	0.610	14.85	0.632	0.239	106.58
	Weibull	4.169	0.005	0.000	4.153	4.180	-0.825	0.768	44.380
	Inverse Gaussian	4.997	1.216	1.478	2.788	11.55	0.793	0.950	90.884
	Burr	4.169	0.005	0.000	4.153	4.181	-0.823	0.766	44.389
	Generalized Pareto	4.170	0.000	0.000	4.170	4.170	-7.872	59.98	44.413
GWO	Rayleigh	4.576	2.271	5.158	0.610	14.85	0.632	0.239	106.58
	Weibull	4.520	2.541	6.457	0.441	17.04	0.813	0.625	118.36
	Inverse Gaussian	4.667	2.618	6.853	1.260	24.64	1.695	4.588	139.65
	Burr	4.521	2.541	6.458	0.442	17.08	0.816	0.634	118.45
	Generalized Pareto	4.476	2.361	5.574	0.974	10.665	0.479	-0.714	104.64
MFO	Rayleigh	4.576	2.271	5.158	0.610	14.85	0.632	0.239	106.57
	Weibull	4.523	2.537	6.437	0.444	17.01	0.810	0.617	118.28
	Inverse Gaussian	4.667	2.618	6.854	1.260	24.64	1.695	4.588	139.69
	Burr	4.153	4.001	16.010	0.059	35.58	1.943	5.539	242.26
	Generalized Pareto	4.080	0.539	0.290	-0.002	4.170	-6.899	48.204	43.125
SSA	Rayleigh	4.576	2.271	5.158	0.610	14.85	0.632	0.239	106.57
	Weibull	4.523	2.537	6.437	0.444	17.01	0.810	0.617	118.27
	Inverse Gaussian	4.702	2.411	5.811	1.409	22.29	1.559	3.874	127.26
	Burr	4.357	4.159	17.297	0.064	36.73	1.917	5.377	273.60
	Generalized Pareto	4.175	1.583	2.505	-4.231	4.720	-3.781	14.666	54.622
WOA	Rayleigh	4.576	2.271	5.158	0.610	14.85	0.632	0.239	106.57
	Weibull	4.523	2.537	6.439	0.444	17.01	0.810	0.617	118.28
	Inverse Gaussian	4.668	2.619	6.858	1.260	24.65	1.695	4.590	139.75
	Burr	4.511	2.798	7.831	0.333	19.38	0.975	1.061	134.19
	Generalized Pareto	4.162	0.059	0.003	3.696	4.170	-7.689	57.944	44.174

EARLY VIEW

The actual $Pavg$ powers at the location are 94.40 and 111.70 W/m2 at 10 m height for the 2019 and 2020 datasets, respectively. The $Pavg$ is directly proportional to the wind hub height. Investigating Tables 6 and 7 regarding the $Pavg$, it observed that the Rayleigh function with the GOA performed the highest GOF with actual data.

Table 8. *Accuracy measures for 2019 dataset*

Optimization method	Distribution	Accuracy measure				Net Fitness	Rank
		MAE	RMSE	R^2	R		
GOA	Rayleigh	0.1411	0.1781	0.9938	0.9984	0.0817	1
	Weibull	1.7871	2.2525	0.0008	0.9169	1.2805	4
	Inverse Gaussian	1.3589	1.6514	0.4630	0.9953	0.8880	2
	Burr	1.7778	2.2416	0.0105	0.9179	1.2727	3
	Generalized Pareto	1.7914	2.2576	-0.0037	0.1528	1.4750	5
GWO	Rayleigh	0.1411	0.1795	0.9937	0.9984	0.0821	3
	Weibull	0.0695	0.1036	0.9979	0.9990	0.0440	1
	Inverse Gaussian	0.2639	0.4198	0.9653	0.9846	0.1834	5
	Burr	0.0727	0.1056	0.9978	0.9989	0.0454	2
	Generalized Pareto	0.1194	0.2636	0.9863	0.9940	0.1007	4
MFO	Rayleigh	0.1411	0.1795	0.9937	0.9984	0.0821	2
	Weibull	0.0694	0.1040	0.9979	0.9990	0.0441	1
	Inverse Gaussian	0.2638	0.4191	0.9654	0.9846	0.1832	3
	Burr	1.2462	1.6359	0.4730	0.9669	0.8606	4
	Generalized Pareto	1.7794	2.2393	0.0125	0.1804	1.4565	5
SSA	Rayleigh	0.1411	0.1795	0.9937	0.9984	0.0821	1
	Weibull	1.7831	2.2478	0.0050	0.9173	1.2771	4
	Inverse Gaussian	0.2638	0.4191	0.9654	0.9846	0.1832	2
	Burr	1.2315	1.7581	0.3913	0.9690	0.9073	3
	Generalized Pareto	1.7914	2.2575	-0.0036	0.1531	1.4748	5
WOA	Rayleigh	0.1411	0.1795	0.9937	0.9984	0.0821	2
	Weibull	0.0695	0.1037	0.9979	0.9990	0.0441	1
	Inverse Gaussian	0.2657	0.4658	0.9573	0.9833	0.1977	3
	Burr	0.1971	0.5219	0.9464	0.9795	0.1983	4
	Generalized Pareto	1.7812	2.2414	0.0107	0.2062	1.4514	5

EARLY VIEW

On the other hand, in general, Weibull distribution based on GWO, MFO and WOA achieved the best GOF with real data. In all cases, Generalized Pareto distribution offered the worst matching with actual data. Rayleigh and Weibull distributions accomplished the best matching for all the statistical descriptor values.

The distribution function with the best match when the deviation between the actual and the expected wind speeds approaches a minimum value. Tables 8 and 9 summarize the GOF of the presented five distributions employing five optimization methods. For all tables, the bold values refer to the optimal per optimization method, whereas the underlined values refer to the best matching.

Examining Tables 8 and 9, it can be observed that Rayleigh distribution based on GOA and SSA optimization methods outperformed the other distributions and occupied the first rank. However, Weibull distribution based on GWO, MFO and WOA did. Except GWO, Generalized Pareto distribution based on other optimization methods presented least matching to real data and generally ranked fifth place. Although, Burr distribution is mainly based on three parameters, but it was not able to achieve a good matching.

To clearly determine the accuracy of the most fitness distribution, the rank is estimated based on the net fitness measure. Distributions are ranked according to the four GOF criteria as mentioned earlier. As given in Table 10, the top-down rank of the five distributions for the 2019 dataset is Weibull, Rayleigh, Inverse Gaussian, Burr, and Generalized Pareto. However, for the 2020 dataset, the rank is Weibull, Rayleigh, Burr, Inverse Gaussian, and Generalized Pareto.

Table 9. *Accuracy measures for 2020 dataset*

Optimization method	Distribution	Accuracy measure				Net Fitness	Rank
		MAE	RMSE	R^2	R		
GOA	Rayleigh	0.2226	0.2707	0.9882	0.9988	0.1266	1
	Weibull	2.0023	2.5045	-0.0109	0.9328	1.3962	4
	Inverse Gaussian	1.1662	1.3861	0.6904	0.9973	0.7162	2
	Burr	2.0022	2.5044	-0.0108	0.9329	1.3962	3
	Generalized Pareto	2.0061	2.5090	-0.0145	0.2244	1.5763	5
GWO	Rayleigh	0.2226	0.2707	0.9882	0.9988	0.1266	4
	Weibull	0.1102	0.1800	0.9948	0.9978	0.0744	1
	Inverse Gaussian	0.3935	0.6295	0.9361	0.9738	0.2782	5
	Burr	0.1109	0.1811	0.9947	0.9978	0.0748	2
	Generalized Pareto	0.1203	0.3181	0.9837	0.9928	0.1155	3
MFO	Rayleigh	0.2226	0.2707	0.9882	0.9988	0.1266	2
	Weibull	0.1101	0.1784	0.9949	0.9979	0.0739	1
	Inverse Gaussian	0.3934	0.6297	0.9361	0.9738	0.2783	3
	Burr	1.2538	1.7935	0.4816	0.9581	0.9019	4
	Generalized Pareto	1.9164	2.4242	0.0529	0.2869	1.5002	5
SSA	Rayleigh	0.2226	0.2707	0.9882	0.9988	0.1266	2
	Weibull	0.1101	0.1784	0.9949	0.9979	0.0739	1
	Inverse Gaussian	0.4100	0.5569	0.9500	0.9792	0.2594	3
	Burr	1.2445	1.9068	0.4140	0.9594	0.9445	4
	Generalized Pareto	1.7372	2.2075	0.2146	0.4976	1.3081	5
WOA	Rayleigh	0.2226	0.2707	0.9882	0.9988	0.1266	2
	Weibull	0.1101	0.1785	0.9949	0.9979	0.0740	1
	Inverse Gaussian	0.3934	0.6301	0.9360	0.9738	0.2784	4
	Burr	0.2350	0.4011	0.9741	0.9954	0.1667	3
	Generalized Pareto	1.9978	2.4960	-0.0041	0.2497	1.5621	5

Again, as shown in Table 10, Weibull and Rayleigh distributions achieved the first and second ranks, respectively. However, from the computation complexity point of view, the uni-parameter Rayleigh is more satisfactory than the Weibull of bi-parameter. The computation time of Weibull was three to five times of Rayleigh, as can be observed from Table 3. Even so, Generalized Pareto distribution was the next to Weibull in terms of computation time, but it was not able to provide a good matching. According to the obtained results in this study, distributions based on different optimization methods provide different degrees of matching. Therefore, it can be deduced that there is no unique, globally accepted, best optimization method to estimate distribution parameters.

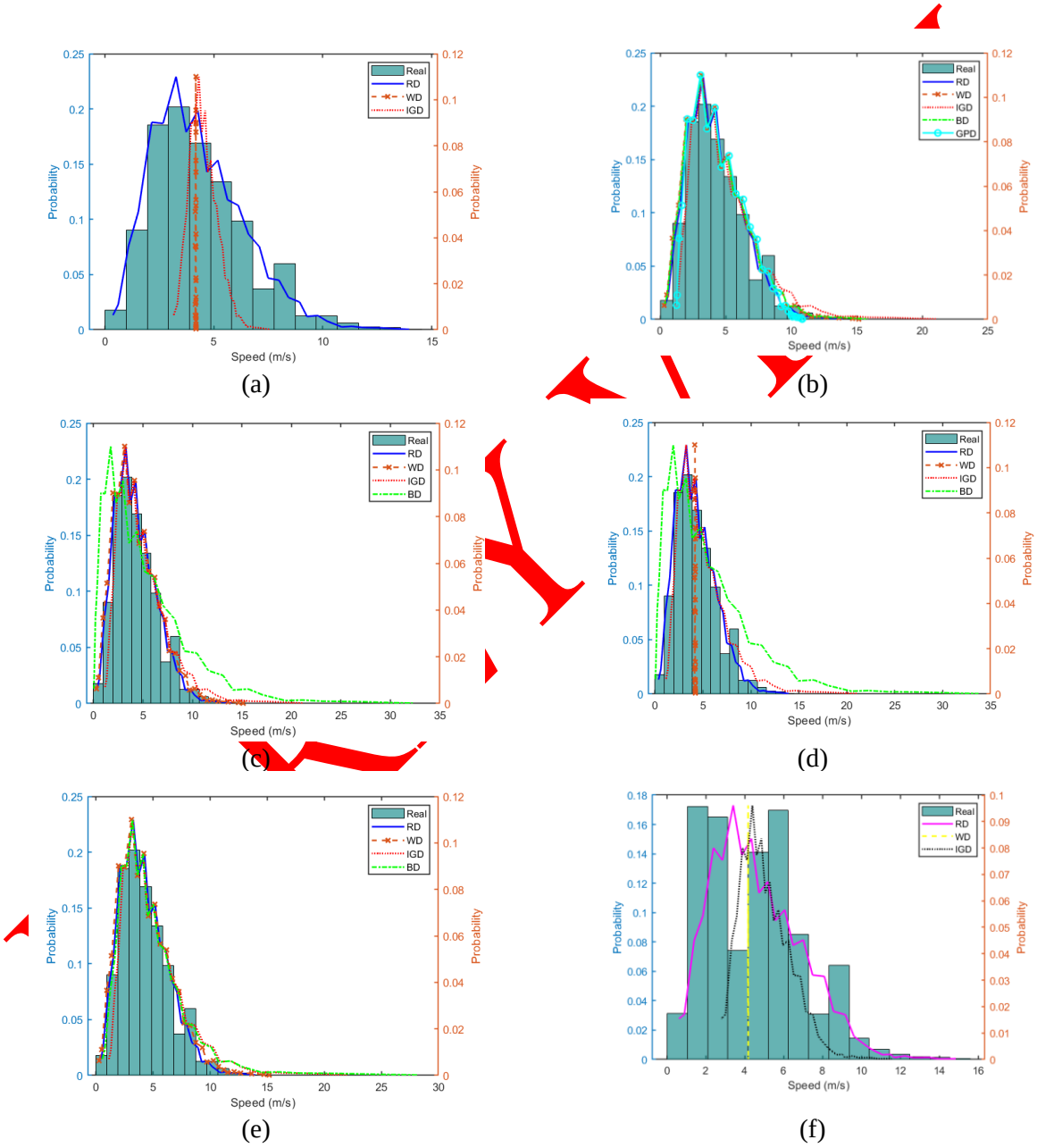
EARLY VIEW

Table 10. *Ranking of distributions*

Distribution	2019					2020				
	1st	2nd	3rd	4th	5th	1st	2nd	3rd	4th	5th
Rayleigh	2	2	1	-	-	1	3	-	1	-
Weibull	3	-	-	2	-	4	-	-	1	-
Inverse Gaussian	-	2	2	-	1	-	1	2	1	1
Burr Type XII	-	1	2	2	-	-	1	2	2	-

Generalized Pareto	-	-	-	1	4	-	-	1	-	4
Best	WD	RD	IGD	BD	GPD	WD	RD	BD	IGD	GPD

To interpret the obtained results visually for all datasets, Figure 2 displays the fitted PDFs. Two different-scale vertical axes are used to represent PDF plots; the left one is for the measured wind data histogram, whereas the right is for the other distributions. These vertical axes represent the probability density. In addition, abbreviations in Figure 2: RD, WD, IGD, BD, and GPD denote Rayleigh, Weibull, Inverse Gaussian, Burr, and Generalized Pareto distributions, respectively. It is noticed that most of the introduced distributions achieved good matching except the Generalized Pareto distribution.



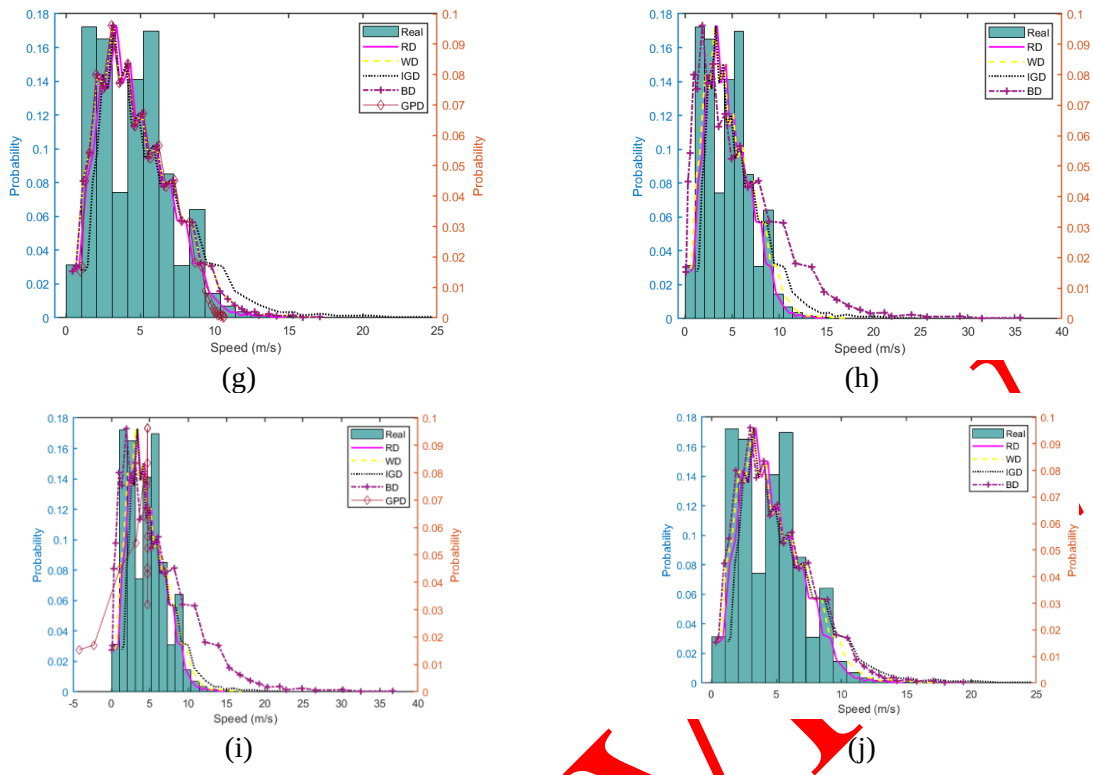


Figure 2. *The PDF curves of used distributions when using (a) GOA-2019, (b) GWO-2019, (c) MFO-2019, (d) SSA-2019, (e) WOA-2019, (f) GOA-2020, (g) GWO-2020, (h) MFO-2020, (i) SSA-2020 and (j) WOA-2020*

Many invaluable conclusions and findings are derived from this comprehensive analysis as follows:

- One of the most critical findings is that the wind regime distribution varies from location to location. Therefore, different distributions should be used to describe the wind distribution pattern accurately. In other words, a particular distribution can accomplish the best GOF at a specific location but not at another.

- The second leading finding is the selection of optimization method which depends on the characteristics of wind data pattern, convergence, and computation complexity. For instance, Rayleigh distribution based on GOA and SSA optimization methods achieved the best matching while Weibull distribution based on GWO, MFO, and WOA did. Consequently, the trade-off between various optimization methods is an indispensable requisite.

- The third crucial finding is the applied error criteria. For example, some error criteria may provide the top matching for a particular distribution but not for another. Consequently, it is indispensable to apply different error measures. Then, the net fitness calculation is essential to determine the best-rank distribution accurately.

- The fourth important finding is parameter numbers per the distribution. In many cases, the distribution functions with higher parameters may better estimate wind patterns. However, the computation burdens significantly increase. For example, based on GWO, the bi-parameter Weibull function provided slightly better matching than the mono-parameter Rayleigh function. Although, the computation time of Weibull distribution parameters selection was about three-five times more than the Rayleigh distribution.

- The fifth important finding is that the scale parameter of some distributions can indicate the wind potentiality at the analyzed site since the scale parameter of these distributions can shape the annual

EARLY

mean wind speed profile. In this study, the Weibull scale parameter can be utilized to predict the annual average wind speed.

- The sixth crucial finding is the skewness and kurtosis statistical descriptors that are also essential to describe the wind regime since they can display the whole wind distribution pattern. In this study, skewness and kurtosis values are positive; therefore, the wind pattern takes the shape of positively skewed and leptokurtic. Thus, the selection of the appropriate distribution can be recognized.

6. CONCLUSION

Selecting convenient distributions for determining the wind speed distribution is crucial for many studies like feasibility studies, wind turbines, and farm design. Therefore, this paper presents five different distributions to describe wind speed distribution. The best parameter values per distribution were assessed employing GOA, GWO, MFO, SSA, and WOA optimization algorithms. The statistical characteristics of the examined location were studied and compared via many statistical descriptors. Besides, the GOF of each distribution based on four different error metrics was measured and compared. In addition, net fitness was computed to determine the distribution with the best matching. The half-hourly recent wind data from the Catalca site in Istanbul were employed to conduct this analysis. Rayleigh distribution based on GOA and SSA outperformed the other distributions. However, Weibull distribution based on GWO, MFO, and WOA did. Besides, Inverse Gaussian and Burr distributions achieved the third and fourth ranks interchangeably for the 2019 and 2020 datasets. Ultimately, the used distribution function, estimation method, and error measure are essential factors in making an accurate decision of the top matching for wind pattern at any site. Finally, this paper provides essential methods to assess the wind potential at any location.

ACKNOWLEDGMENTS

The author would like to thank Ali ORHAN and Rabia KILIC for their efforts in collecting the raw data for this study.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1] Wadi, M., Elmasry, W., "Statistical analysis of wind energy potential using different estimation methods for Weibull parameters: a case study", Electrical Engineering, 103(6): 2573-2594, (2021).

[2] Wadi, M., Elmasry, W., "Different Statistical Distributions and Genetic Algorithms", International Conference on Electric Power Engineering–Palestine (ICEPE)-IEEE, 1–7, (2021).

[3] Shi, J., Erdem, E., "Estimation of wind energy potential and prediction of wind power", In Wind Energy Engineering, Academic Press, 25-49, (2017).

[4] Pishgar, S., Keyhani, A., and Sefeedpari, P., "Wind speed and power density analysis based on Weibull and Rayleigh10distributions (a case study: Firouzkooh county of Iran", Renewable and Sustainable Energy Reviews, 42: 313–322, (2015).

[5] Morgan, E.C., Lackner, M., Vogel R.M., and Baise, L.G., "Probability distributions for offshore wind speeds", Energy Conversion and Management, 52(1): 15–26, (2011).

[6] Crutcher, H.L., Baer, L., "Computations from elliptical wind distribution statistics", Journal of Applied Meteorology and Climatology, 1(4): 522–530, (1962).

EARLY VIEW

[7] Dutta, S., Genton, M.G., "A non-Gaussian multivariate distribution with all lower-dimensional Gaussians and related families", Journal of Multivariate Analysis, 132: 82–93, (2014).

[8] Yuan, K., Zhang, K., Zheng, Y., Li, D., Wang, Y., and Yang, Z., "Irregular distribution of wind power prediction", Journal of Modern Power Systems and Clean Energy, 6(6): 1172–1180, (2018).

[9] Garcia, A., Torres, J., Prieto, E., and De Francisco, A., "Fitting wind speed distributions: a case study", Solar Energy, 2062(2): 139–144, (1998).

[10] Scerri, E., Farrugia, R., "Wind data evaluation in the Maltese Islands", Renewable Energy, 7(1): 109–114, (1996).

[11] Ahsanullah, M., Alzaatreh, A., "Some Characterizations of the Log-Logistic Distribution", Stochastics and Quality Control, 33(1): 23–29, (2018).

[12] Yilmaz, V., Celik, H. E., "A statistical approach to estimate the wind speed distribution: the case of Gelibolu region", Dogus Universitesi Dergisi, 9(1): 122-132, (2011).

[13] Alavi, O., Mohammadi, K., and Mostafaeipour, A., "Evaluating the suitability of wind speed probability distribution models: A case of study of east and southeast parts of Iran", Energy Conversion and Management, 119: 101–108, (2016).

[14] Mert, I., Karakus, C., "A statistical analysis of wind speed data using Burr, generalized gamma, and Weibull distributions in Antakya, Turkey", Turkish Journal of Electrical Engineering & Computer Sciences, 23(6): 1571-1586, (2015).

[15] Rajabi, M., Modarres, R., "Extreme value frequency analysis of wind data from Isfahan, Iran", Journal of Wind Engineering and Industrial Aerodynamics, 96(1): 78–82, (2008).

[16] El-Shanshoury, G.I., Ramadan, A., "Estimation of extreme value analysis of wind speed in the North-Western coast of Egypt", Arab Journal of Nuclear Sciences and Applications, 45: 265–274, (2012).

[17] Nagatsuka, H., Balakrishnan, N., "A method for estimating parameters and quantiles of the three-parameter inverse Gaussian distribution based on statistics invariant to unknown location", Journal of Statistical Computation and Simulation, 84(11): 2361–2377, (2014).

[18] Alayat, M.M., Kassem, Y., and Camur H., "Assessment of wind energy potential as a power generation source: A case study of eight selected locations in Northern Cyprus", Energies, 11(10): 2697, (2018).

[19] Lee, D., Baldick, R., "Probabilistic wind power forecasting based on the laplace distribution and golden search", IEEE/PES Transmission and Distribution Conference and Exposition, 1–5, (2016).

[20] Wallner, M., "A half-normal distribution scheme for generating functions", European Journal of Combinatorics, 187: 103138, (2020).

[21] Gomez, Y.M., Vidal, I., "A generalization of the half-normal distribution", Applied Mathematics-A Journal of Chinese Universities, 31(4), 409-424, (2016).

[22] Ayuketang, A.N., Joseph, E., "Generalized extreme value distribution models for the assessment of seasonal wind energy potential of Debuncha, Cameroon", Journal of Renewable Energy, (2016).

EARLY VIEW

[23] Sarkar, A., Deep, S., Datta, D., Vijaywargiya, A., Roy, R., and Phanikanth, V., "Weibull and Generalized Extreme Value Distributions for Wind Speed Data Analysis of Some Locations in India", KSCE Journal of Civil Engineering, 823(8): 3476–3492, (2019).

[24] Singh, V.P., Guo, H., "Parameter estimation for 3-parameter generalized Pareto distribution by the principle of maximum entropy (POME)", Hydrological Sciences Journal, 40(2): 165–181, (1995).

[25] D'Amico, G., Petroni, F., and Prattico, F., "Wind speed prediction for wind farm applications by extreme value theory and copulas", Journal of Wind Engineering and Industrial Aerodynamics, 145: 229–236, (2015).

[26] Zhang, J., Chen, T., and Xu, L., "Wind power fluctuation characteristics of wind farms", In Atlantis Press, 24: 1478-1481, (2015).

[27] Sohoni, V., Gupta, S., and Nema, R., "A comparative analysis of wind speed probability distributions for wind power assessment of four sites", Turkish Journal of Electrical Engineering & Computer Sciences, 24(6): 4724–4735, (2016).

[28] Wadi, M., Elmasry, W., Shobole, A., Tur, M.R., Bayindir, R., and Shahinzadeh, H., " Wind Energy Potential Approximation with Various Metaheuristic Optimization Techniques Deployment", 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS)-IEEE, 1-6, (2021).

[29] Akinci, T.C., Nogay, H.S., "Wind speed correlation between neighboring measuring stations", Arabian Journal for Science and Engineering, 37(4): 1007-1019, (2012).

[30] Jung, C., Schindler, D., "Global comparison of the goodness-of-fit of wind speed distributions", Energy Conversion and Management, 133: 216–234, (2017).

[31] Saxena, B.K., Rao, K.V.S., "Comparison of Weibull parameters computation methods and analytical estimation of wind turbine capacity factor using polynomial power curve model: case study of a wind farm", Renewables:Wind, Water, and Solar, 2(1): 1–11, (2015).

[32] Pobocikova, I., Sedliackova, Z., and Michalkova, M., "Application of four probability distributions for wind speed modeling", Procedia Engineering, 192: 713-718, (2017).

[33] Drobinski, P., Coulais, C., and Jourdier, B., "Surface wind-speed statistics modelling: Alternatives to the Weibull distribution and performance evaluation", Boundary-Layer Meteorology, 157(1): 97–123, (2015).

[34] Abolpour, B., Abolpour, B., Bakhshi, H., and Yaghobi, M., "An Appropriate Extreme Value Distribution for the Annual Extreme Gust Winds Speed", Journal of Fundamentals of Renewable Energy and Applications, 7(1): 2-4, (2017).

[35] Quan, Y., Wang, F., and Gu, M., "A method for estimation of extreme values of wind pressure on buildings based on the generalized extreme-value theory", Mathematical Problems in Engineering, (2014).

[36] Zhao, X., Zhang, Z., Cheng, W., and Zhang, P., "A new parameter estimator for the Generalized Pareto distribution under the peaks over threshold framework", Mathematics, 7(5): 406, (2019).

EARLY VIEW

[37] Wadi, M., Baysal, M., and Shobole, A., "Reliability and Sensitivity Analysis for Closed-Ring Distribution Power Systems", Electric Power Components and Systems, 49(6-7): 696-714, (2022).

[38] Krishnamoorthy, R., Udhayakumar, K., Raju, K., Elavarasan, R.M., and Mihet-Popa, L., "An Assessment of Onshore and Offshore Wind Energy Potential in India Using Moth Flame Optimization", Energies, 13(12): 1–41, (2020).

[39] Zhang, L., Li, Q., Guo, Y., Yang, Z., and Zhang L., "An investigation of wind direction and speed in a featured wind farm using joint probability distribution methods", Sustainability, 10(12): 4338, (2018).

[40] Ahsanullah, M., Alzaatreh, A., "Parameter estimation for the log-logistic distribution based on order statistics", REVSTAT Statistical Journal, 16(4): 429–443, (2018).

[41] Lin, L., Ang, A.H., Fan, W., and Xia, D., "A probability-based analysis of wind speed distribution and related structural1response in southeast China", Structure and Infrastructure Engineering, 15(1): 14–26, (2019).

[42] Markose, S., Alentorn, A., "The generalized extreme value distribution, implied tail index, and option pricing", The Journal of Derivatives, 18(3): 35–60, (2011).

[43] Kang, S., Song, J., "Parameter and quantile estimation for the generalized Pareto distribution in peaks over threshold framework", Journal of the Korean Statistical Society, 46: 487–501, (2017).

[44] Brabson, B., Palutikof, J., "Tests of the generalized Pareto distribution for predicting extreme wind speeds", Journal of Applied Meteorology, 39(9): 1627–1640, (2000).

[45] Holmes, J., Moriarty, W., "Application of the generalized Pareto distribution to extreme value analysis in wind engineering", Journal of Wind Engineering and Industrial Aerodynamics, 83(1-3): 1–10, (1999).

[46] Steinkohl, C., Davis, R.A., and Kluppelberg, C., "Extreme value analysis of multivariate high-frequency wind speed data", Journal of Statistical Theory and Practice, 7(1): 73-94, (2013).

[47] Ersoz, S., Akinci, T.C., Nogay, H.S., and Dogan, G., "Determination of wind energy potential in Kirklareli-Turkey", International Journal of Green Energy, 10(1): 103-116, (2013).

[48] Sohoni, V., Gupta, S., and Nema, R., "A comparative analysis of wind speed probability distributions for wind power assessment of four sites", Turkish Journal of Electrical Engineering & Computer Sciences, 24(6): 4724–4735, (2016).

[49] Arikan, Y., Arslan, O.P., and Cam, E., "The analysis of wind data with rayleigh distribution and optimum turbine and cost analysis in Elmadag, Turkey", Istanbul University-Journal of Electrical and Electronics Engineering, (2015).

[50] Bidaoui, H., El -Abbassi, I., El-Bouardi, A., and Darcherif, A., "Wind speed data analysis using Weibull and Rayleigh distribution functions, case study: five cities northern Morocco", Procedia Manufacturing, 32: 786–793, (2019).

[51] Maleki-Jebely, F., Zare, K., and Deiri, E., "Efficient estimation of the PDF and the CDF of the inverse Rayleigh distribution", Journal of Statistical Computation and Simulation, 88(1): 75–88, (2018).

EARLY VIEW

[52] Kumar, M.B.H., Balasubramaniyan, S., Padmanaban, S., and Holm-Nielsen, J.B., "Wind Energy Potential Assessment by Weibull Parameter Estimation Using Multiverse Optimization Method: A Case Study of Tirumala Region in India", Energies, 12(11): 2158, (2019).

[53] Justus, C., Hargraves, W., Mikhail, A., and Graber, D., "Methods for estimating wind speed frequency distributions", Journal of Applied Meteorology, 17(3): 350–353, (1978).

[54] Khan, M.S., Pasha, G., and Pasha, A.H., "Theoretical analysis of inverse Weibull distribution", WSEAS Transactions on Mathematics, 7(2): 30–38, (2008).

[55] Dokur, E., Kurban, M., and Ceyhan, S., "Wind speed modelling using inverse weibull distrubition: a case study for Bilecik, Turkey", International Journal of Energy Applications and Technologies, 3(2): 55–59, (2016).

[56] Seshadri, V., "The inverse Gaussian distribution: some properties and characterizations", Canadian Journal of Statistics, 11(2): 131–136, (1983).

[57] Pant, M.D., Headrick, T.C., "A Method for Simulating Burr Type III and Type XII Distributions through Moments and Correlations", International Scholarly Research Notices, (2013).

[58] Tsogt, K., Zandraabal, T., and Lin, C., "Diameter and height distributions of natural even-aged pine forests (Pinus sylvestris) in Western Khentey, Mongolia", Taiwan Journal of Forest Science, 28(1): 29–41, (2013).

[59] Ismail, N.H.B., Khalid, Z.B.M., "EM algorithm in estimating the 2-and 3-parameter Burr Type III distributions", American Institute of Physics, 1605(1): 881–887, (2014).

[60] Kumar, D., "The Burr type XII distribution with some statistical properties", Journal of Data Science, 15(3): 509-533, (2017).

[61] Kim, C., Kim, W., "Estimation of the parameters of burr type III distribution based on dual generalized order statistics", The Scientific World Journal, (2014).

[62] Wadi, M., Baysal, M., and Shobole, A., "Comparison between open-ring and closed-ring grids reliability", 4th International Conference on Electrical and Electronic Engineering (ICEEE), 290-294, (2017).

[63] Van-Montfort M., Witter, J., "The generalized Pareto distribution applied to rainfall depths", Hydrological Sciences Journal, 31(2): 151–162, (1986).

[64] Lenz, R., "Generalized Pareto distributions application to autofocus in automated microscopy", IEEE Journal of Selected Topics in Signal Processing, 10(1): 92–98, (2015).

[65] Brabson, B., Palutikof, J., "Tests of the generalized Pareto distribution for predicting extreme wind speeds", Journal of Applied Meteorology, 39(9): 1627–1640, (2000).

[66] El-Din, M., Sadek, A., and Sharawy, A.M., "Characterization of the generalized Pareto distribution by general progressively Type-II right censored order statistics", Journal of the Egyptian Mathematical Society, 25(4): 369-374, (2017).

[67] Guha, D., Roy, P.K., and Banerjee, S., "Load frequency control of large scale power system using quasi-oppositional grey wolf optimization algorithm", Engineering Science and Technology, an International Journal, 19(4): 1693–1713, (2016).

EARLY VIEW

[68] Saremi, S., Mirjalili, S., and Lewis, A., "Grasshopper optimisation algorithm: theory and application", Advances in Engineering Software, 105: 30–47, (2017).

[69] Neve, A.G., Kakandikar, G.M., and Kulkarni, O., "Application of grasshopper optimization algorithm for constrained and unconstrained test functions", International Journal of Swarm Intelligence and Evolutionary Computation, 186(3): 1–7, (2017).

[70] Ghulanavar, R., Dama, K.K., and Jagadeesh, A., "Diagnosis of faulty gears by modified AlexNet and improved grasshopper optimization algorithm (IGOA)", Journal of Mechanical Science and Technology, 34(10): 4173–4182, (2020).

[71] Mirjalili, S., Mirjalili, S.M., and Lewis, A., "Grey wolf optimizer", Advances in Engineering Software, 69: 46–61, (2014).

[72] Mirjalili, S., "Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm", Knowledge-based Systems, 89: 228–249, (2015).

[73] Talaat, M., Alsayyari, A.S., Farahat, M.A., and Said, T., "Moth-flame algorithm for accurate simulation of a non-uniform electric field in the presence of dielectric barrier", IEEE Access, 7: 3836–3847, (2018).

[74] Mirjalili, S., Gandomi, A.H., Mirjalili, S.Z., Saremi, S., Faris, H., and Mirjalili, S.M., "Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems", Advances in Engineering Software, 114: 163–191, (2017).

[75] Mirjalili, S., Lewis, A., "The whale optimization algorithm", Advances in Engineering Software, 95: 51–67, (2016).

[76] VC, V.R., "Optimal renewable resources placement in distribution networks by combined power loss index and whale optimization algorithms", Journal of Electrical Systems and Information Technology, 5(2): 175–191, (2018).

[77] Wadi, M., Kekezoglu, B., Baysal, M., Tur, M.R., and Shobolec, A., "Feasibility Study of Wind Energy Potential in Turkey: Case Study of Catalca District in Istanbul", 2nd International Conference on Smart Grid and Renewable Energy (SGRE)-IEEE, 1-6, (2019).

[78] Elmasry, W., Wadi, M., "EDLA-EFDS: A Novel Ensemble Deep Learning Approach For Electrical Fault Detection Systems", Electric Power Systems Research, 207: 107834, (2022).

[79] Wadi, M., "Fault detection in power grids based on improved supervised machine learning binary classification", Journal of Electrical Engineering, 72(5), 315-322, (2021).

[80] Wadi, M., Elmasry, W., "An anomaly-based technique for fault detection in power system networks", International Conference on Electric Power Engineering–Palestine (ICEPE-P)-IEEE, 1-6, (2021).

[81] Willmott, C.J., Matsuura, K., "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance", Climate Research, 30(1): 79–82, (2005).

[82] Hyndman, R.J., Koehler, A.B., "Another look at measures of forecast accuracy", International Journal of Forecasting, 4222(4): 679–688, (2006).

EARLY VIEW

[83] Gul, M., Tai, N., Huang, W., Nadeem, M.H., and Yu, M., "Evaluation of Wind Energy Potential Using an Optimum Approach based on Maximum Distance Metric", Sustainability, 12(5): 1999, (2020).

[84] Wadi, M., "Five different distributions and metaheuristics to model wind speed distribution", Journal of Thermal Engineering, 7(14): 1898-1920, (2021).

[85] Elmasry, W., Wadi, M., "Enhanced Anomaly-Based Fault Detection System in Electrical Power Grids", International Transactions on Electrical Energy Systems, (2022).

EARLY VIEW