Stochastic Environmental Research and Risk Assessment (2024) 38:2213–2230
https://doi.org/10.1007/s00477-024-02676-5

ORIGINAL PAPER

Comparison of probability distributions used for harnessing the wind energy potential: a case study from India

Gaurav Kumar Gugliani[1] · Christophe Ley[2] · Najmeh Nakhaei Rad[3,4] · Andriette Bekker[3,4]

Accepted: 26 January 2024 / Published online: 4 March 2024

Abstract
Modeling wind speed data is the prime requirement for harnessing the wind energy potential at a given site. While the Weibull distribution is the most commonly employed distribution in the literature and in practice, numerous scientific articles have proposed various alternative continuous probability distributions to model the wind speed at their convenient sites. Fitting the best distribution model to the data enables the practitioners to estimate the wind power density more accurately, which is required for wind power generation. In this paper we comprehensively review fourteen continuous probability distributions, and investigate their fitting capacities at seventeen locations of India covering the east and west offshore corner as well as the mainland, which represents a large variety of climatological scenarios. A first main finding is that wind speed varies a lot inside India and that one should treat each site individually for optimizing wind power generation. A second finding is that the wide acceptance of the Weibull distribution should at least be questioned, as it struggles to represent wind regimes with heterogeneous data sets exhibiting multimodality, high levels of skewness and/or kurtosis. Our study reveals that mixture distributions are very good alternative candidates that can model difficult shapes and yet do not require too many parameters.

Keywords Mixture distribution · Unimodal distribution · Weibull distribution · Wind energy · Wind speed

1 Introduction

1.1 Historical development of wind speed modeling with probability distributions

Mid 20th Century was an epoch when the world started exploring the wind energy potential. With its emerging demand for power and vulnerability to oil crises, India started its wind energy program in 1983–84. The success of the wind energy program lies in the accuracy of the assessment of wind flow patterns at a potential site. Wind energy is a site-specific and intermittent source of power. Therefore, an extensive wind resources assessment is an essential prerequisite for harnessing the wind power potential at a given site. Wind resource assessment estimates wind flow patterns based on several factors: available wind data, topographical conditions, meteorological conditions, etc. This necessitates the involvement of statistics in evaluating wind flow patterns, which is critical in designing mega-structures and optimizing energy generation from the wind. In statistical terms, the wind flow pattern is not stable in the short term. Nonetheless, it exhibits a consistent and stable pattern over the long term (except for radical and lasting changes

✉ Najmeh Nakhaei Rad
najmeh.nakhaeirad@up.ac.za

Gaurav Kumar Gugliani
gaurav.gugliani007@gmail.com

Christophe Ley
christophe.ley@uni.lu

Andriette Bekker
andriette.bekker@up.ac.za

1 Department of Mechanical Engineering, Mandsaur University, Mandsaur, MP 458001, India

2 Department of Mathematics, University of Luxembourg, 4364 Esch-sur-Alzette, Luxembourg

3 Department of Statistics, University of Pretoria, Pretoria 0002, South Africa

4 DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), Johannesburg, South Africa

due to climate change, which can be spotted by statistical approaches such as change-point detection Aminikhanghahi and Cook (2017)). According to Zhang (2015), wind statistics is a scientific field that examines the wind patterns over a significant duration. Wind data is viewed as a continuous random variable, leading to the use of continuous probability distributions to model the predictable wind pattern. Since the 1940s till today, several papers have been published that used different continuous probability distributions to describe the wind speed, including (Akpinar and Akpinar 2004a, 2009; Jaramillo and Borja 2004; Carta and Ramirez 2007a, b; Carta and Mentado 2007; Gokcek et al. 2007; Vicente 2008; Akdağ et al. 2010; Fyrippis et al. 2010; Safari and Gasore 2010; Chang 2011b; Morgan et al. 2011; Safari 2011; Soukissian 2013; Zhang et al. 2013; Alavi et al. 2016; Hu et al. 2016; Jung et al. 2017; Kantar et al. 2018) and Mohammadi et al. (2017). Distributions with up to 2 parameters were used to model unimodal data, while data exhibiting bimodality have been modeled using multi-parameter distributions, in particular as mixtures of 2-parameter distributions.

Sherlock (1951) recommended utilizing the Pearson type III distribution which is essentially the Gamma distribution, due to its successful and widespread use. The distribution employs two parameters, a scale parameter and a shape parameter and performs well in modeling natural phenomena specificity velocity data.

Luna and Church (1974) used the 2-parameter log-normal distribution for studying air pollution level. The same distribution was implemented by Kaminsky (1977) and Justus (1978) for wind speed analyses, but since then it has only very rarely been considered for fitting this type of data (Tar 2008; Garcia et al. 1998; Bogardi and Matyasovzky 1996).

In the 1970s, the Rayleigh (R) and Weibull (W) distributions entered the scene to model the wind speed (Hennessey Jr 1978). Until the late 1990s, the Weibull distribution proves to be superior to earlier distributions with a low (<= 2) number of parameters (Morgan 1995; Akpinar and Akpinar 2004b, 2005; Pishgar-Komleh et al. 2015; Pishgar-Komleh and Akram 2017). Therefore, it has been a part of widely used computer modeling softwares such as HOMER (Rehman et al. 2007; Van Alphen et al. 2007) and WASP (Hunter and Elliot 1994; Sahin et al. 2005). For instance, the Weibull distribution showed better fitting than Rayleigh (Bidaoui et al. 2019), exponential, square root normal, log-normal and Gamma distributions (Chang 2011b). However, it is well known that the Weibull distribution is not suitable for fitting bimodal data or data with high volume (> 15 % of total wind data set) of low wind speed (meaning 0 m/s) (Carta and Ramírez 2007a, b). Therefore a lot of research efforts have been oriented to find alternatives for and modifications to the Weibull distribution (e.g., Chadee and Sharma (2001); Carta and Mentado (2007); Bali and Theodossiou (2008); Akpinar and Akpinar (2009); Akdağ et al. (2010); Chang (2011b); Chellali et al. (2012); Akgül et al. (2016); Bracale et al. (2017); Aries et al. (2018)). A 3-parameter Weibull distribution with an added location parameter is another suitable alternative to fit wind data of low wind speed (Chalamcharla and Doraiswamy 2016). However, Chadee and Sharma (2001) noted that including the extra location parameter in the estimation process creates challenges, and a positive value for this parameter results in an unrealistic condition of zero probability of wind speeds less than the parameter value. To address high probabilities of zero wind speeds, Carta et al. (2008) proposed using a singly truncated normal (TN) distribution. In cases where wind speed data has infrequent low speeds, Bardsley (1980) recommended the use of the inverse Gaussian distribution as a viable alternative to the 3-parameter Weibull distribution with a positive location parameter. Bivona et al. (2003) fitted all non-zero wind speeds with the Weibull distribution and treated zero wind speeds separately. Table 1 shows comparative studies of the most classical 2-parameter distributions for wind speed modeling.

As there was no universal acceptance of the Weibull distribution (Carta et al. 2009), the search for other distributions intensified, leading to numerous studies, and new distributions were developed. Ouarda et al. (2016) revealed the importance of skewness and kurtosis while modeling the data sets. Some new distributions which were earlier used in other applications were also tested for goodness-of-fit of wind speed data. Soukissian (2013) has introduced the 4-parameter Johnson S_B distribution for wind speed data modeling and compared it with the Weibull distribution. He revealed that indiscriminate use of the Weibull distribution is unjustified and found that the Johnson S_B distribution is a much more suitable model for 11 and 8 buoys of Eastern and Western Mediterranean Sea, respectively.

Various authors have proposed using two-component mixture distributions with different weight proportions to model bimodal wind speed data. Most proposed mixture distributions comprise a Weibull component (Jaramillo and Borja 2004; Carta and Ramírez 2007a;b; Akpinar and Akpinar 2009; Akdağ et al. 2010; Shin et al. 2016) and are typically of the type Weibull–Weibull, truncated normal-Weibull, and Gamma-Weibull. Table 2 provides a summary of articles in which mixture models have been applied for wind speed modeling.

1.2 Review of methods used for parameter estimation in continuous probability distributions

There exist numerous different methods to estimate the parameters of continuous probability distributions, such as for instance the method of moments (MoM), the least square method (LSM), the L-moment method and the maximum likelihood method (ML). Various articles have compared distinct methods in the context of wind data modeling, including (Akdağ and Dinler 2009; Carta et al. 2009; Bagiorgas et al. 2011; Chang 2011a; Morgan et al. 2011; Saleh et al. 2012; Arslan et al. 2014; Azad et al. 2014; De Andrade et al. 2014; Akdağ and Güler 2015; Mohammadi et al. 2016). In the next paragraph we shall briefly point out the pros and cons of these methods, and refer the reader to Gugliani et al. (2018) for a more detailed analysis.

For the LSM, the best estimator is the one that minimizes the sum of squared errors between the observed and the corresponding theoretical values from the distribution. The LSM thus is based on the cumulative distribution function (cdf) of a continuous random variable, function which describes the probability for this random variable to be smaller than a given value. This can lead to complex calculations, in which case it is recommended to solve the equation using a nonlinear technique such as Levenberg–Marquardt (Akdağ et al. 2010).

The MoM is the simplest computational method as it estimates the distribution's parameters using the sample moments. In this way, the parameters are estimated by equating the theoretical moments with the sample moments. However, the method has certain drawbacks (e.g., it can lack robustness), as pointed out in Akdağ and Dinler (2009).

Hosking (1990) proposed the L-moment as another important parameter estimation method. The L-moments are more robust than conventional moments to outliers in the data and enable more secure inferences to be made from small samples about an underlying probability distribution. They are less susceptible to sampling variability, which makes it more suitable for modeling extreme data. Several authors (Gubareva 2011; Murshed et al. 2011; Strupczewski et al. 2011; Papalexiou and Koutsoyiannis 2013; Rutkowska et al. 2017; Ul Hassan et al. 2019; Nerantzaki and Papalexiou 2022) have utilized this method to fit generalized extreme value distributions to rainfall, flood, streamflow data across various parts of the world. Comparative studies of this method with other two methods, viz., MLE and MoM reveals that it is equivalently good as compared to MLE (Rowinski et al. 2002; Gubareva and Gartsman 2010; Hu et al. 2020) and outperforms MoM in parameter estimation (Murshed et al. 2011; Vivekanandan 2015).

The ML method selects those values of parameters that maximize the probability under that distribution of obtaining the randomly observed sample. Suppose $(v_1, v_2, \dots, v_n)$ is the vector of the observations and $\boldsymbol{\theta}$ is the vector of the parameters. The likelihood function is defined as the product of probability density functions (pdfs), which we denote here as f, evaluated at each individual observation

$$L = \prod_{i=1}^{n} f\left(v_i; \boldsymbol{\theta}\right).$$

Subsequently, the log-likelihood function is obtained as

$$\ln L = \sum_{i=1}^{n} \ln f\left(v_i; \boldsymbol{\theta}\right).$$

By setting the partial derivatives of the log-likelihood function with respect to $\boldsymbol{\theta}$ to zero

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln L = 0,$$

the maximum likelihood estimates (MLEs) of the parameters are obtained by solving the system of equations, however solving the likelihood equations can be tricky and require numerical methods (Chang 2011a).

1.3 Model selection criteria

In the statistical literature, there exist various criteria to identify the best-fitting distribution for a given data set. In studies about wind speed data, the most commonly used are the coefficient of determination (R^2), the root mean square error (RMSE) (Akdağ et al. 2010; Aries et al. 2018), the chi-square (χ^2) (Akpinar and Akpinar 2009) and the Kolmogorov–Smirnov (K–S) goodness-of-fit tests (Ayyub and McCuen 2016; Chang 2011a).

In this paper, the K–S goodness-of-fit test has been used to measure the closeness of the fitted cdf with the cumulative relative frequency of the sample wind speed data and to indicate whether or not a distribution is suitable to fit a given data set. The K–S test is defined as the max-error between two cumulative distribution functions

$$Q = \max |F(v) - G(v)|$$

where $F(v)$ is a fitted cdf and $G(v)$ is the cumulative relative frequency of a sample.

However, among all acceptable distributions, they do not tell which one fits best (p-values only serve to reject a distribution for fit or not, but one should not compare p-values among themselves to rank distributions). Such a ranking is provided by information criteria such as the Akaike Information Criterion (AIC) which is based on a compromise between the goodness-of-fit of a distribution in terms of the likelihood function and the number of parameters to estimate, and this compromise is obtained via a penalization on that number. The mathematical expression for AIC is given as

$$AIC = 2N - 2log(L)$$

where L is the likelihood and N is the number of parameters of the model. The smallest AIC value indicates the most suitable distribution. This distribution, however, is not guaranteed to give a reasonable fit (the AIC would also choose the best among bad-fitting distributions). Therefore the best strategy is to first compare distributions via the AIC and then test the best-fitting distribution through the K–S goodness-of-fit test.

1.4 Contribution of the present paper

Given the plethora of different proposals of distributions for modeling wind speed, it is not obvious to know which one to use. As indicated in Tables 1 and 2, comparisons have already been made, but so far no paper has done a really exhaustive comparison taking also mixture components into account. Moreover, the comparison of distributions in the literature is mostly along the coastal line of countries, and often a single distribution performs better than others for all the locations under consideration. However, this factor is not valid for a vast country like India. The main land of India is full of various hills across its geography. These hilly regions also have the capability to harness the wind energy potential. The additional advantage of these hilly regions is that they are not prone to cyclones thanks to their higher topography compared to surrounding land. Once wind turbines have been installed, they can operate at their full capacity producing uninterrupted power supply, without the fear of damage to the power plant. Therefore, in this study, we have taken as many as seventeen locations covering the main land, western coastal region, and eastern coastal region of India, so that the manufacturers can identify the most suitable distributions for their probable site before installing the power plant.

The wind speed data utilized in this study were recorded at a height of 10 m by the Indian Meteorological Department (IMD), Pune, assuming no density variation occurs vertically with height for long slender structures. The vertical change in wind speed with altitude typically follows the power law. Upon estimating the mean wind speed at a height of 10 m using parameters from the most appropriate distribution, the mean wind speed at higher altitudes will be computed using the power law. This computation will enable a judicious selection of the type of wind turbine required for installation at a specific height. Consequently, a comparative evaluation of various distributions is necessary to assess the most suitable distribution for site-specific wind speed data. This evaluation is crucial in determining the appropriate wind turbine for optimal performance at different altitudes.

Following these motivations, in this paper we review and compare various probability distributions that have been suggested over the years by different researchers for wind resources assessment. Moreover, we shall also include some novel distributions that have primarily been used in other domains such as economics and reliability analysis or financial assessment. Many more (too many) probability distributions exist for describing positive data, see Sinner et al. (2022) for an overview. We restrict to those fourteen that we judge most important for modeling wind speed (see Sect. 2), and we estimate their parameters by the ML method, yielding precise estimates with minimal variance. The ML method allows us to use the AIC as criterion to compare the distributions. According to Ley et al. (2021), a good probability distribution should both be versatile, i.e. fit various distinct shapes, and parameter parsimonious, hence ideally not have too many parameters as this complicates interpretation and can lead to overfitting. Therefore, we choose the AIC as a goodness-of-fit criterion and the Kolmogorov–Smirnov test as a goodness-of-fit test. As case study we consider long-term wind speed data of seventeen onshore locations in India, which are described in Sect. 3. Nine sites lie in the seven windy states of India, and eight sites are from the state with zero wind power generation (as per physical report published by the Ministry of New & Renewable Energy). See Fig. 1. The results are presented and discussed in Sect. 4, while Sect. 5 provides a conclusion.

2 Overview of the considered continuous distributions

The distributions relevant for this study are briefly described in what follows. We start from 1- and 2-parameter distributions and end with 4- and 5-parameter distributions.

2.1 Weibull distribution

The 2-parameter Weibull distribution (W) is a classical distribution for wind speed data analysis, in particular unimodal frequency distributions. Originally the Weibull-distribution has 3 parameters, the third being known as the location parameter used for defining the lowest value in a data set. Since for wind speed this corresponds to 0, the location parameter can be dropped (or, say, implicitly equated to 0) and the 3-parameter Weibull distribution simplifies to the 2-parameter Weibull distribution. This 2-parameter Weibull distribution has been extensively employed to estimate the wind power potential or, to be more specific, in the estimation of wind characteristics, see e.g. Bivona et al. (2003); Akpinar and Akpinar (2004a, 2005); Gokcek et al. (2007); Fyrippis et al. (2010); Safari and Gasore (2010); Dursun et al. (2011); Baseer et al. (2017). The pdf and cdf of the 2-parameter Weibull distribution are respectively given by

$$f(v; k, s) = \left(\frac{k}{s}\right)\left(\frac{v}{s}\right)^{k-1} \exp\left[-\left(\frac{v}{s}\right)^{k}\right], v, k, s > 0, \quad (1)$$

and

Fig. 1 The locations of the seventeen considered stations in India (https://www.surveyofindia.gov.in/pages/outline-maps-of-india)

Table 1 Comparative studies of the most classical 2-parameter distributions for wind speed modeling, where * indicates particularly relevant distributions for the considered study

Author(s)	Well-known models for wind speed data					Other possible candidates for wind speed data
Tar (2008)	Normal	Weibull	Rayleigh	Log-normal	Gamma	Square root normal
Zhou et al. (2010)		Weibull	Rayleigh	Log-normal	Gamma	Inverse Gamma, Inverse Gaussian, Erlang
Brano et al. (2011)		Weibull	Rayleigh	Log-normal	Gamma	Inverse Gaussian, Pearson type V and Burr
Zamani and Badri (2015)	Normal	Weibull*	Rayleigh	Log-normal		
Yin (1997)	Normal*	Weibull*	Rayleigh			
Sohoni et al. (2016)		Weibull*	Rayleigh*	Log-normal	Gamma*	Inverse Gaussian
Philippopoulos et al. (2012)		Weibull*	Rayleigh	Log-normal	Gamma*	Inverse Gaussian
de Lima Leite and das Virgens Filho (2011)	Normal	Weibull*	Rayleigh		Gamma	Beta
Safari (2011)	Normal	Weibull*	Rayleigh	Log-normal	Gamma*	
Kiss and Jánosi (2008)		Weibull	Rayleigh	Log-normal	Gamma*	
Amaya-Martínez et al. (2014)		Weibull	Rayleigh	Log-normal	Gamma	

$$F(v;k,s) = 1 - \exp\left[-\left(\frac{v}{s}\right)^k\right] ,$$

where k is the non-dimensional shape parameter and s the scale parameter whose dimension is the same as that of the variable v. For clarification purposes we mention that the variable v stands for the wind speed that we wish to model. Besides a reasonably good fit to wind speed data, there are two further reasons for the popularity of the Weibull distribution: (a) there exist formulas that allow a vertical extrapolation of the wind characteristics (Tar 2008; Safari and Gasore 2010), and (b) it is very practical for calculating

Table 2 Overview of papers that have used mixture models for representing wind speed data, where the number of *'s indicates the most relevant distributions for the considered study

Author(s)	Duration of observation period	Types of probability density function				
Akpinar and Akpinar (2009)	8 years	Weibull	Weibull–Weibull*	Truncated normal-Weibull**		
Vicente (2008)	1 year	Weibull	Weibull–Weibull	Truncated normal-Weibull		
Chang (2011a)	2 years	Weibull	Weibull–Weibull*	Truncated normal-Weibull**	Truncated normal-normal	Gamma-Weibull ***
Akdağ et al. (2010)	1 year (3 hly avg.)	Weibull	Weibull–Weibull**			
Carta and Mentado (2007)	16 years	Weibull	Weibull–Weibull**	Truncated normal-Weibull**		
Jaramillo and Borja (2004)	1 year	Weibull	Weibull–Weibull**			
Rajapaksha and Perera (2016)	1 year	Weibull	Weibull–Weibull**	Log-normal-Weibull *		Gamma-Weibull

the capacity factor, power coefficient, and output power of wind turbines (Jangamshetti and Rau 1999; Dursun et al. 2011; Gugliani et al. 2021).

2.2 Rayleigh distribution

The Rayleigh (R) distribution is a 1-parameter distribution that arises as a special case of the Weibull distribution whose shape parameter is fixed to 2. Consequently, the expression of pdf and cdf of the Rayleigh distribution are given as

$$f(v;s) = \frac{2v}{s^2}\exp\left[-\left(\frac{v}{s}\right)^2\right], v, s > 0,$$

and

$$F(v;s) = 1 - \exp\left[\frac{-v^2}{s^2}\right].$$

2.3 Birnbaum–Saunders distribution

The 2-parameter Birnbaum–Saunders (BS) distribution is known as fatigue life distribution and was promoted in the two papers Birnbaum and Saunders (1969a, b). It has been developed by making a monotonic transformation of the standard normal random variable. The pdf and cdf of the BS distribution are given as

$$f(v;\alpha,\beta) = \frac{1}{2\sqrt{2\pi}\alpha\beta}\left[\left(\frac{\beta}{v}\right)^{1/2} + \left(\frac{\beta}{v}\right)^{3/2}\right] \times \exp\left[-\frac{1}{2\alpha^2}\left(\frac{v}{\beta} + \frac{\beta}{v} - 2\right)\right], v, \alpha, \beta > 0,$$

and

$$F(v;\alpha,\beta) = \Phi\left[\frac{1}{\alpha}\left\{\left(\frac{v}{\beta}\right)^{1/2} - \left(\frac{\beta}{v}\right)^{1/2}\right\}\right],$$

where β is a scale parameter, α is a shape parameter and Φ (.) is the standard normal cdf.

2.4 Gamma distribution

The Gamma (G) distribution is a 2-parameter distribution whose curve drops off much more gradually than that of the exponential distribution for shape parameters $\zeta > 1$ and more quickly for $\zeta < 1$. The pdf and cdf are

$$f(v;\zeta,\beta) = \frac{v^{\zeta-1}}{\beta^\zeta \Gamma(\zeta)}\exp\left[-\frac{v}{\beta}\right], v, \beta, \zeta > 0, \tag{2}$$

and

$$F(v;\zeta,\beta) = \int_0^v \frac{x^{\zeta-1}}{\beta^\zeta\Gamma(\zeta)}\exp\left[-\frac{x}{\beta}\right] dx,$$

where ζ and β are the shape and scale parameters, respectively, and $\Gamma(.)$ is the gamma function. The chi-squared distribution is a special case of the Gamma corresponding to $\beta = 2$ and $\zeta = k/2$ for some positive integer k.

2.5 Nakagami distribution

The 2-parameter Nakagami (Na) distribution is strongly related to the G distribution (Nakagami 1960) and it is extensively used in communication engineering (Pajala et al. 2006; Beaulieu and Cheng 2005). Suppose V has the

G distribution in (2), then $\sqrt{V/\zeta}$ follows the Na distribution. The pdf and cdf of the Nakagami distribution are given as

$$f(v;\zeta,\beta)=\frac{2\zeta^{\zeta}}{\beta^{\zeta}\Gamma(\zeta)}v^{2\zeta-1}\exp\left(-\frac{\zeta}{\beta}v^{2}\right),\zeta>1/2,\ v,\beta>0,$$

and

$$F(v;\zeta,\beta)=\frac{P\left(\zeta,\frac{\zeta}{\beta}v^{2}\right)}{\Gamma(\zeta)},$$

where P and Γ are the upper incomplete gamma and gamma functions, respectively.

2.6 Log-normal distribution

If a random variable V follows the log-normal (LN) distribution, then $Y=\ln V$ has the normal distribution. The expressions for the pdf and cdf of the log-normal distribution are

$$f(v;\mu,\sigma)=\frac{1}{v\sigma\sqrt{2\pi}}\exp\left[-\frac{(\ln v\ -\mu)^{2}}{2\sigma^{2}}\right],v,\sigma>0,\mu\in\mathbb{R},$$

and

$$F(v;\mu,\sigma)=\frac{1}{2}+\frac{1}{2}erf\left[\frac{\ln v-\mu}{\sqrt{2}\sigma}\right].$$

2.7 Truncated normal distribution

If the support of the density of a normal random variable Y is truncated on the left at zero, the resulting truncated random variable $V>0$ follows the truncated normal (TN) distribution with pdf and cdf

$$f(v;\mu,\sigma)=\frac{1}{\Phi(\mu/\sigma)\sigma\sqrt{2\pi}}\exp\left[-\frac{(v-\mu)^{2}}{2\sigma^{2}}\right],v,\mu,\sigma>0,\tag{3}$$

and

$$F(v;\mu,\sigma)=\frac{1}{\Phi(\mu/\sigma)}\left\{\frac{1}{2}+\frac{1}{2}erf\left[\frac{v-\mu}{\sqrt{2}\sigma}\right]-\Phi(-\mu/\sigma)\right\}.$$

2.8 Inverse Gaussian distribution (Wald distribution)

The inverse Gaussian (IG) is a skewed, 2-parameter distribution which is similar to the Gamma distribution with greater skewness and a sharper peak. The name is misleading in the sense that an IG random variable is not obtained as inverse of a normal random variable, but it is related to two distinct quantities of Brownian motions that the IG and normal describe. This distribution is suitable to fit unimodal and positively skewed data sets. The pdf of the IG is given as

$$f(v;\mu,\lambda)=\left(\frac{\lambda}{2\pi v^{3}}\right)^{\frac{1}{2}}\exp\left[\frac{-\lambda(v-\mu)^{2}}{2\mu^{2}v}\right],v>0,$$

where μ 0 is the mean and λ 0 is the shape parameter. The cdf can be expressed in terms of the standard normal distribution function $\Phi(.)$ by

$$\begin{aligned}F(v;\mu,\lambda)\ =&\Phi\left(\left(\frac{\lambda}{v}\right)^{\frac{1}{2}}\left(-1+\frac{v}{\mu}\right)\right)\\&+\exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\left(\frac{\lambda}{v}\right)^{\frac{1}{2}}\left(1+\frac{v}{\mu}\right)\right).\end{aligned}$$

The IG has the property that if a random variable V follows the inverse Gaussian distribution with parameters μ and λ, then a scalar multiple cV with $c>0$ follows the same distribution with parameters $c\mu$ and $c\lambda$, respectively.

2.9 Johnson S_B distribution

The Johnson S_B (JSB) distribution is one of the Johnson distributions (Johnson 1949) and remarkably flexible. This flexibility is owed to the presence of 4 parameters; as of now, we move indeed from 2-parameter distributions to distributions with 4 or more parameters. The pdf and cdf of the JSB distribution are given by

$$\begin{aligned}f(v;\gamma,\delta,\xi,\lambda)=&\frac{\delta\lambda}{\sqrt{2\pi}(v-\xi)(\xi+\lambda-v)}\\&\times\exp\left\{-\frac{1}{2}\left[\gamma+\delta\ln\left(\frac{v-\xi}{\xi+\lambda-v}\right)\right]^{2}\right\},\end{aligned}$$

and

$$F(v;\gamma,\delta,\xi,\lambda)=\Phi\left(\gamma+\delta\ln\left(\frac{v-\xi}{\xi+\lambda-v}\right)\right),$$

where $\xi\le v\le\xi+\lambda$, ξ is the location parameter, $\lambda>0$ is the scale parameter and γ and $\delta>0$ are shape parameters. The JSB distribution actually also has been derived from a normal distribution. Indeed, if a random variable V follows the JSB distribution, then $Z=\gamma+\delta\ln\left(\frac{Y}{1-Y}\right)$ with $Y=(V-\xi)/\lambda$ follows the standard normal distribution.

2.10 Generalized beta distribution of the second kind

The generalized beta distribution of the second kind (GB2) introduced by McDonald and Xu (1995) is a 4-parameter flexible distribution which is mostly applied as a size distribution of income in economics. The pdf of the GB2 is given by

$$f(v;a,b,p,q) = \frac{av^{ap-1}}{b^{ap}B(p,q)\left(1+\left(\frac{v}{b}\right)^{a}\right)^{p+q}},\ v>0,$$

where $a,p,q>0$ are shape parameters, $b>0$ is a scale parameter and $B(p,q)=\frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ is the beta function. The cdf of the GB2 is

$$F(v;a,b,p,q) = 1 - I_{\left(1+\left(\frac{v}{b}\right)^{a}\right)^{-1}}(p,q), \tag{4}$$

where $I_x(p,q)=\frac{B_x(p,q)}{B(p,q)}$ is the incomplete beta function.

2.11 Generalized hyperbolic distribution

The generalized hyperbolic (GH) distribution is a 5-parameter distribution introduced by Barndorff-Nielsen (1978) and contains numerous well-known special cases such as variance-gamma, Laplace, hyperbolic, Student's t, Cauchy, normal inverse Gaussian and normal distributions. It can model skew as well as light- and heavy-tailed data. Through a location-scale transformation, the pdf of the GH distribution is given as

$$\begin{aligned} f(v;\lambda,\alpha,\beta,\mu,\sigma) \\ = \frac{\sqrt{\alpha}\left(1-\beta^2\right)^{\frac{\lambda}{2}-\frac{1}{4}}}{\sqrt{2\pi}\delta\sigma K_\lambda(\alpha)}\left[\left(\frac{v-\mu}{\delta\sigma}\right)^2+1\right]^{\frac{\lambda}{2}-\frac{1}{4}} \\ \times K_{\lambda-1/2}\left(\frac{\alpha}{\sqrt{1-\beta^2}}\sqrt{\left(\frac{v-\mu}{\delta\sigma}\right)^2+1}\right) \\ \exp\left[\frac{\alpha\beta}{\sqrt{1-\beta^2}}\left(\frac{v-\mu}{\delta\sigma}\right)\right], v>0, \end{aligned}$$

with

$$\delta = \left[\frac{K_{\lambda+1}(\alpha)}{\alpha K_\lambda(\alpha)} + \frac{\beta^2}{1-\beta^2}\left(\frac{K_{\lambda+2}(\alpha)}{K_\lambda(\alpha)} - \frac{K_{\lambda+1}(\alpha)^2}{K_\lambda(\alpha)^2}\right)\right]^{-\frac{1}{2}},$$

and

$$\mu = m - \delta\sigma\frac{\beta}{\sqrt{1-\beta^2}}\frac{K_{\lambda+1}(\alpha)}{K_\lambda(\alpha)},$$

where m and δ are the mean and variance of the distribution, respectively, where $K_\lambda(.)$ denotes the modified Bessel function of the third kind with order $\lambda \in \mathbb{R}$, λ and $\alpha>0$ are shape parameters, $|\beta|<1$ is a skewness parameter, $\sigma>0$ a scale parameter and $\mu \in \mathbb{R}$ a location parameter.

2.12 Mixture distributions

Consider a finite set of pdfs $f_1(v),...,f_k(v)$ and weights $w_1,...,w_k$ such that $w_i \geq 0$ and $\sum_{i=1}^k w_i = 1$. A mixture distribution $f(v)$ is then represented by

$$f(v) = \sum_{i=1}^{k} w_i f_i(v).$$

Mixture distributions are useful for modeling heterogeneous wind data, see, e.g., Jaramillo and Borja (2004); Carta and Ramírez (2007a, b); Akpinar and Akpinar (2009); Akdağ et al. (2010); Qin et al. (2009, 2012) or Alonzo et al. (2017). The following mixture distributions are investigated in this paper as two-component mixture models ($k=2$ in (4)).

2.12.1 Weibull–Weibull distribution

The Weibull–Weibull distribution (WW) consists of two Weibull components with different weight proportions. Jaramillo and Borja (2004) used this distribution for the first time for wind speed data analysis of La Ventosa, Mexico, while Akdağ et al. (2010) analyzed the wind speed data of nine buoys located in the Ionian and Aegean Sea (Eastern Mediterranean) with the WW distribution and compared it with the conventional Weibull distribution.

2.12.2 Truncated normal-Weibull distribution

The truncated normal-Weibull distribution (TNW) is based on the truncated normal (3) and the Weibull (1) distributions. Carta and Ramírez (2007a, b) analyzed the wind speed data of 16 locations of the Canarian Archipelago that comprised both unimodal and bimodal distributions with the TNW, WW and W distributions.

2.12.3 Truncated normal-Gamma distribution

The truncated normal-Gamma distribution (TNG) is a mixture of the truncated normal (3) and the Gamma (2) distributions. Gugliani et al. (2017) found this distribution to model best the wind speed data at the Trivandrum site in India.

3 Data description

We have considered long-term wind speed data from the Indian Meteorology Department Pune, IMD, that has a significant number of weather stations across India. Dyne pressure tube anemometer is the instrument employed by IMDs to record wind speed data. It is located at the height of 10 m above the mean ground level at a position utterly free from obstructions to the airflow. Typically, these observatories are established near airports to take advantage of open terrain. In this paper, wind speed data of seventeen stations, namely Bangalore, Dolphin Nose, Amritsa, Palam, New Delhi, Jaipur, Lucknow, Allahabad, Gaya, New Kandla, Ahmedabad, Bhopal, Indore, Jamshedpur, Calcutta, Hyderabad and Tuticorin, have been considered for the case study (see Fig. 1). Table 3 provides some information about the geographical coordinates of stations and the wind speed observations for each station. In this study, the impact of null wind speed has been checked and found to be occurring in less than 15% of the cases. This is considered to be marginally low (Takle and Brown 1978; Razika and Marouane 2014), therefore any null values have been removed from the hourly data series.

Table 4 shows the statistical description of wind speed data for the seventeen considered locations in India. From Table 4 it has been revealed that New Kandla and Calcutta have maximum wind speed. The two stations are located in India's western and eastern offshore and near the Arabian Sea and Bay of Bengal, respectively. The New Kandla and Indore are two stations showing mean and median wind speed well above the cut-in (2–3 m/s) wind speed of wind turbines at 10 m height, followed by Tuticorin. These sites are therefore the most probable sites for installing a wind farm. Tuticorin has the smallest skewness (in absolute value), whereas Bangalore has the highest among all stations. The Dolphin Nose exhibits negative skewness, however this might be associated with the fact that we have less than 10,000 observations at that station. The kurtosis of all stations is higher than 3 except for Indore

Table 3 Geographical coordinates, number of wind speed observations, observation period, and missing years for the selected reference stations in India (IMD, Pune)

Station	Altitude (m)	Latitude	Longitude	Number of observations	Period	Missing observations
Bangalore	915	13.20°	77.71°	249,666	1969–2000	1969[Feb, Aug]
Dolphin Nose	1814	10.2091°	77.4872°	8604	1985	Nil
Amritsar	234	31.6340°	74.8723°	219,124	1969–2006	1977[Jul–Aug]; 1986 [Jun–Dec]; 1987; 1988; 1994; 1995[Apr]; 2000[May–Aug]; 2004[Mar]
Palam(A)	237	28.57°	77.10°	204,527	1969–2006	1974[Jan]; 1980[Feb]; 1981[May]; 1990
New Delhi	216	28.6139°	77.2090°	211,877	1969–2006	1983;1987; 1988
Jaipur	432	26.92°	75.82°	162,057	1969–2002	1971[Jun]; 1978[Sep]; 1979[Jan]; 1980[May]; 1982[Dec]; 1983[Jan–May]; 1988[Dec];1990 [Sep]; 1991[Aug–Dec]; 1993[Jun-Aug];1996[Jul–Dec]; 1998[Apr–Oct]
Lucknow	123	26.8467°	80.9462°	119,167	1969–2000	1973[Oct, Dec]; 1974[Mar, Oct–Nov]; 1979[Jun–Oct]; 1981; 1988–1997; 1998[Jan–Mar, Nov–Dec]; 1999[Apr]
Allahabad	98	25.4358°	81.8463°	13,770	1969–1971[Jan–Mar]	Nil
Gaya	111	24.7914°	85.0002°	40,024	1969–1977	1970; 1974
New Kandla	550	23.0134°	70.2144°	168,960	1969–1991	1984[Mar];1985[Jun, July]; 1991[Nov–Dec]
Ahmedabad	53	23.0225°	72.5714°	145,834	1969–1990	1979[Nov–Dec]; 1981; 1990[Jan]; 1990[Nov–Dec]
Bhopal	527	23.2599°	77.4126°	153,008	1969–1990	1973[Feb–Mar]; 1982[Apr]; 1990[Oct–Dec]
Indore	550	22.7196°	75.8577°	142,968	1971–1998	1972[Jan–May]; 1978[Feb]; 1981[Oct–Dec];1982; 1983; 1984; 1998[May]
Jamshedpur	159	22.8046°	86.2029°	13,338	1969–1974	1970; 1971[Jan, Oct–Dec]; 1972
Calcutta	9.14	22.5726°	88.3639°	147,608	1969–1994	1983[Dec]; 1986[Jul]; 1987[Sep–Dec];1988–1990; 1991[Jan]
Hyderabad(A)	536.0	17.37°	78.48°	194,000	1969–1998	1975[Apr–Jun]; 1981; 1985; 1986[Jan–Mar]; 1988[Jan–Apr]; 1995[Jan–Apr]; 1998[Jan–Mar]; 1998 [Nov–Dec]
Tuticorin	4	8.7642°	78.1348°	138,384	1989–2006	1994[May]; 1997[Oct–Dec]; 1998[Jan–Jun]

Table 4 Statistical description of wind speed data for the seventeen considered locations in India

Station	Max	Mean	Median	Stdev	CV	Skewness	Kurtosis
Bangalore	25.0000	2.5313	2.2222	1.4057	0.5553	3.0054	42.8016
Dolphin Nose	27.5000	22.3506	23.3333	3.7080	0.1659	− 0.9933	3.4919
Amritsar	25.5556	2.4323	1.9444	1.9864	0.8167	1.9076	9.7166
Palam(A)	21.9444	3.0424	2.7778	1.8959	0.6232	1.1682	5.1370
New Delhi	22.2222	2.4216	2.2222	1.7467	0.7213	1.4089	6.0431
Jaipur	19.4444	1.9491	1.6667	1.5288	0.7844	1.7038	7.6952
Lucknow	21.1111	2.9433	2.7778	2.0429	0.6941	1.0686	4.4142
Allahabad	20.0000	1.5419	1.1111	1.3986	0.9071	1.8711	9.1666
Gaya	23.3333	2.6510	2.2222	2.0748	0.7827	1.5752	6.4443
New Kandla	26.6667	5.6891	5.2778	2.9643	0.5210	0.6518	3.3330
Ahmedabad	19.4444	2.8650	2.7778	1.6243	0.5669	0.8490	4.1811
Bhopal	22.2222	3.8370	3.3333	2.2075	0.5753	0.7520	3.6700
Indore	19.4444	5.1389	5.2778	2.5710	0.5003	0.1884	2.7585
Jamshedpur	14.4444	2.1224	1.6667	1.4953	0.7045	1.5621	7.1206
Calcutta	26.6667	2.2426	1.6667	1.8895	0.8425	1.6286	7.3651
Hyderabad(A)	20.2778	2.7399	2.2222	1.9527	0.7127	1.2808	5.1310
Tuticorin	22.7778	4.6705	4.4444	2.4467	0.5239	0.3718	3.0436

which is a land lock fastest growing city. The high kurtosis value reveals that all the stations' wind speed histogram is leptokurtic except for Indore. Furthermore, the CV is maximum for Allahabad, followed by Calcutta and Amritsar, and least for Indore. A high CV value means a wide variation in wind speed from the mean wind speed.

4 Result and discussions

The ML method was used to estimate the parameters of the fourteen distributions analyzed in the seventeen locations in India. To compare the performance of different models, we used the AIC as a goodness-of-fit criterion and the Kolmogorov–Smirnov (K–S) test as a goodness-of-fit test. Tables 6 and 7 in the Appendix contain the estimated values of the parameters for the fourteen distributions in the seventeen locations, while Table 8 in the Appendix shows the AIC values. The distribution with minimum AIC is the most suitable distribution for the given data set. Among 5-parameter distributions, the truncated normal-Gamma distribution outperforms all other distributions at four locations, followed by the Weibull–Weibull distribution at three stations, the generalized hyperbolic distribution at two locations and the truncated normal-Weibull at one location. At four locations the 4-parameter generalized beta distribution of second kind has the best performance as a wind speed model. At three locations have 2-parameter distributions come out as most suitable, twice the Birnbaum–Saunders distribution and once the Gamma distribution. If we compare only 2-parameter distributions among themselves, then the Gamma is judged the most suitable six times, the Birnbaum–Saunders four times, the Weibull distribution three times, and the Nakagami and the truncated normal distributions respectively two times. Note however that the Weibull is sometimes only beaten by a very small margin, in particular by the Nakagami distribution. Nevertheless, these findings reveal the interesting fact that one should not blindly use the Weibull distribution out of convenience, as better choices are definitely available, even among 2-parameter distributions. For each station, the best fitted model along with the corresponding AIC, the p-value of the K–S test, the coefficient of determination R^2 and RMSE are summarized in Table 5. We see that, at 5% level, only 4 best-fitting distributions would be rejected, while none would be rejected at 3% level, showing, especially at such a high number of observations, that the selected distributions are very suitable models for the data under investigation.

For visual inspection, we provide the wind speed histograms and empirical cdfs along with pdfs and cdfs of the best fitted models for four locations in India, namely Bangalore, Hyderabad, Jaipur and New Kandla, see Figs. 2 and 3. The plots for other distributions and stations are of course available upon request, as we did not want to render the paper unnecessarily long. Note that we chose as class width for the bins 2 km/h following the recent suggestions by Deep et al. (2020), who illustrated the appropriateness of a 2 km/h class width for removing the sampling error. As general conclusion, we find that, for highly skewed and leptokurtic histograms, distributions with more than 2 parameters are more suitable, which explains why multiparameter or mixture models are such good choices. The reader is referred to Gugliani et al. (2018) to calculate the wind power density by knowing the pdf of different distributions.

Table 5 The best fitted model with the corresponding *p*-value of the K–S test, the AIC, R^2 and the RMSE for the seventeen considered locations in India

Station	Best model	*P*-value of K–S	AIC	R^2	RMSE
Bangalore	Weibull–Weibull	0.0720	813588.3628	0.9952	0.0218
Dolphin Nose	Weibull–Weibull	0.0347	43623.1022	0.9990	0.0102
Amritsar	Generalized beta II	0.0709	801932.6258	0.9938	0.0165
Palam(A)	Generalized beta II	0.0700	787954.6508	0.9955	0.0185
New Delhi	Generalized hyperbolic	0.0746	746002.4678	0.9935	0.0194
Jaipur	Birnbaum–Saunders	0.0837	504797.2540	0.9938	0.0168
Lucknow	Truncated normal-Gamma	0.0535	470688.2468	0.9962	0.0163
Allahabad	Birnbaum–Saunders	0.1402	36991.8347	0.9757	0.0309
Gaya	Generalized hyperbolic	0.0735	149394.3200	0.9933	0.0207
New Kandla	Truncated normal-Weibull	0.0336	829182.6197	0.9983	0.0139
Ahmedabad	Truncated normal-Gamma	0.0513	531873.6712	0.9970	0.0161
Bhopal	Generalized beta II	0.0457	653209.2921	0.9976	0.0153
Indore	Truncated normal-Gamma	0.0516	662655.7425	0.9983	0.0149
Jamshedpur	Generalized beta II	0.0657	42964.1920	0.9917	0.0242
Calcutta	Weibull–Weibull	0.0697	516581.4881	0.9905	0.198
Hyderabad(A)	Gamma	0.0419	737440.2027	0.9964	0.0152
Tuticorin	Truncated normal-Gamma	0.0557	628521.3462	0.9973	0.0180

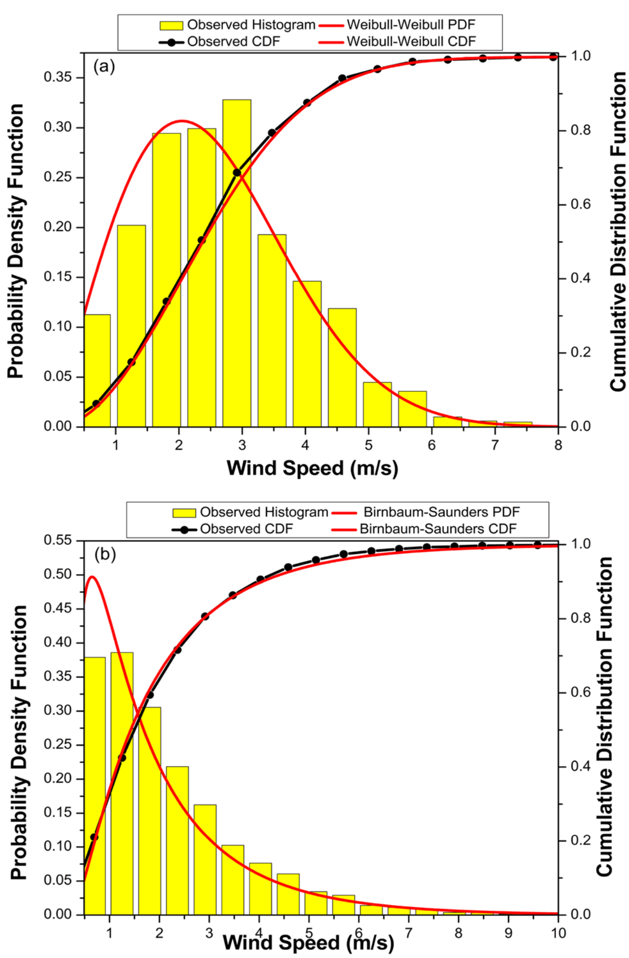

Fig. 2 Wind speed histograms and empirical cdfs at **a** Banglore and **b** Jaipur stations along with the pdfs and cdfs of the best fitted distributions

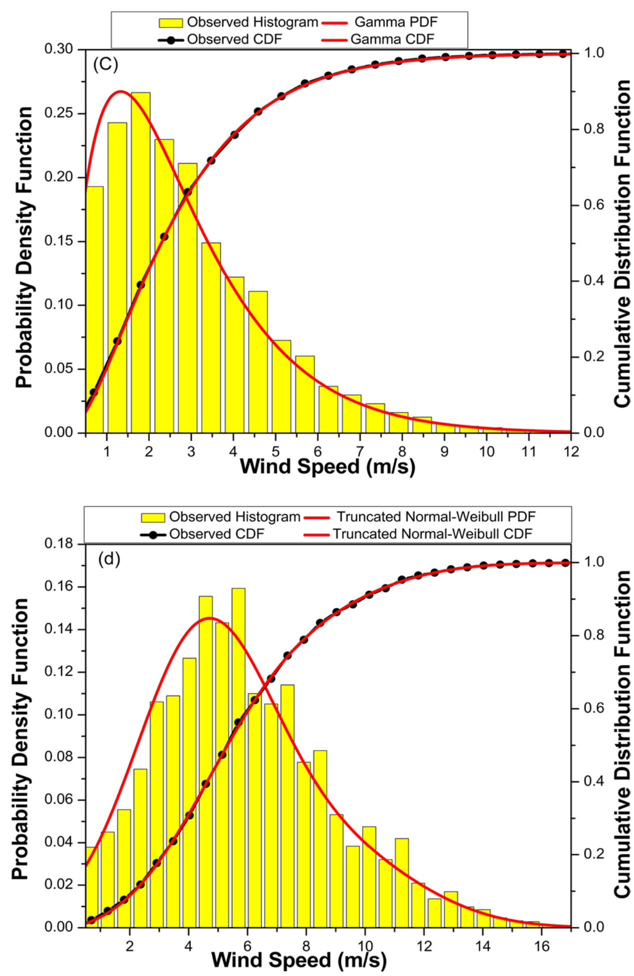

Fig. 3 Wind speed histograms and empirical cdfs at **c** Hyderabad and **d** New Kandla stations along with the pdfs and cdfs of the best fitted distributions

5 Conclusions

Fourteen continuous probability distributions have been reviewed and compared for modeling wind speed data at seventeen locations in India covering the east and west offshore corner and the mainland of India, hence a large variety of distinct climatological situations. Our aim was to identify the site-specific best distribution that can model the wind speed data with minimum amount of parameters and maximum agreement with the given wind data set. The Maximum Likelihood method has been used to estimate the parameters of the distributions. We determined the most suitable distribution by means of the AIC and checked the suitability with the Kolmogorov–Smirnov test. We found that out of the seventeen locations, four wind speed sites have been best modeled by the truncated normal-Gamma distribution, four by the generalized beta distribution of second kind, three by the Weibull–Weibull distribution, two respectively by the generalized hyperbolic and the Birnbaum–Saunders distributions, and one respectively by the truncated normal-Weibull distribution and Gamma distribution.

Our study reveals two main important messages, namely (i) that wind speed varies quite a lot within India from one location to another and that one should treat each geographic situation individually for best wind power generation, and (ii) that the wide acceptance of the Weibull distribution should at least be questioned, as it cannot perfectly represent all the wind regimes for wind speed modeling, especially wind regimes with heterogeneous data sets exhibiting multimodality, high levels of skewness and/or kurtosis. Instead, more suitable probability distributions such as those presented in this paper should be selected for each wind regime to minimize errors in the estimation of the wind energy potential at a given site. Our study shows that mixture distributions are very good candidates.

Appendix

See Tables 6, 7 and 8.

Table 6 Estimated parameters of 1, 2 and 4-parameter distributions for fitting wind speed data of the seventeen considered locations in India

Station	Distribution							
	Truncated normal		Birnbaum–Saunders		Nakagami		Weibull	
	μ	σ	α	β	β	ζ	k	s
Bangalore	2.2577	1.6336	2.0565	0.6705	0.9696	8.3835	1.8756	2.8472
Dolphin Nose	22.3506	3.7078	21.9489	0.1912	8.5307	513.2981	8.0714	23.8169
Amritsar	0.0000	3.1403	1.6277	0.9780	0.5246	9.8618	1.2931	2.6398
Palam(A)	2.3080	2.4143	2.3547	0.7545	0.7880	12.8505	1.6848	3.4175
New Delhi	0.7392	2.6691	1.7737	0.8451	0.6344	8.9150	1.4611	2.6849
Jaipur	0.0000	2.4771	1.4233	0.8549	0.5792	6.1361	1.3708	2.1441
Lucknow	1.4081	2.9482	2.0637	0.9051	0.6343	12.8365	1.4749	3.2581
Allahabad	0.0000	2.0817	1.0361	0.9871	0.4619	4.3333	1.1798	1.6392
Gaya	0.0000	3.3664	1.9202	0.8667	0.5767	11.3324	1.3671	2.9142
New Kandla	5.2966	3.3197	4.4571	0.7280	0.9917	41.1528	1.9953	6.4119
Ahmedabad	2.5072	1.914	2.2414	0.7342	0.8821	10.8467	1.8353	3.2265
Bhopal	3.3063	2.6286	2.9012	0.7865	0.8402	19.5961	1.7922	4.3118
Indore	4.8738	2.8235	3.6549	0.8660	0.9388	33.0182	2.0204	5.7568
Jamshedpur	0.8860	2.2045	1.6111	0.7903	0.6741	6.7401	1.5134	2.3656
Calcutta	0.0000	2.9324	1.4924	0.9916	0.4982	8.5992	1.2440	2.4141
Hyderabad(A)	0.9984	2.9298	1.9624	0.8761	0.6326	11.3198	1.4632	3.0350
Tuticorin	4.3359	2.7477	3.3947	0.8387	0.9028	27.8001	1.9371	5.2390

Table 6 (continued)

Station	Distribution							
	Gamma		Log-normal		Inverse Gaussian		Rayleigh	
	ζ	β	μ	σ	μ	λ	s	–
Bangalore	2.0474	2.5313	2.0565	0.6705	2.5313	5.0624	2.0474	–
Dolphin Nose	30.8345	0.7249	3.0906	0.1894	2.2350	6.0591	16.0203	–
Amritsar	1.5940	1.5260	0.5435	0.8979	2.4323	2.0527	2.2206	–
Palam(A)	2.4982	1.2178	0.8993	0.7028	3.0423	4.6789	2.5348	–
New Delhi	1.9782	1.2241	0.6109	0.7878	2.4215	2.8771	2.1113	–
Jaipur	1.8360	1.0616	0.3710	0.8007	1.9490	2.2551	1.7516	–
Lucknow	1.8948	1.5533	0.7930	0.8330	2.9433	2.9833	2.5334	–
Allahabad	1.4059	1.0968	0.0370	0.9220	1.5418	1.2724	1.4720	–
Gaya	1.8160	1.4598	0.6750	0.8084	2.6509	2.9713	2.3804	–
New Kandla	3.0766	1.8491	1.5673	0.6574	5.6891	9.4816	4.5361	–
Ahmedabad	2.7481	1.0426	0.8597	0.6821	2.8650	4.6843	2.3288	–
Bhopal	2.5583	1.4999	1.1367	0.7200	3.8370	5.3735	3.1302	–
Indore	2.6796	1.9178	1.4388	0.7536	5.1388	5.77730	4.0631	–
Jamshedpur	2.1588	0.9831	0.5034	0.7422	2.1223	2.9395	1.8358	–
Calcutta	1.4977	1.4974	0.4380	0.9162	2.2425	1.8309	2.0736	–
Hyderabad(A)	1.9405	1.4119	0.7286	0.8084	2.7398	2.9953	2.3791	–
Tuticorin	2.6251	1.7792	1.3389	0.7449	4.6705	5.6509	3.7283	–

Station	Distribution							
	Generalized Beta II				Johnson S_B			
	a	b	p	q	γ	δ	ζ	λ
Bangalore	3.7625	4.1064	0.4905	2.0080	−0.4817	1.0512	1.9605	0.7653
Dolphin Nose	49.4137	28.7742	0.1048	28.6291	−1.2419	0.9250	6.6210	20.9908
Amritsar	0.8400	132.4121	2.2148	67.2178	6.7738	1.7230	−0.8434	145.6430
Palam(A)	1.2351	41.8468	1.7928	44.3631	4.4346	1.9830	−1.1158	39.4417
New Delhi	0.8285	182.6785	2.8238	106.0801	3.7796	1.6330	−0.7107	30.5787
Jaipur	0.4572	198.3910	8.6925	78.2316	4.1441	1.5261	−0.4790	33.1396
Lucknow	1.3562	140.6683	1.1550	197.9257	2.5926	1.5614	−1.0319	22.4283
Allahabad	2.0000	1.9597	0.6723	1.4318	5.4423	1.6262	−0.6929	55.7250
Gaya	0.5043	196.9846	7.0499	67.5031	2.8172	1.2791	−0.3697	25.0001
New Kandla	2.6223	15.5207	0.7117	7.1381	2.0972	1.8848	−2.1481	30.1947
Ahmedabad	2.1978	9.1743	0.8254	8.4510	6.6518	2.8724	−2.2402	54.3772
Bhopal	2.3070	14.3330	0.7142	10.8763	2.9476	2.1029	−2.1407	28.7593
Indore	6.4400	9.1562	0.2237	1.7759	0.7988	2.2667	−5.2861	25.0254
Jamshedpur	0.7043	196.5487	4.2289	108.6955	5.0759	1.7628	−0.5813	44.6038
Calcutta	0.6371	198.6996	3.4756	66.3805	4.4871	1.6289	−0.9481	45.9953
Hyderabad(A)	1.0142	175.3738	1.9175	130.7861	2.6212	1.3963	−0.5423	21.4290
Tuticorin	5.9816	7.8998	0.2403	1.4153	2.3516	2.8293	−5.5568	33.1669

Table 7 Estimated parameters of 5-parameter distributions for fitting wind speed data of the seventeen considered locations in India

Station	Generalised hyperbolic					Truncated normal-Weibull				
	λ	α	β	μ	σ	w	μ	σ	k	s
Bangalore	1.2446	6.6285	0.9567	−1.0300	1.4057	0.9853	2.3655	1.3771	1.1479	5.1158
Dolphin Nose	1.8948	0.3559	−0.9031	27.2632	3.7080	0.0000	23.2546	2.2524	8.0713	23.8169
Amritsar	0.0928	1.9661	0.9836	−0.3975	1.9864	0.2500	1.6516	1.4083	1.2433	2.7795
Palam(A)	−4.2264	5.5674	0.9888	−1.5814	1.8959	0.0937	1.8245	0.6906	1.7052	3.5571
New Delhi	1.6351	0.0013	0.9769	0.2143	1.7467	0.0812	0.8378	0.3395	1.5340	2.8538
Jaipur	−2.5356	2.3982	0.9929	−0.4963	1.5288	0.0000	1.7617	0.8055	1.3708	2.1440
Lucknow	1.3463	2.1043	0.9807	−0.3884	2.0429	0.0111	7.2333	1.2394	1.4844	3.2077
Allahabad	−1.6029	1.7223	0.9917	−0.2766	1.3986	0.7747	0.0172	2.3481	3.0206	0.4274
Gaya	1.1307	0.6896	0.9825	2.6510	2.0748	0.0000	1.8394	2.1486	1.3671	2.9142
New Kandla	−2.6095	14.7953	0.9735	−5.3929	2.9643	0.6714	4.3522	2.2688	2.9797	9.0926
Ahmedabad	−3.0641	9.5302	0.9770	−2.0437	1.6243	0.0727	3.1442	0.9480	1.7770	3.1984
Bhopal	−3.5972	10.6223	0.9794	−3.2435	2.2075	0.0000	3.0986	2.4391	1.7921	4.3116
Indore	−19.4973	122.5250	0.8802	−19.5578	2.5710	0.3202	0.0072	4.6793	2.9370	6.5021
Jamshedpur	0.2934	2.2948	0.9862	−0.1974	1.4953	0.1275	0.9273	0.3668	1.5954	2.5693
Calcutta	0.1391	1.4805	0.9835	−0.1396	1.8895	0.2868	3.1385	2.8565	1.4160	1.7576
Hyderabad(A)	0.4000	2.8510	0.9834	−0.6120	1.9527	0.0000	3.1019	0.4274	1.4632	3.0349
Tuticorin	−3.0369	38.0386	0.9428	−9.5116	2.4467	0.6147	5.2218	2.0907	1.3762	4.0849

Station	Truncated normal-Gamma					Weibull–Weibull				
	w	μ	σ	ζ	β	w	k_1	s_1	k_2	s_2
Bangalore	0.4555	2.7290	1.0909	2.4708	0.9511	0.9989	2.0500	2.8390	1.7634	18.2708
Dolphin Nose	1.0000	22.3506	3.7078	37.7793	0.5858	0.4795	6.9374	20.8927	22.3658	25.5638
Amritsar	0.2500	1.8051	1.5999	1.4672	1.7113	0.1328	1.4117	4.7063	1.3961	2.3619
Palam(A)	0.2500	3.1427	1.5786	2.2660	1.3155	0.2963	1.8698	4.6887	1.8557	2.8996
New Delhi	0.0459	5.1124	1.9114	2.0499	1.1177	0.5956	1.6745	3.4322	1.7821	1.6666
Jaipur	0.1791	3.4814	2.1916	2.2367	0.6963	0.6440	1.6012	2.7880	2.0348	1.0931
Lucknow	0.1806	4.1958	2.0160	1.8348	1.4423	0.7500	1.7590	3.8986	1.5863	1.5094
Allahabad	0.7500	0.3156	2.2461	8.8020	0.0415	0.7504	1.4654	2.1433	3.0109	0.4210
Gaya	0.2581	4.4137	2.7650	2.3908	0.8048	0.6354	1.6281	3.8457	2.0781	1.4533
New Kandla	0.7086	5.4298	3.7751	7.7367	0.6349	0.0978	4.8101	5.4981	1.9303	6.4854
Ahmedabad	0.2819	3.3949	1.2849	2.3771	1.1152	0.0911	3.3671	3.4158	1.7680	3.1994
Bhopal	0.4222	4.3542	2.0889	2.1875	1.5490	0.1931	2.5412	4.9998	1.6849	4.1378
Indore	0.9667	5.1455	2.6213	7.4603	0.0565	0.7930	2.8027	6.4693	1.1547	2.8993
Jamshedpur	0.0000	2.1949	1.4921	2.1588	0.9831	0.5097	1.6909	3.0985	1.8565	1.6477
Calcutta	0.1770	4.0113	2.5621	1.6815	1.0667	0.8836	1.3937	2.7315	3.4941	0.4898
Hyderabad(A)	0.0001	2.7871	2.5710	1.9404	1.4120	0.6438	1.6500	83.7450	1.7102	1.8499
Tuticorin	0.7461	5.0721	2.2627	1.4969	2.1864	0.6675	2.7681	5.95667	1.2886	3.6757

Table 8 Akaike Information Criterion for the seventeen considered locations in India

Distribution	Station					
	Bangalore	Dolphin Nose	Amritsar	Palam (A)	New Delhi	Hyderabad
Birnbaum–Saunders	867462.8257	49056.4520	809072.9251	809175.7264	757834.2378	752187.7312
Gamma	825698.4175	48196.3165	802639.4347	788064.9237	750750.0918	**737,440.2027**
Log-normal	855480.8665	48970.8078	812849.6314	804055.1376	759050.2037	750752.9408
Nakagami	833094.7251	47567.3274	819219.2102	793751.4765	763173.6840	745282.2252
Rayleigh	833248.0428	59492.7883	899309.3597	802111.0504	798236.9877	777852.2713
Weibull	831012.6281	45547.9689	805888.9356	790678.0358	755218.0241	739666.8554
Johnson S_B	875117.2253	43824.7393	816020.1684	792044.4383	758904.3357	742095.3042
Generalized hyperbolic	824095.0570	44014.5137	809436.0020	792993.6692	**746,002.4678**	742236.4743
Truncated normal	852268.4818	46973.2211	819585.4821	805494.7569	770247.4078	751140.5971
Truncated normal-Weibull	820463.3320	45553.9689	805884.2600	789802.4769	753720.2714	739672.8562
Truncated normal-Gamma	817541.3528	46977.2211	803780.2500	788540.3090	750486.8344	737446.2428
Weibull–Weibull	**813,588.3628**	**43,623.1022**	803324.5836	788604.7413	751881.86901	737978.0019
Inverse Gaussian	876,137,9849	49070.9294	821525.6134	816647.7999	765656.3139	761785.5063
Generalized Beta II	815136.9880	43782.1787	**801,932.6258**	**787,954.6508**	750268.3496	737495.4553
Distribution	Station					
	Jaipur	Lucknow	Allahabad	Gaya	New Kandla	Tuticorin
Birnbaum–Saunders	**504,797.2540**	484024.6219	**36,991.8347**	150012.8082	884660.2061	696490.7826
Gamma	510563.0593	471674.0255	38580.3750	150954.5513	837420.2050	647134.7176
Log-normal	508080.8444	483672.6167	37861.4243	150589.5639	867374.4984	681753.1808
Nakagami	526838.9362	472724.0906	40117.8566	154706.0530	830205.7092	633660.9918
Rayleigh	567232.7572	492466.2746	47817.9481	164862.0374	830211.3917	634636.4780
Weibull	516178.7388	471197.7258	38865.3866	152236.9689	830211.8370	634417.6390
Johnson S_B	515918.6415	477144.9511	40472.8809	151389.1250	831257.7389	635885.4345
Generalized hyperbolic	513639.1558	475046.5118	39074.3392	**149,394.3200**	832360.5380	636534.9547
Truncated normal	529246.0974	475591.4879	40183.8055	155267.0731	833848.5356	629534.2326
Truncated normal-Weibull	516184.7396	471151.5693	37927.2839	152242.9693	**829,182.6197**	628547.5488
Truncated normal-Gamma	508932.4421	**470,688.2468**	37434.1950	150341.8156	829574.7824	**628,521.3462**
Weibull–Weibull	509205.5512	470889.0170	37032.6018	150271.4093	829270.3620	628669.2364
Inverse Gaussian	508459.2512	491474.3856	37290.3299	151048.9151	893883.7201	708899.2724
Generalized Beta II	506666.3564	471128.6091	37867.4427	150124.8767	830016.4575	628842.6855
Distribution	Station					
	Ahmedabad	Bhopal	Indore	Jamshedpur	Calcutta	–
Birnbaum–Saunders	558434.9063	689872.6837	756800.9405	43265.2282	516787.9889	–
Gamma	534739.9602	658092.2441	693841.5655	43052.0866	520603.1664	–
Log-normal	553029.7126	681565.5179	736239.9503	43329.2990	522358.3866	–
Nakagami	532474.4352	653311.7492	675900.8426	44051.2534	531871.7872	–
Rayleigh	534038.8171	656552.5576	676277.1299	45659.1584	596485.3175	–
Weibull	532157.8093	653375.7550	676257.1174	43494.1219	522885.2775	–
Johnson S_B	536254.0812	657798.8377	673605.2002	43375.7415	536539.5082	–
Generalized hyperbolic	536066.2566	658835.4402	675300.2866	43066.3941	522623.7059	–
Truncated normal	538350.3674	657432.9365	667041.6426	44694.8338	531873.1733	–
Truncated normal-Weibull	532000.2240	653381.7628	665662.1289	43322.0542	521941.4965	–
Truncated normal-Gamma	**531,873.6712**	653700.0014	**662,655.7425**	43058.0870	519829.2662	–
Weibull–Weibull	531991.6100	653226.6144	665260.5735	43201.1833	**516,581.4881**	–
Inverse Gaussian	564801.9432	69716.5465	772071.8416	43616.5492	523589.0754	–
Generalized Beta II	531919.6096	**653,209.2921**	666987.1595	**42,964.1920**	519042.8030	–

The bold values show the minimum AIC values, indicating the distributions with the best fit

Acknowledgements The first author is thankful to Prof. Arnab Sarkar from IIT BHU for motivating him to conduct research in climatology. We would like to thank the Indian Institute of Technology for providing the facilities, the Indian Meteorological Department (IMD), Pune, for supplying the data used in this research, and the Bhabha Atomic Research Center (BARC), Mumbai, for providing the necessary funds to carry out the reported research work. However, the opinions expressed in this manuscript are those of the authors and not of these agencies. The research of the third and fourth authors is supported in part by the RDP grant at university of Pretoria, the National Research Foundation (NRF) of South Africa, Ref.: RA210106581084, grant No. 150170; ref. SRUG2204203865, RA171022270376 (Grant No: 119109), the South African DST-NRF-MRC SARChI Research Chair in Biostatistics (Grant No. 114613) and DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), South Africa. The opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the CoE-MaSS or the NRF.

Author contributions The authors have contributed equally to this work.

Funding Open access funding provided by University of Pretoria. Board of Research in Nuclear Sciences (Project Code: M-21-114) The authors are grateful to the Indian Meteorological Department, Pune, for the supply of wind data to carry out this research work and BRNS for providing funding to get these data. We are also thankful to our respective organizations, for motivating us, to carry out research in the field of climatology.

Data availability The data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to the private policy by the Indian Meteorological Department.

Code availability The code is available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

Ethical approval The authors have declared that the research work carried out in this manuscript is the original work of the authors and no data have been collected without the consent of the competent authority.

Consent to participate All the authors have willingly participated in this research work.

Consent for publication All the authors have given their consent for submission and subsequent publication of the manuscript in Nature Energy.

References

Akdağ SA, Dinler A (2009) A new method to estimate Weibull parameters for wind energy applications. Energy Convers Manag 50(7):1761–1766

Akdağ SA, Güler Ö (2015) A novel energy pattern factor method for wind speed distribution parameter estimation. Energy Convers Manag 106:1124–1133

Akdağ S, Bagiorgas H, Mihalakakou G (2010) Use of two-component Weibull mixtures in the analysis of wind speed in the Eastern Mediterranean. Appl Energy 87(8):2566–2573

Akgül FG, Şenoğlu B, Arslan T (2016) An alternative distribution to Weibull for modeling the wind speed data: inverse Weibull distribution. Energy Convers Manag 114:234–240

Akpinar E, Akpinar S (2004a) Determination of the wind energy potential for Maden, Turkey. Energy Convers Manag 45(18–19):2901–14

Akpinar E, Akpinar S (2004b) Statistical analysis of wind energy potential on the basis of the Weibull and Rayleigh distributions for Agin-Elazig, Turkey. Proc Inst Mech Eng Part A J Power Energy 218(8):557–565

Akpinar EK, Akpinar S (2005) An assessment on seasonal analysis of wind energy characteristics and wind turbine characteristics. Energy Convers Manag 46(11–12):1848–1867

Akpinar S, Akpinar EK (2009) Estimation of wind energy potential using finite mixture distribution models. Energy Convers Manag 50(4):877–884

Alavi O, Mohammadi K, Mostafaeipour A (2016) Evaluating the suitability of wind speed probability distribution models: a case of study of east and southeast parts of Iran. Energy Convers Manag 119:101–108

Alonzo B, Ringkjob HK, Jourdier B et al (2017) Modelling the variability of the wind energy resource on monthly and seasonal timescales. Renew Energy 113:1434–1446

Amaya-Martínez PA, Saavedra-Montes AJ, Arango-Zuluaga EI (2014) A statistical analysis of wind speed distribution models in the Aburrá Valley, Colombia. CT &F-Ciencia, Tecnología y Futuro 5(5):121–136

Aminikhanghahi S, Cook DJ (2017) A survey of methods for time series change point detection. Knowl Inf Syst 51(2):339–367

Aries N, Boudia SM, Ounis H (2018) Deep assessment of wind speed distribution models: a case study of four sites in Algeria. Energy Convers Manag 155:78–90

Arslan T, Bulut YM, Yavuz AA (2014) Comparative study of numerical methods for determining Weibull parameters for wind energy potential. Renew Sustain Energy Rev 40:820–825

Ayyub BM, McCuen RH (2016) Probability, statistics, and reliability for engineers and scientists. CRC Press

Azad AK, Rasul MG, Yusaf T (2014) Statistical diagnosis of the best Weibull methods for wind power assessment for agricultural applications. Energies 7(5):3056–3085

Bagiorgas HS, Giouli M, Rehman S et al (2011) Weibull parameters estimation using four different methods and most energy-carrying wind speed analysis. Int J Green Energy 8(5):529–554

Bali TG, Theodossiou P (2008) Risk measurement performance of alternative distribution functions. J Risk Insur 75(2):411–437

Bardsley W (1980) Note on the use of the inverse Gaussian distribution for wind energy applications. J Appl Meteorol Climatol 19(9):1126–1130

Barndorff-Nielsen O (1978) Hyperbolic distributions and distributions on hyperbolae. Scand J Stat 5(3):151–157

Baseer MA, Meyer JP, Rehman S et al (2017) Wind power characteristics of seven data collection sites in Jubail, Saudi Arabia using Weibull parameters. Renew Energy 102:35–49

Beaulieu NC, Cheng C (2005) Efficient Nakagami-m fading channel simulation. IEEE Trans Veh Technol 54(2):413–424

Bidaoui H, El Abbassi I, El Bouardi A et al (2019) Wind speed data analysis using Weibull and Rayleigh distribution functions, case study: five cities northern Morocco. Procedia Manuf 32:786–793

Birnbaum ZW, Saunders SC (1969a) Estimation for a family of life distributions with applications to fatigue. J Appl Probab 6(2):328–347

Birnbaum ZW, Saunders SC (1969b) A new family of life distributions. J Appl Probab 6(2):319–327

Bivona S, Burlon R, Leone C (2003) Hourly wind velocity analysis in Sicily. Renew Energy 28(9):1371–1385

Bogardi I, Matyasovzky I (1996) Estimating daily wind speed under climate change. Sol Energy 57(3):239–248

Bracale A, Carpinelli G, De Falco P (2017) A new finite mixture distribution and its expectation-maximization procedure for extreme wind speed characterization. Renew Energy 113:1366–1377

Brano VL, Orioli A, Ciulla G et al (2011) Quality of wind speed fitting distributions for the urban area of Palermo, Italy. Renew Energy 36(3):1026–1039

Carta JA, Mentado D (2007) A continuous bivariate model for wind power density and wind turbine energy output estimations. Energy Convers Manag 48(2):420–432

Carta J, Ramirez P (2007a) Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions. Renew Energy 32(3):518–531

Carta JA, Ramírez P (2007b) Use of finite mixture distribution models in the analysis of wind energy in the Canarian Archipelago. Energy Convers Manag 48(1):281–291

Carta JA, Ramírez P, Velázquez S (2008) Influence of the level of fit of a density probability function to wind-speed data on the WECS mean power output estimation. Energy Convers Manag 49(10):2647–2655

Carta JA, Ramirez P, Velazquez S (2009) A review of wind speed probability distributions used in wind energy analysis: case studies in the Canary Islands. Renew Sustain Energy Rev 13(5):933–955

Chadee JC, Sharma C (2001) Wind speed distributions: a new catalogue of defined models. Wind Eng 25(6):319–337

Chalamcharla SC, Doraiswamy ID (2016) Mathematical modeling of wind power estimation using multiple parameter Weibull distribution. Wind Struct 23(4):351–366

Chang TP (2011a) Estimation of wind energy potential using different probability density functions. Appl Energy 88(5):1848–1856

Chang TP (2011b) Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application. Appl Energy 88(1):272–282

Chellali F, Khellaf A, Belouchrani A et al (2012) A comparison between wind speed distributions derived from the maximum entropy principle and Weibull distribution. Case of study; six regions of Algeria. Renew Sustain Energy Rev 16(1):379–385

De Andrade CF, Neto HFM, Rocha PAC et al (2014) An efficiency comparison of numerical methods for determining Weibull parameters for wind energy applications: a new approach applied to the northeast region of Brazil. Energy Convers Manag 86:801–808

de Lima Leite M, das Virgens Filho JS (2011) Adjustment of models of probability distribution to hourly wind speed series for Ponta Grossa, Paraná State. Acta Sci Technol 33(4):447

Deep S, Sarkar A, Ghawat M et al (2020) Estimation of the wind energy potential for coastal locations in India using the Weibull model. Renew Energy 161:319–339

Dursun B, Alboyaci B, Gokcol C (2011) Optimal wind-hydro solution for the Marmara region of Turkey to meet electricity demand. Energy 36(2):864–872

Fyrippis I, Axaopoulos PJ, Panayiotou G (2010) Wind energy potential assessment in Naxos Island, Greece. Appl Energy 87(2):577–586

Garcia A, Torres J, Prieto E et al (1998) Fitting wind speed distributions: a case study. Solor Energy 62(2):139–144

Gokcek M, Bayulken A, Bekdemir K (2007) Investigation of wind characteristics and wind energy potential in Kirklareli, Turkey. Renew Energy 32(10):1739–1752

Gubareva TS (2011) Types of probability distributions in the evaluation of extreme floods. Water Resour 38:962–971

Gubareva TS, Gartsman BI (2010) Estimating distribution parameters of extreme hydrometeorological characteristics by L-moments method. Water Resour 37:437–445

Gugliani G, Sarkar A, Mandal S et al (2017) Location wise comparison of mixture distributions for assessment of wind power potential: a parametric study. Int J Green Energy 14(9):737–753

Gugliani G, Sarkar A, Ley C et al (2018) New methods to assess wind resources in terms of wind speed, load, power and direction. Renew Energy 129:168–182

Gugliani GK, Sarkar A, Ley C et al (2021) Identification of optimum wind turbine parameters for varying wind climates using a novel month-based turbine performance index. Renew Energy 171:902–914

Hennessey JP Jr (1978) A comparison of the Weibull and Rayleigh distributions for estimating wind power potential. Wind Eng 2:156–164

Hosking JR (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. J R Stat Soc Ser B Stat Methodol 52(1):105–124

Hu Q, Wang Y, Xie Z et al (2016) On estimating uncertainty of wind energy with mixture of distributions. Energy 112:935–962

Hu L, Nikolopoulos EI, Marra F, Anagnostou EN (2020) Sensitivity of flood frequency analysis to data record, statistical model, and parameter estimation methods: an evaluation over the contiguous United States. J Flood Risk Manag 13(1):e12580

Hunter R, Elliot G (1994) Wind-diesel systems: a guide to the technology and its implementation. Cambridge University Press

Jangamshetti SH, Rau VG (1999) Site matching of wind turbine generators: a case study. IEEE Trans Energy Convers 14(4):1537–1543

Jaramillo O, Borja M (2004) Wind speed analysis in La Ventosa, Mexico: a bimodal probability distribution case. Renew Energy 29(10):1613–1630

Johnson NL (1949) Systems of frequency curves generated by methods of translation. Biometrika 36(1/2):149–176

Jung C, Schindler D, Laible J et al (2017) Introducing a system of wind speed distributions for modeling properties of wind speed regimes around the world. Energy Convers Manag 144:181–192

Justus CG (1978) Winds and wind system performance. Research supported by the National Science Foundation and Energy Research and Development Administration Philadelphia

Kaminsky F (1977) Four probability densities/log-normal, gamma, Weibull, and Rayleigh/and their application to modelling average hourly wind speed. In: international solar energy society, annual meeting, 19_6–19_10

Kantar YM, Usta I, Arik I et al (2018) Wind speed analysis using the extended generalized Lindley distribution. Renew Energy 118:1024–1030

Kiss P, Jánosi IM (2008) Comprehensive empirical analysis of ERA-40 surface wind speed distribution over Europe. Energy Convers Manag 49(8):2142–2151

Ley C, Babić S, Craens D (2021) Flexible models for complex data with applications. Annu Rev Stat Appl 8:369–391

Luna R, Church H (1974) Estimation of long-term concentrations using a "universal'' wind speed distribution. J Appl Meteorol Climatol 13(8):910–916

McDonald JB, Xu YJ (1995) A generalization of the beta distribution with applications. J Econom 66:133–152
Mohammadi K, Alavi O, Mostafaeipour A et al (2016) Assessing different parameters estimation methods of Weibull distribution to compute wind power density. Energy Convers Manag 108:322–335
Mohammadi K, Alavi O, McGowan JG (2017) Use of Birnbaum–Saunders distribution for estimating wind speed and wind power probability distributions: a review. Energy Convers Manag 143:109–122
Morgan VT (1995) Statistical distributions of wind parameters at Sydney, Australia. Renew Energy 6(1):39–47
Morgan EC, Lackner M, Vogel RM et al (2011) Probability distributions for offshore wind speeds. Energy Convers Manag 52(1):15–26
Murshed MS, Kim S, Park JS (2011) Beta-K distribution and its application to hydrologic events. Stoch Environ Res Risk Assess 25:897–911
Nakagami M (1960) The m-distribution–a general formula of intensity distribution of rapid fading. Statistical methods in radio wave propagation. Elsevier, pp 3–36
Nerantzaki SD, Papalexiou SM (2022) Assessing extremes in hydroclimatology: a review on probabilistic methods. J Hydrol 605:127302
Ouarda TB, Charron C, Chebana F (2016) Review of criteria for the selection of probability distributions for wind speed data and introduction of the moment and l-moment ratio diagram methods, with a case study. Energy Convers Manag 124:247–265
Pajala E, Isotalo T, Lakhzouri A et al. (2006) An improved simulation model for Nakagami-m fading channels for satellite positioning applications. In: 3rd workshop on position. Navigation and communication, Hannover, Germany, pp 81–89
Papalexiou SM, Koutsoyiannis D (2013) Battle of extreme value distributions: a global survey on extreme daily rainfall. Water Resour Res 49(1):187–201
Philippopoulos K, Deligiorgi D, Karvounis G (2012) Wind speed distribution modeling in the greater area of Chania, Greece. Int J Green Energy 9(2):174–193
Pishgar-Komleh S, Akram A (2017) Evaluation of wind energy potential for different turbine models based on the wind speed data of Zabol region, Iran. Sustain Energy Technol Assess 22:34–40
Pishgar-Komleh S, Keyhani A, Sefeedpari P (2015) Wind speed and power density analysis based on Weibull and Rayleigh distributions (a case study: Firouzkooh county of Iran). Renew Sustain Energy Rev 42:313–322
Qin X, Zhang J, Yan X (2009) A finite mixture three-parameter Weibull model for the analysis of wind speed data. In: 2009 international conference on computational intelligence and software engineering, pp 1–3
Qin X, Zhang JS, Yan Xd (2012) Two improved mixture Weibull models for the analysis of wind speed data. J Appl Meteorol Climatol 51(7):1321–1332
Rajapaksha K, Perera K (2016) Wind speed analysis and energy calculation based on mixture distributions in Narakkalliya, Sri Lanka. J Natl Sci Found Sri Lanka 44(4):409
Razika NII, Marouane M (2014) Comparison between hybrid Weibull and MEP methods for calculating wind speed distribution. In: IEEE (ed) 2014 5th international renewable energy congress (IREC), pp 1–6
Rehman S, El-Amin I, Ahmad F et al (2007) Feasibility study of hybrid retrofits to an isolated off-grid diesel power plant. Renew Sustain Energy Rev 11(4):635–653
Rowinski PM, Strupczewski WG, Singh VP (2002) A note on the applicability of log-Gumbel and log-logistic probability distributions in hydrological analyses. Hydrol Sci J 47(1):107–122
Rutkowska A, Żelazny M, Kohnová S, Łyp M, Banasik K (2017) Regional L-moment-based flood frequency analysis in the Upper Vistula River Basin, Poland. Pure Appl Geophys 174:701–721
Safari B (2011) Modeling wind speed and wind power distributions in Rwanda. Renew Sustain Energy Rev 15(2):925–935
Safari B, Gasore J (2010) A statistical investigation of wind characteristics and wind energy potential based on the Weibull and Rayleigh models in Rwanda. Renew Energy 35(12):2874–2880
Sahin B, Bilgili M, Akilli H (2005) The wind power potential of the eastern Mediterranean region of Turkey. J Wind Eng Ind Aerodyn 93(2):171–183
Saleh H, Abou El-Azm Aly A, Abdel-Hady S (2012) Assessment of different methods used to estimate Weibull distribution parameters for wind speed in Zafarana wind farm, Suez Gulf, Egypt. Energy 44(1):710–719
Sherlock R (1951) Analyzing winds for frequency and duration. On atmospheric pollution. Springer, pp 42–49
Shin JY, Ouarda TB, Lee T (2016) Heterogeneous mixture distributions for modeling wind speed, application to the UAE. Renew Energy 91:40–52
Sinner C, Dominicy Y, Trufin J, Waterschoot W, Weber P, Ley C (2023) From Pareto to Weibull–a constructive review of distributions on $\mathbb{R}+$. Int Stat Rev 91(1):35–54
Sohoni V, Gupta S, Nema R (2016) A comparative analysis of wind speed probability distributions for wind power assessment of four sites. Turk J Electr Eng Comput Sci 24(6):4724–4735
Soukissian T (2013) Use of multi-parameter distributions for offshore wind speed modeling: the Johnson SB distribution. Appl Energy 111:982–1000
Strupczewski WG, Kochanek K, Markiewicz I, Bogdanowicz E, Weglarczyk S, Singh VP (2011) On the tails of distributions of annual peak flow. Hydrol Res 42(2–3):171–192
Takle ES, Brown JM (1978) Note on the use of Weibull statistics to characterize wind-speed data. J Appl Meteorol 1962–1982(17):556–559
Tar K (2008) Some statistical characteristics of monthly average wind speed at various heights. Renew Sustain Energy Rev 12(6):1712–1724
Ul Hassan M, Hayat O, Noreen Z (2019) Selecting the best probability distribution for at-site flood frequency analysis; a study of Torne River. SN Appl Sci 1:1–10
Van Alphen K, van Sark WG, Hekkert MP (2007) Renewable energy technologies in the Maldives–determining the potential. Renew Sustain Energy Rev 11(8):1650–1674
Vicente RT (2008) Influence of the fitted probability distribution type on the annual mean power generated by wind turbines: a case study at the Canary Islands. Energy Convers Manag 49(8):2047–2054
Vivekanandan N (2015) Flood frequency analysis using method of moments and L-moments of probability distributions. Cogent Eng 2(1):1018704
Yin J (1997) A comparative study of the statistical distributions of wave heights. China Ocean Eng 3:285–304
Zamani A, Badri M (2015) Wave energy estimation by using a statistical analysis and wave buoy data near the southern Caspian Sea. China Ocean Eng 29(2):275–286
Zhang MH (2015) Wind resource assessment and micro-siting: science and engineering. Wiley
Zhang J, Chowdhury S, Messac A et al (2013) A multivariate and multimodal wind distribution model. Renew Energy 51:436–447
Zhou J, Erdem E, Li G et al (2010) Comprehensive evaluation of wind speed distribution models: a case study for North Dakota sites. Energy Convers Manag 51(7):1449–1458