Journal of the Nigerian Statistical Association, Vol. **29**, 2017, 14– 34

Probability models for low wind speeds zones

P. Osatohanmwen[a*], F. O. Oyegue[b], B. Ajibade[c] and R. Idemudia[d]
[a,b,d]*Department of Mathematics, University of Benin, Benin City, Nigeria*; [c]*Department of General Studies, Petroleum Training Institute, Efurrun, Nigeria*

This paper focuses on the comparison of some probability distributions for low wind speeds regimes using some goodness-of-fit criteria. Results obtained from the analysis indicated that while the 2-parameter Weibull distribution reported the best fit in terms of its highest p-value of the Kolmogorov-Smirnov (K-S) statistic, the Akaike information criterion (AIC) reported the 4-parameter transformed beta distribution as the best model for the wind speed observations. Further, the 1-parameter Maxwell distribution presented the best Q-Q plot of all the fitted distributions. The results from the study also suggest that several other probability distributions can be used as efficient alternative to the conventional Weibull distribution in fitting wind speeds data from low wind speeds zones.

Keywords: AIC; Q-Q plot; K-S statistic; wind speed; maximum likelihood; probability distribution.

1. Introduction

Over the years, there has risen a widespread interest in the statistical analysis of meteorological variables. Statistical analysis of meteorological variables includes the day-to-day or month-to-month or even the annual study of variations and frequencies of these variables in any location. These studies involve the systematic observation, recording and analysis of rainfall, temperature, cloud cover, relative humidity, sunshine, flood frequencies, earthquakes and wind regimes across locations. Central to meteorological studies in recent times is the study of wind distributions across locations. These studies have been carried out with an aim to create a predictive base for wind pattern and distributions which grossly affect many activities of men like air travels, sea travels, satellite technology and radio communications etc. On a more practical end, these studies have helped to reinforce the drive to harness the wind as an optimal energy alternative to most non-renewable energy means widely relied on by most countries. Renewable energy sources such as the wind, as an alternative to fossil fuels, is infinitely available, renewable, widely distributed, clean and presents no harmful emissions to the environment(Slootweg et al., 2001). In Nigeria today, despite the availability of abundance of energy sources, the country is still in short supply of electrical power. Of her population spanning over 180 million people, less than 50% have access to grid electricity. The hydro power and fossil fuels which the country majorly rely on as sources of energy are grossly affected by seasonal variations in volume of water and the activities of militants in the Niger Delta respectively, as well as a host of other factors which by implication has impaired the power output of the nation. The electricity supply to the consumers that are connected to the hydro-driven grid is erratic and epileptic and hence, consumers satisfaction is on low ebb. This has spurred the need to harness other alternative form of energy and in particular renewable energy for power generation. The heightened concerns about global warming and the continued apprehensions about nuclear power around the world had also been a driving factor stimulating the drive for renewable energy alternatives.

Wind is a natural phenomenon related to the movement of air masses influenced primarily by the differential solar heating of the earth (Sambo, 2005). Wind is caused by differences in the atmospheric pressure. When there exists a difference in atmospheric pressure, air moves

*Corresponding author. Email: profpato2014@gmail.com

from the higher to the lower pressure area, resulting in winds of diverse speeds. Wind is a classic example of a stochastic variable. Due to its stochastic nature, wind energy cannot be controlled but can be managed (Agbetuyi et al., 2012). The need to manage wind resources stems from the fact that wind power is available only when the wind speed is above certain threshold or baseline (Brady, 2009). Even though wind power is very consistent from year to year, it is also subject to significant variations over shorter time scales; hourly, daily or seasonally (Odo et al., 2012). Bearing in mind that instantaneous electrical generation and consumption must remain in balance to maintain grid stability, this variability can lead to substantial challenges to incorporating large amounts of wind power into a grid system. Statistical analysis is thus required to be able to study the pattern of wind flow in order to guarantee optimal wind energy power generation. More so, the extent to which wind can be exploited as a source of energy depends on the probability density of occurrence of different wind speeds at any generating site. To optimize the design of a wind energy conversion device, data on wind speed range over which the device must operate to maximize energy extraction is required. This in turn requires the knowledge of the frequency distribution of the wind speed. Thus for a wind energy conversion device like the wind turbine, explicit knowledge of the distribution of the wind speed is highly required to enhance output and stability of system as well as for proper wind system design. Several probability distributions have been used in the literature in carrying out wind speed frequency analysis in many studies. The 2-parameter Weibull distribution has been used extensively (Wentink, 1976; Petersen, 1981; Ulgen and Hepbasli, 2002; Celik, 2004; Zaharim et al., 2009; Sarkar and Kasperki, 2009; Gupta and Biswa, 2010; Odo et al., 2012; Osatohanmwen et al., 2016) and it has been found to fit a wide collection of wind regimes. Even though the 2-parameter Weibull distribution has been employed extensively more or less as the conventional wind speed model, it has also been found to be inadequate in fitting some wind speed data (Jaramillo and Borja, 2004; Masseran et al., 2013; Datta and Datta, 2013) and inadequate in fitting the upper tail of most wind speed distributions (Perrin et al., 2006; Sarkar, 2011). Other probability distributions often employed in wind speed studies include the Rayleigh, Burr XII, gamma, inverse gamma, Gaussian, inverse Gaussian, exponential, log-normal, Erlang, exponentiated Weibull, 3-parameter beta, log-logistics, Pearson V, Pearson VI and uniform distributions (Yilmaz and Celik, 2008; Safari, 2011; Masseran et al., 2013; Datta and Datta, 2013). The main aim of all these studies has been to obtain the most appropriate theoretical wind speed distribution of a particular location in other to induce practical decision making and optimal wind energy generation. In this paper, a comparison of several probability distributions used in fitting wind speed data is considered using some goodness-of-fit criteria .

The paper is organized in seven sections. Method employed in the paper is presented in Section 2, a note on the data used for the study is contained in Section 3, and Section 4 contains a discussion on the various probability distributions used for the comparison. Analysis of data is undertaken in Section 5 with discussion of results and conclusion in Sections 6 and 7 respectively.

2. Method

Here we present some statistical tools cogent to the analysis that will be undertaken in this paper. We start by looking at the probability density and cumulative distribution functions. The Maximum likelihood method of parameter estimation for parametric distributions and some goodness-of-fit measures are also considered.

2.1 *Probability density function and cumulative distribution function*

Associated with any continuous random variable X is a probability distribution. A continuous probability distribution is usually defined by a probability density function (PDF) f

and a cumulative distribution function (CDF) F. The PDF f is a function of the values assumed by the random variable X denoted x and a vector of parameter(s) Θ which characterizes the distribution. The PDF gives the probability that the random variable X assumes any real value x on its support. Integrating the PDF gives the CDF of a distribution. The CDF gives the cumulative probability that the random variable X assume values less than and equal to x. In this paper, wind speed is taken to be a continuous random variable X with a PDF $f(x;\Theta)$ and CDF $F(x;\Theta)$.The CDF is related to the PDF by the expression

$$F(x;\Theta) = \int_{-\infty}^{x} f(t;\Theta)\,dt. \tag{2.1}$$

2.2 *Maximum likelihood estimation of wind speed distribution parameters*

Estimating the parameter(s) of a given wind speed distribution is an essential task to be undertaken when carrying out wind speed studies. The maximum likelihood estimation approach is widely used in many fields of study due to some of its robust properties that include efficiency and consistency. For a random independent wind speed sample $x_1, x_2, , x_n$ from a given probability distribution, the approach involves the maximization of the log-likelihood function

$$L = \sum_{i=1}^{n} \log\left(f(x_i;\Theta)\right). \tag{2.2}$$

When carrying out the maximization of (2.2), the solution of some of the resulting systems of equations may not be analytically tractable. To circumvent this, iterative numerical procedures are used to obtain the estimates of the parameters of specific distributions. Some of these iterative schemes are well implemented in some statistical software packages like the R programming language(R Development Core Team, 2009). In this paper, the maximum likelihood estimation technique will be employed in fitting wind speed data to specific distributions.

2.3 *Goodness-of-fit measures*

Goodness-of-fit measures are statistics or graphics used to determine whether observed samples of a random variable fits into a specific theoretical distribution. They are also used to measure the relative performance of a particular theoretical distribution in fitting a given data set when compared with some other family of theoretical distribution(s). In this paper, the goodness-of-fit measures we shall involve for our analysis are the Akaike information criterion (AIC) (Akaike, 1974), the *p-value* of the Kolmogorov-Smirnov (K-S) statistic and the Q-Q plot.

(a) **Akaike information criterion (AIC).** The AIC is a model identification and performance measure which depends on the log-likelihood function of a distribution. The AIC value is calculated using the relation

$$\text{AIC}\left(\hat{\Theta}\right) = 2k - 2L, \tag{2.3}$$

where $\hat{\Theta}$ is the maximum likelihood estimate of the parameter vector Θ, k is the number of parameters of the distribution in question, and L is the value of the log-likelihood. Among several competing families of distributions, the one with the smallest AIC value is considered the best model for the data set.

(b) **Kolmogorov-Smirnov (K-S) statistic.** The Kolmogorov-Smirnov (K-S) statistic is one of the most powerful non-parametric test statistics in the statistical literature. The statistic is obtained by making use of ranks of wind speed observations rather than the actual observed values. The K-S test statistic is used to compare a theoretical cumulative distribution function $F(x;\Theta)$ of a continuous random variable X to an empirical cumulative distribution function (ECDF) $\hat{F}_n(x;\Theta)$ of a random sample of size n. The ECDF is based on the *order statistics*

$$X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$$

with $X_{(i)}$ denoting the ith order statistic. The ECDF $\hat{F}_n(x;\Theta)$ is defined as the number of data points less than or equal x divided by the sample size n. It can be expressed in terms of the order statistics as

$$\hat{F}_n(x;\Theta) = \begin{cases} 0; & x \leq X_{(1)} \\ \frac{j}{n}; & X_{(j)} \leq x < X_{j+1} \\ 1; & x \geq X_{(n)}. \end{cases} \tag{2.4}$$

The K-S statistic is given as

$$D = \max\left[\hat{F}_n(x;\Theta) - F(x;\Theta)\right]. \tag{2.5}$$

When using the K-S statistic, two hypotheses are constructed; the null hypothesis (H_0) and an alternative hypothesis (H_1) for a given $\alpha-$ level of significance. Under H_0, $F(x;\Theta) = \hat{F}_n(x;\Theta)$ while H_1 specifies that $F(x;\Theta) \neq \hat{F}_n(x;\Theta)$. The distribution of the observed sample is taken to be the same as that of the theoretical distribution $F(x;\Theta)$ (i.e. H_0 is true) if

$$D < D_\alpha, \tag{2.6}$$

where D_α is the tabulated K-S critical value read off from the K-S test statistic table for a defined $\alpha-$ level of significance.

Associated with the K-S statistic is a quantity called the *p-value* which is sometimes referred to as the observed level of significance. The *p-value* is defined as the probability of observing a value of the test statistic D as extreme as or more extreme than the one that is observed, when H_0 is true. A *p-value* less than or equal to the given $\alpha-$ level of significance will lead to the rejection of H_0.

(c) **Q-Q plot.** The Q-Q plot is a graphical measure of goodness-of-fit of a given theoretical probability distribution used in fitting a data set. The Q-Q plot graphically compares the empirical quantiles (based on order statistics of the observed sample) of a given data set to that of the theoretical quantiles of the distribution used in fitting the data. The empirical quantiles appears on a 45^0 line on the graph while the theoretical quantiles are scattered along this line. The theoretical distribution quantiles that lies closest to the line is taken to fit the data best among competing distributions.

3. Wind data

The analysis in this paper is based on 3 years (2012 – 2014) daily mean wind speed recordings obtained from the recording station of the National Center for Energy and Environment (NCEE) Benin City, Edo State, South-South Nigeria. The wind speed observations were obtained at a height of 10 meters with 988 sample observations available out of a possible 1096 sample observations. The missing observations were unavoidably due to shut down and repair/replacement of recording systems. Since the missing observations are less than 10% of the possible wind speed observations, results to be obtained using the available sample will still be highly valid. The highest and lowest wind speed observations are $7.18m/s$ and $0.1m/s$ respectively which indicates that the city lies on the low wind speed zone in the country. A look at the wind speed trend in Figure 1 clearly shows that there has been a decrease in the mean daily wind speeds measurement over the period covered. This is largely due to the increased construction of structures which are highly concentrated around the measuring site. Wind speed as we know it, tends to be higher over freer landscape than congested environments. Other environmental factors may have also contributed to the downward moving trend over the years. The wind speed data is available upon request from the corresponding author.

4. Theorectical wind speed distributions

In this section, a discussion on some theoretical wind speed probability distributions is presented. Our discussion shall cover some commonly used wind speed distributions like the Weibull, Rayleigh, lognormal and gamma distributions as well as some not-too-popular wind speed distributions like the transformed beta, paralogistic, transformed gamma and inverse paralogistic distributions.

4.1 *Normal distribution (N-2)*

The normal distribution is the most important probability distribution in statistics. It is a 2-parameter distribution with PDF and CDF given respectively as

$$f\left(x;\mu,\sigma\right)=\frac{1}{\sqrt{2\pi\sigma^{2}}}\mathrm{e}^{-z^{2}/2}, \tag{4.1}$$

$$F\left(x;\mu,\sigma\right)=\frac{1}{2}\pm\frac{1}{2}\gamma\left(\frac{1}{2},\frac{z^{2}}{2}\right)/\sqrt{\pi}, \tag{4.2}$$

$$z=\left(x-\mu\right)/\sigma,-\infty<x<\infty,-\infty<\mu<\infty,\sigma>0,\pi=3.141593,$$

where $\gamma\left(.,.\right)$ is the lower incomplete gamma function. The positive sign in (8) is valid for $z\geq 0$ and the negative sign for $z<0$. The parameters μ and σ are the mean (location parameter) and standard deviation (scale parameter) of the wind speed random variable X respectively (Walck, 2007). The maximum likelihood estimators of the parameters μ and σ are given respectively as

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_{i}, \tag{4.3}$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{\mu}\right)^2}. \tag{4.4}$$

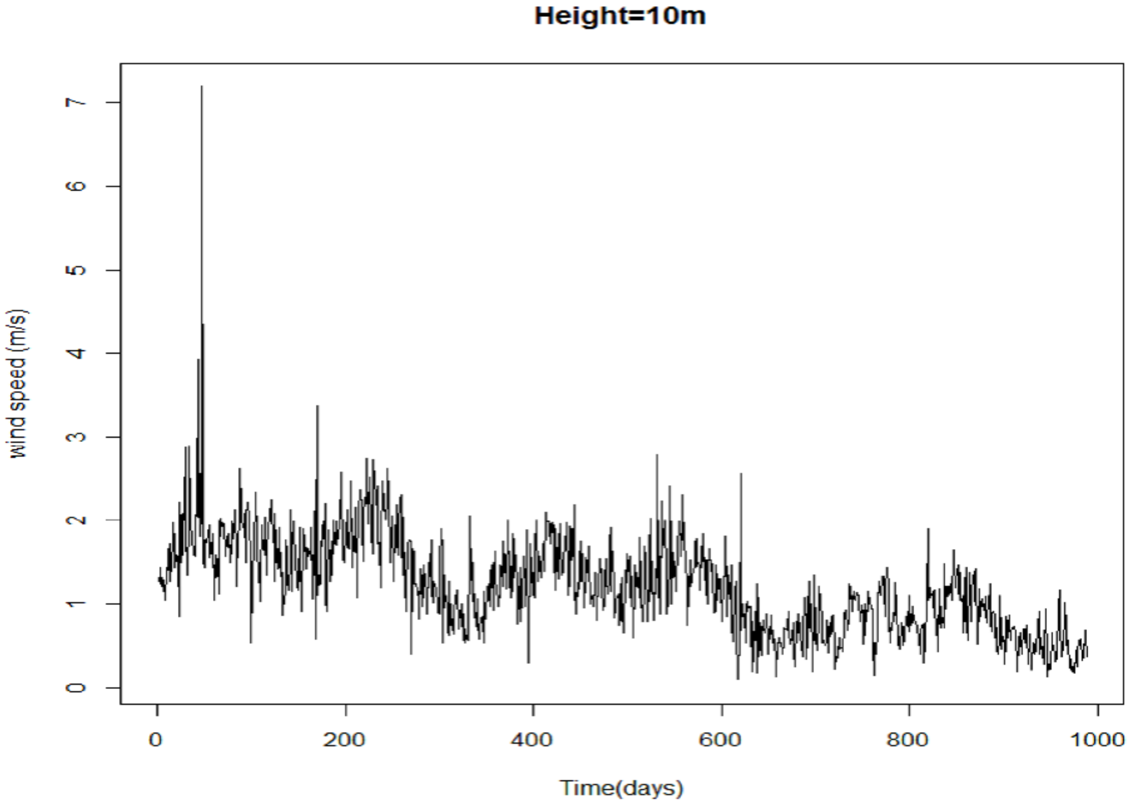

Figure 1. Wind speed data for Benin City (2012 - 2014)

4.2 *Lognormal distribution (LN-2)*

The PDF and CDF of the 2-parameter lognormal (LN-2)distribution are given respectively as

$$f\left(x;\mu,\sigma\right) = \frac{1}{x\sqrt{2\pi\sigma^2}}\mathrm{e}^{-z^2/2}, \tag{4.5}$$

$$F\left(x;\mu,\sigma\right) = \frac{1}{2} \pm \frac{1}{2}\gamma\left(\frac{1}{2},\frac{z^2}{2}\right)/\sqrt{\pi}, \tag{4.6}$$

$$z = \left(\ln x - \mu\right)/\sigma, x, \mu, \sigma > 0, \pi = 3.141593,$$

where $\gamma\left(.,.\right)$ is the lower incomplete gamma function. The positive sign in (12) is valid for $z \geq 0$ and the negative sign for $z < 0$. The parameters μ and σ are the mean and standard deviation of the natural logarithm of the wind speed random variable X respectively (Walck, 2007). The maximum likelihood estimators of the parameters μ and σ are given respectively as

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\ln x_i, \tag{4.7}$$

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\ln x_i - \hat{\mu}\right)^2}. \tag{4.8}$$

4.3 *Maxwell distribution*

The PDF and CDF of the 1-parameter Maxwell (M-1) distribution is given respectively as

$$f\left(x;\alpha\right) = \sqrt{\frac{2}{\alpha^6\pi}}x^2\mathrm{e}^{-x^2/2\alpha^2}, \tag{4.9}$$

$$F\left(x;\alpha\right) = 2\gamma\left(\frac{3}{2},\frac{x^2}{2\alpha^2}\right)/\sqrt{\pi}, \tag{4.10}$$

$$x =\geq 0, \alpha > 0, \pi = 3.141593,$$

where $\gamma\left(.,.\right)$ is the lower incomplete gamma function and α a scale parameter (Walck, 2007). The maximum likelihood estimator of the scale parameter α is given as

$$\hat{\alpha} = \sqrt{\frac{1}{3n}\sum_{i=1}^{n}x_i^2}. \tag{4.11}$$

4.4 *Rayleigh distribution (R-1)*

The PDF and CDF of the 1-parameter Rayleigh (R-1) distribution is given respectively as

$$f\left(x;\sigma\right) = \frac{x}{\sigma^2}\mathrm{e}^{-x^2/2\sigma^2}, \tag{4.12}$$

$$F\left(x;\sigma\right) = \gamma\left(1,\frac{x^2}{2\sigma^2}\right), \tag{4.13}$$

$$x \geq 0, \sigma > 0,$$

where $\gamma\left(.,.\right)$ is the lower incomplete gamma function and σ a scale parameter (Johnson et al., 1995; Walck, 2007). The maximum likelihood estimator of the scale parameter σ is given as

$$\hat{\sigma} = \sqrt{\frac{1}{2n}\sum_{i=1}^{n}x_i^2}. \tag{4.14}$$

4.5 *Transformed gamma distribution (TG-3)*

The transformed gamma distribution is a 3-parameter distribution with PDF and CDF given respectively as

$$f\left(x;\alpha,\beta,c\right)=\frac{\beta\left(x/c\right)^{\alpha\beta}\mathrm{e}^{-(x/c)^{\beta}}}{x\Gamma\left(\alpha\right)}, \tag{4.15}$$

$$F\left(x;\alpha,\beta,c\right)=\frac{\gamma\left(\beta(\alpha-1)+1,(x/c)^{\beta}\right)}{\Gamma\left(\alpha\right)}, \tag{4.16}$$

$$x\geq 0,\alpha,\beta,c>0,$$

where $\Gamma\left(.\right)$ and $\gamma\left(.,.\right)$ are the gamma and lower incomplete gamma functions respectively. The parameters α and β are shape parameters and c a scale parameter (Klugmann et al., 2008). The maximum likelihood estimates of the scale parameter c and shape parameters α and β are obtained by solving the systems of equations

$$\frac{n}{\beta}+\alpha\sum_{i=1}^{n}\mathrm{ln}x_i-n\alpha\mathrm{lnc}-\sum_{i=1}^{n}\left(\frac{x_i}{c}\right)^{\beta}\ln\left(\frac{x_i}{c}\right)=0$$

$$-\frac{n}{c}-\frac{n(\alpha\beta-1)}{c}+\frac{\beta}{c}\sum_{i=1}^{n}\left(\frac{x_i}{c}\right)^{\beta}=0$$

$$\beta\sum_{i=1}^{n}\mathrm{ln}x_i-n\beta\mathrm{lnc}-n\psi\left(\alpha\right)=0$$

for α,β and c. Where $\psi\left(.\right)$ is the digamma function which is the same as the derivative of the natural logarithm of $\gamma\left(.\right)$.

4.6 *Gamma distribution (G-2)*

The 2-parameter gamma distribution is the 3-parameter transformed gamma distribution with $\beta=1$. Its PDF and CDF are given respectively as

$$f\left(x;\alpha,c\right)=\frac{\left(x/c\right)^{\alpha}\mathrm{e}^{-(x/c)}}{x\Gamma\left(\alpha\right)}, \tag{4.17}$$

$$F\left(x;\alpha,c\right)=\frac{\gamma\left(\alpha,(x/c)\right)}{\Gamma\left(\alpha\right)}, \tag{4.18}$$

$$x\geq 0,\alpha,c>0,$$

where $\Gamma\left(.\right)$ and $\gamma\left(.,.\right)$ are the gamma and lower incomplete gamma functions respectively. The parameters α and c are shape and scale parameters respectively (Klugmann et al., 2008). The maximum likelihood estimates of the scale parameter c and shape parameter α are obtained by solving the systems of equations

$$-\frac{n}{c}-\frac{n(\alpha-1)}{c}+\frac{1}{c^2}\sum_{i=1}^{n}x_i=0$$

$$\sum_{i=1}^{n}\ln x_i-n\ln c-n\psi\left(\alpha\right)=0$$

for α and c.

4.7 *Weibull distribution (W-2)*

The Weibull distribution is the most widely used theoretical distribution in wind speed studies. The 2-parameter Weibull distribution is the 3-parameter transformed gamma distribution with $\alpha = 1$. Its PDF and CDF are given respectively as

$$f\left(x;\beta,c\right)=\frac{\beta\left(x/c\right)^{\beta}\mathrm{e}^{-(x/c)^{\beta}}}{x}, \tag{4.19}$$

$$F\left(x;\beta,c\right)=\gamma\left(1,\left(x/c\right)^{\beta}\right), \tag{4.20}$$

$$x\geq 0,\beta,c>0,$$

where $\gamma\left(.,.\right)$ is the lower incomplete gamma function. The parameters β and c are shape and scale parameters respectively (Klugmann et al., 2008). The maximum likelihood estimates of the scale parameter c and shape parameter β are obtained by solving the systems of equations

$$\frac{n}{\beta}+\sum_{i=1}^{n}\ln x_i-n\ln c-\sum_{i=1}^{n}\left(\frac{x_i}{c}\right)^{\beta}\ln\left(\frac{x_i}{c}\right)=0$$

$$-\frac{n}{c}-\frac{n(\beta-1)}{c}+\frac{\beta}{c}\sum_{i=1}^{n}\left(\frac{x_i}{c}\right)^{\beta}=0$$

for β and c.

4.8 *Exponential distribution (E-1)*

The 1-parameter exponential distribution is the 3-parameter transformed gamma distribution with $\alpha = \beta = 1$. Its PDF and CDF are given respectively as

$$f\left(x;c\right)=\frac{1}{c}\mathrm{e}^{-(x/c)}, \tag{4.21}$$

$$F\left(x;c\right)=\gamma\left(1,\left(x/c\right)\right), \tag{4.22}$$

$$x\geq 0, c>0,$$

where $\gamma\left(.,.\right)$ is the lower incomplete gamma function and the parameter c, a scale parameter (Klugmann et al., 2008). The maximum likelihood estimator of the scale parameter c is given as

$$\hat{c}=\frac{1}{n}\sum_{i=1}^{n}x_i. \tag{4.23}$$

4.9 *Chi-square distribution (CS-1)*

The 1-parameter chi-square distribution is a special case of the gamma distribution. Its PDF and CDF are given respectively as

$$f\left(x;p\right)=\frac{1/2\left(\frac{x}{2}\right)^{\frac{p}{2}-1}\mathrm{e}^{-\left(x/2\right)}}{\Gamma\left(p/2\right)}, \tag{4.24}$$

$$F\left(x;p\right)=\frac{\gamma\left(p/2,x/2\right)}{\Gamma\left(p/2\right)}, \tag{4.25}$$

$$x>0, p\geq 0,$$

where $\Gamma\left(.\right)$ and $\gamma\left(.,.\right)$ are the gamma and lower incomplete gamma functions respectively. The parameter p is the degree of freedom. The maximum likelihood estimates of p is obtained by solving the equation

$$\psi\left(p/2\right)=\frac{1}{2n}\sum_{i=1}^{n}\ln\left(x_i/2\right)$$

for p.

4.10 *Transformed beta distribution (TB-4)*

The 4-parameter transformed beta distribution is defined by its PDF and CDF given respectively as

$$f\left(x;a,b,c,s\right)=\frac{b/s\left(x/s\right)^{bc-1}\left[1+\left(x/s\right)^{b}\right]^{-\left(a+c\right)}}{B\left(a,c\right)}, \tag{4.26}$$

$$F\left(x;a,b,c,s\right)=\frac{B_t\left(a,c\right)}{B\left(a,c\right)}, \tag{4.27}$$

$$t = (x/s)^b, x > 0, a, b, c, s > 0,$$

where $B(.,.)$ and $B_t(.,.)$ are the beta and incomplete beta functions respectively, and $B(.,.) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$, p and q are real numbers (Klugmann et al. 2008). The parameters a, b and c are shape parameters while s is a scale parameter. The maximum likelihood estimates of a, b, c and s are obtained by solving the systems of equations

$$n\psi(a+c) - n\psi(a) - \sum_{i=1}^{n} \ln\left[1 + (x_i/s)^b\right] = 0$$

$$\frac{n}{b} + c\sum_{i=1}^{n} \ln x_i - nc\ln s - (a+c)\sum_{i=1}^{n} \frac{(x_i/s)^b \ln(x_i/s)}{1 + (x_i/s)^b} = 0$$

$$n\psi(a+c) - n\psi(c) + b\sum_{i=1}^{n} \ln x_i - nb\ln s - \sum_{i=1}^{n} \ln\left[1 + (x_i/s)^b\right] = 0$$

$$-\frac{nbc}{s} + \frac{b(a+c)}{s^{b+1}} \sum_{i=1}^{n} \frac{x_i^b}{1 + (x_i/s)^b} = 0$$

for a, b, c and s.

4.11 *Burr distribution (B-3)*

The 3-parameter Burr distribution also known as the Burr XII distribution (Johnson et al., 1995) is the 4-parameter transformed beta distribution with $c = 1$. Its PDF and CDF are given respectively as

$$f(x; a, b, s) = \frac{b/s\,(x/s)^{b-1}\left[1 + (x/s)^b\right]^{-(a+1)}}{B(a, 1)}, \tag{4.28}$$

$$F(x; a, b, s) = \frac{B_t(a, 1)}{B(a, 1)}, \tag{4.29}$$

$$t = (x/s)^b, x > 0, a, b, s > 0,$$

where $B(.,.)$ and $B_t(.,.)$ are the beta and incomplete beta functions respectively (Klugmann et al., 2008). The parameters a and b are shape parameters while s is a scale parameter. The maximum likelihood estimates of a, b and s are obtained by solving the systems of equations

$$\frac{n}{a} - \sum_{i=1}^{n} \ln\left[1 + (x_i/s)^b\right] = 0$$

$$\frac{n}{b}+\sum_{i=1}^{n}\text{ln}x_i-n\text{ln}s-(a+1)\sum_{i=1}^{n}\frac{(x_i/s)^b\ln(x_i/s)}{1+(x_i/s)^b}=0$$

$$-\frac{nb}{s}+\frac{b(a+1)}{s^{b+1}}\sum_{i=1}^{n}\frac{x_i^b}{1+(x_i/s)^b}=0$$

for a, b and s.

4.12 *Inverse Burr distribution (IB-3)*

The 3-parameter inverse Burr distribution also known as Burr III distribution (Johnson et al., 1995) is the 4-parameter transformed beta distribution with $a = 1$. Its PDF and CDF are given respectively as

$$f(x;b,c,s)=\frac{b/s\,(x/s)^{bc-1}\left[1+(x/s)^b\right]^{-(c+1)}}{B(1,c)}, \tag{4.30}$$

$$F(x;b,c,s)=\frac{B_t(1,c)}{B(1,c)}, \tag{4.31}$$

$$t=(x/s)^b, x>0, b, c, s>0,$$

where $B(.,.)$ and $B_t(.,.)$ are the beta and incomplete beta functions respectively (Klugmann et al., 2008). The parameters b and c are shape parameters while s is a scale parameter. The maximum likelihood estimates of b, c and s are obtained by solving the systems of equations

$$\frac{n}{b}+c\sum_{i=1}^{n}\text{ln}x_i-nc\text{ln}s-(c+1)\sum_{i=1}^{n}\frac{(x_i/s)^b\ln(x_i/s)}{1+(x_i/s)^b}=0$$

$$\frac{n}{c}+b\sum_{i=1}^{n}\text{ln}x_i-nb\text{ln}s-\sum_{i=1}^{n}\ln\left[1+(x_i/s)^b\right]=0$$

$$-\frac{nbc}{s}+\frac{b(c+1)}{s^{b+1}}\sum_{i=1}^{n}\frac{x_i^b}{1+(x_i/s)^b}=0$$

for b, c and s.

4.13 *Log logistic distribution (LL-2)*

The 2-parameter log logistic distribution also known as Fisk distribution is the 4-parameter transformed beta distribution with $a = c = 1$. Its PDF and CDF are given respectively as

$$f(x; b, s) = b/s\,(x/s)^{b-1}\left[1 + (x/s)^b\right]^{-2}, \tag{4.32}$$

$$F(x; b, s) = B_t(1, 1), \tag{4.33}$$

$$t = (x/s)^b, x > 0, b, s > 0,$$

where $B_t(.,.)$ is the incomplete beta function (Klugmann et al., 2008). The parameter b is a shape parameter while s is a scale parameter. The maximum likelihood estimates of b and s are obtained by solving the systems of equations

$$\frac{n}{b} + \sum_{i=1}^{n} \ln x_i - n \ln s - 2\sum_{i=1}^{n} \frac{(x_i/s)^b \ln(x_i/s)}{1 + (x_i/s)^b} = 0$$

$$-\frac{nb}{s} + \frac{2b}{s^{b+1}} \sum_{i=1}^{n} \frac{x_i^b}{1 + (x_i/s)^b} = 0$$

for b and s.

4.14 *Paralogistic distribution (PL-2)*

The 2-parameter paralogistic distribution is the 4-parameter transformed beta distribution with $b = a$ and $c = 1$. Its PDF and CDF are given respectively as

$$f(x; a, s) = \frac{a^2}{s}(x/s)^{a-1}\left[1 + (x/s)^a\right]^{-(a+1)}, \tag{4.34}$$

$$F(x; a, s) = aB_t(a, 1), \tag{4.35}$$

$$t = (x/s)^a, x > 0, a, s > 0,$$

where $B_t(.,.)$ is the incomplete beta function (Klugmann et al., 2008). The parameter a is a shape parameter while s is a scale parameter. The maximum likelihood estimates of a and s are obtained by solving the systems of equations

$$\frac{2n}{a} + \sum_{i=1}^{n} \ln x_i - n \ln s - (a+1)\sum_{i=1}^{n} \frac{(x_i/s)^a \ln(x_i/s)}{1 + (x_i/s)^a} - \sum_{i=1}^{n} \ln\left[1 + (x_i/s)^a\right] = 0$$

$$-\frac{na}{s}+\frac{a(a+1)}{s^{a+1}}\sum_{i=1}^{n}\frac{x_i^a}{1+(x_i/s)^a}=0$$

for a and s.

4.15 *Inverse paralogistic distribution (IPL-2)*

The 2-parameter inverse paralogistic distribution is the 4-parameter transformed beta distribution with $a=1$ and $c=b$. Its PDF and CDF are given respectively as

$$f\left(x;b,s\right)=\frac{b^2}{s}\left(x/s\right)^{b^2-1}\left[1+\left(x/s\right)^b\right]^{-(b+1)}, \tag{4.36}$$

$$F\left(x;b,s\right)=bB_t(1,b), \tag{4.37}$$

$$t=\left(x/s\right)^b, x>0, b, s>0,$$

where $B_t(.,.)$ is the incomplete beta function (Klugmann et al., 2008). The parameter b is a shape parameter while s is a scale parameter. The maximum likelihood estimates of b and s are obtained by solving the systems of equations

$$\frac{2n}{b}+2b\sum_{i=1}^{n}\ln x_i-2nb\ln s-(b+1)\sum_{i=1}^{n}\frac{\left(x_i/s\right)^b\ln\left(x_i/s\right)}{1+\left(x_i/s\right)^b}-\sum_{i=1}^{n}\ln\left[1+\left(x_i/s\right)^b\right]=0$$

$$-\frac{nb^2}{s}+\frac{b(b+1)}{s^{b+1}}\sum_{i=1}^{n}\frac{x_i^b}{1+\left(x_i/s\right)^b}=0$$

for b and s.

5. Data analysis

For the analysis, the theoretical wind speed distributions discussed in Section 4 were used to fit the wind speed data. The maximum likelihood estimates of the parameter(s) of the distributions alongside the goodness-of-fit statistics discussed in section two is presented in Table 1. The 5% level of significance is used for the analysis. The Newton-Raphson types Numerical optimization procedure were used in obtaining the maximum likelihood estimates of some distribution parameters. These iterative numerical techniques were implemented in the R programming software. The R programmes for the analysis can be made available to interested readers upon request from the corresponding author. In Figure 2 (a-o), density plot of all the fitted distribution is presented. The Q-Q plots of all the theoretical distributions are given in Figure 3 (a-o). The purpose of the tabular and graphical presentations is to determine among the several wind speed probability models considered in this paper, the one(s) that is/are most adequate in carrying out wind speed analysis for the given location. The need to also explore the performance of all the distributions considered is also a rationale for the analysis.

6. Discussion of results

The normal distribution, as one of the distributions used for the analysis presents a relatively poor fit for the wind speed observations. This is because, while its K-S statistic value and *p-value* is indicating a good fit at the 5% level of significance, its AIC value is relatively high. Again, the Q-Q plot for the normal distribution clearly shows that the lower and upper quantiles of the wind speeds are poorly estimated. The lognormal distribution reported a lower AIC value than the normal distribution even though its K-S statistic value and *p-value* indicates that the distribution is different from the distribution of the observed wind speeds at the 5% level of significance.

Table 1: Maximum likelihood estimates of the parameters of wind speed distributions (Standard error of estimates in parenthesis)

Distributions	Parameter estimate(s)	AIC	K-S statistic	*p-value*
N-2	$\hat{\mu}$=1.2239 (0.0186) $\hat{\sigma}$=0.5846 (0.0186)	1746.941	0.0386	0.1056
LN-2	$\hat{\mu}$=0.0756 (0.0171) $\hat{\sigma}$=0.5381 (0.0121)	1732.703	0.0859	9.25×10^{-7}
M-1	$\hat{\alpha}$=0.7831 (0.0424)	1663.822	0.0621	0.00097
R-1	$\hat{\sigma}$=0.9591 (0.0153)	1663.462	0.0512	0.0114
TG-3	$\hat{\alpha}$=2.0355 (0.3710) $\hat{\beta}$=1.4880 (0.1527) $\hat{c}$=0.7995 (0.1492)	1631.179	0.0397	0.0888
G-2	$\hat{\alpha}$=4.1141 (0.3710) $\hat{c}$=0.2975 (0.0137)	1639.53	0.0548	0.0053
W-2	$\hat{\beta}$=2.1780 (0.0498) $\hat{c}$=1.3802 (0.0212)	1652.202	0.0297	0.3489
E-1	$\hat{c}$=1.2239 (0.0260)	2377.28	0.2533	2.2×10^{-16}
CS-1	$\hat{p}$=1.9518 (0.04487)	2579.935	0.2861	2.2×10^{-16}

TB-4	$\hat{a}$=2.5240 (0.9898) $\hat{b}$=4.0263 (0.8083) $\hat{c}$=0.5216 (0.1363) $\hat{s}$=2.0447 (0.2122)	1610.149	0.0343	0.1958
B-3	$\hat{a}$=6.3806 (1.7060) $\hat{b}$=2.5179 (0.0849) $\hat{s}$=2.7513 (0.3704)	1613.713	0.0311	0.2964
IB-3	$\hat{b}$=0.2813 (0.0325) $\hat{c}$=6.7452 (0.5469) $\hat{s}$=1.7268 (0.0495)	1614.652	0.0387	0.1033
LL-2	$\hat{b}$=3.3334 (0.0881) $\hat{s}$=1.1278 (0.0188)	1716.131	0.0606	0.0014
PL-2	$\hat{a}$=2.7833 (0.0623) $\hat{s}$=1.8204 (0.0282)	1625.622	0.0391	0.0963
IPL-2	$\hat{b}$=2.3436 (0.0466) $\hat{s}$=0.6896 (0.0133)	1849.468	0.1022	2.16×10^{-9}

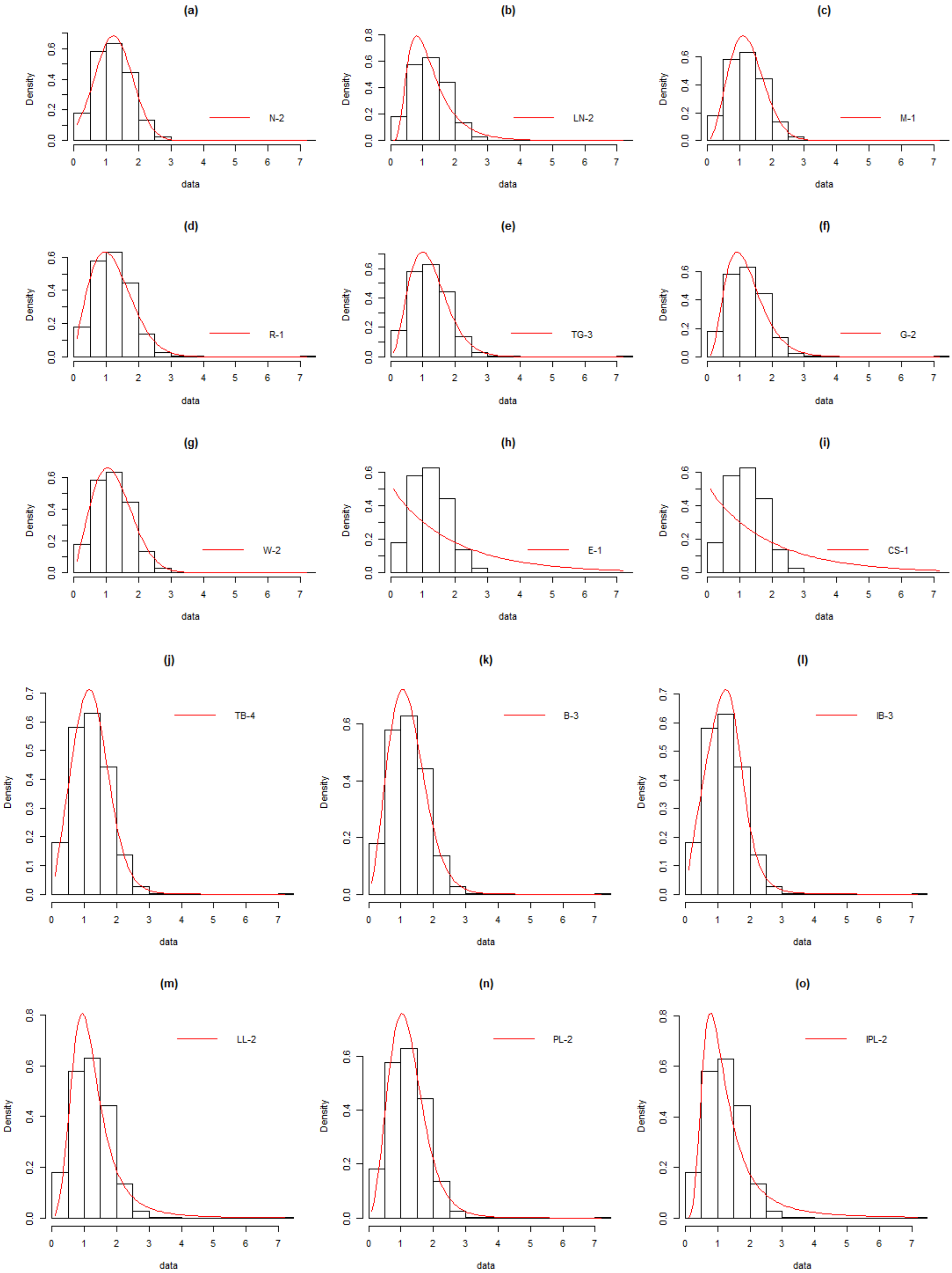

Figure 2(a-o). Density plots of fitted wind speeds distributions

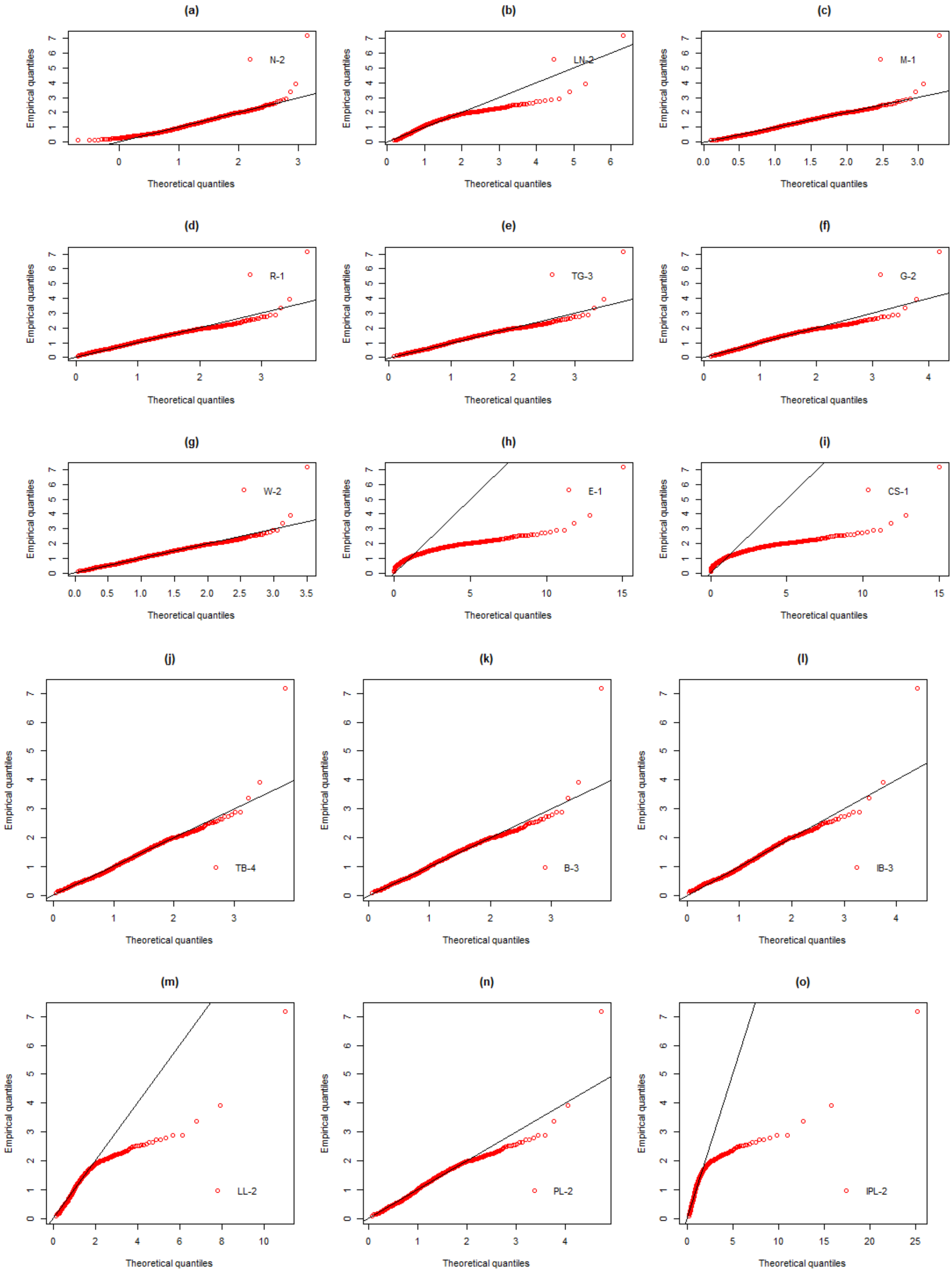

Figure 5(a-o). Q-Q plots of fitted wind speed distributions

The upper quantiles of the wind speed distribution was poorly estimated by the lognormal distribution as indicated by its Q-Q plot.

The AIC value of the Maxwell distribution is smaller than that of the normal and log-normal distributions even though its K-S statistic value and *p-value* indicates that the

Maxwell distribution is different from the distribution of the observed wind speeds at the 5% level of significance. The Q-Q plot of the Maxwell distribution is best among all the distributions considered. The Rayleigh distribution is observed to be the best distribution for the data among the 1-parameter distributions considered for the analysis. The Q-Q plot for the Rayleigh distribution also attests to the adequacy of the distribution.

The transformed gamma distribution which generalizes the gamma, Weibull, exponential, and chi-square distributions offered an adequate fit for the wind speed data. The AIC value, the K-S statistic value, the *p-value* and the Q-Q plot for the transformed gamma distribution support a very good fit of the data to the distribution. The gamma distribution also presents a very good fit to the data. The conventional Weibull distribution also proved to be a very good distribution for the data with the highest p-value. The other goodness-of-fit statistics and graphics also supports a very good fit of the Weibull distribution to the data. Meanwhile, the exponential and chi-square distributions reported the poorest fit to the data among the distributions used for the analysis.

The 4-parameter transformed beta distribution gave the best fit to the wind speeds data among all the distributions considered for the study based on its lowest AIC value. The K-S statistic value and the *p-value* also confirmed its adequacy in fitting the data. The Burr and inverse Burr distributions also gave good fits to the data. The paralogistic distribution gave a very good fit to the data while the inverse paralogistic and log logistic distributions gave very poor fits.

7. Conclusion

Fitting of specific probability distributions to observed wind speed samples of various locations has been carried out extensively in the literature. While the 2-parameter Weibull distribution has been adopted more or less as the conventional wind speed model, studies which include the one carried out in this paper have suggested that other probability distributions may also be adequate. The over-arching purpose of these studies has been to show which probability distribution best captures the wind condition of a particular location in order to enhance policy formulations, engineering design of wind energy conversion devices and systems.

We recommend based on the results obtained from the study that the gamma, Raleigh, Burr, transformed gamma, paralogistic and Weibull distributions be considered as suitable candidate models when undertaking wind power analysis and preliminary assessment and design of wind turbines for low wind speeds zones. More so, the Weibull distribution offered the best model from the study and should be adopted as the leading model in the analysis of wind speed in the low wind speeds zones.

Acknowledgement

The authors are very grateful to the Editor-in-Chief and the referees who offered useful comments for the improvement of this paper. We would also like to express our sincere gratitude to the National Center for Energy and Environment (NCEE), University of Benin, Benin City, for supplying us with the data used for the analysis.

References

Agbetuyi, A.F., Akinbulire, T.O., Abdul Kareem, A.O., and Awosope, C.O.A. (2012). Wind energy potential in Nigeria. *International Electrical Engineering Journal (IEEJ)*, 3: 595-601.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19: 716723.

Brady, T.F. (2009). A simulation solution of the integration of wind power into an electricity generating network. *Proceedings of the 2009 winter conference*, 1523-1529.

Celik, A.N. (2004). A statistical analysis of wind power density based on the Weibull and Rayleigh models at the Southern region of Turkey. *Renewable Energy*, 29: 593-604.

Datta, D., and Datta, D. (2013). Comparison of Weibull distribution and exponentiated Weibull distribution based estimation of mean and variance of Wind data. *International Journal of Energy, Information and Communications*, 4: 1-12.

Gupta, R., and Biswa, A. (2010). Wind data analysis of Silchar (Assam, India) by Rayleigh and Weibull methods. *Journal of Mechanical Engineering Research*, 2: 10-24.

Jaramillo, O.A., and Borja, M.A. (2004). Wind speed analysis in La ventosa, Mexico: A Bi-model probability distribution case. *Renewable Energy*, 29: 1613-1630.

Johnson, N.L., Kotz, S., and Balakrishnan, N. (1995). *Continuous univariate distributions* (second edition, vol. 2). New York, John Wiley & sons, Inc.

Klugman, S.A., Panjer, H.H., and Wilmot, G.E. (2008). *Loss Models, From Data to Decisions* (Third Edition), Wiley.

Masseran, N., Razal, A.M. Ibrahim, K., Zaharim A., and Sopian, K. (2013). The probability distribution model of wind speed over East Malaysia. *Research Journal of Applied Sciences, Engineering and Technology*, 6: 1774-1779.

Odo, F.C., Offiah, S.U., and Ugwuoke, P.E. (2012). Weibull distribution-based model for prediction of wind potential in Enugu, Nigeria. *Advances in Applied Science Research*, 3: 1202-1208.

Osatohanmwen, P., Oyegue, F.O., and Ogbonmwan, S.M. (2016). Statistical analysis of wind energy potential in Benin City using the 2-parameter Weibull distribution. *International Journal for Renewable Energy and Environment*, 2: 22-31.

Perrin O., Rootzén H., Taesler R. (2006). A discussion of statistical methods used to estimate extreme wind speeds. *Theoretical and Applied Climatology*, 85: 203-215.

Petersen, T.L., Troen, I., Frandsen, S., and Hedegaard, K. (1981). *Wind Atlas for Denmark*, RISO, Denmark.

R Development Core Team.*R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria; 2009. http://www.R-project.org.

Safari, B. (2011). Modeling wind speed and wind power distribution in Rwanda. *Renewable and Sustainable Energy*, Revision, 15: 925-935.

Sambo, A.S. (2005). Renewable energy for rural development: The Nigerian perspective. *ISESCO Science and Technology Vision*, 1: 16-18.

Sarkar, A., and Kasperki, M. (2009). Weibull parameters for Wind Speed distribution in India. *Proceedings of 5th National conference on Wind Engineering*, pp. 134-158.

Sarkar A., Singh, S. and Mitra, D. (2011). Wind Climate modeling using Weibull and Extreme value distribution. *International Journal of Engineering Science and Technology*, 3: 100-106.

Slootweg, J.C., Haan, S.W.H., Polinder, H., and Kling, W.L. (2001). Modeling wind turbines in power system dynamics simulation. *Proceedings of the Power Engineering Society Summer Meeting Conference*, 1: 22-26.

Ulgen, K., and Hepbasli, A. (2002). Determination of Weibull parameters for wind Energy analysis of Izmir, Turkey. *International Journal of Energy Research*, 26: 495-506.

Walck, C. (2007). *Hand-book on statistical distributions for experimentalists.* Internal Report SUF-PFY/96-01.

Wentink, T. (1976). Study of Alaskan wind power potential and its possible application. *Final Rep.*, Rep No. NSF/RANN/SE/AER 74-0023/FR 76/1, Geophysical Institute, University of Alaska.

Yilmaz, V., and Celik, H.E. (2008). A statistical approach to estimate the wind speed distribution: The Case of Gelibolu region. Dogus University, *Dergisi*, 9: 122-132.

Zaharim, A., Razali, A.M., Abidin, R.Z., and Sopian, K. (2009). Fitting of statistical distributions to wind speed data in Malaysia. *European Journal of Scientific Research*, 26: 6-12.