International Journal of Probability and Statistics 2022, 11(1): 19-27
DOI: 10.5923/j.ijps.20221101.03

Estimation of Wind Power Using a New Two Variable Copula-based Model in South Eastern Nigeria

Opabisi Adeyinka Kosemoni[1,*], Adeyemo Samuel O.[1], Ohaegbulam Promise O.[2]

[1]Department of Mathematics/Statistics, Federal Polytechnic Nekede, Owerri, Imo State, Nigeria
[2]Department of Food Technology, Federal Polytechnic Nekede, Owerri, Imo State, Nigeria

Abstract This study examines the use of other probability models as alternative for modeling wind speed data. The 2-parameter Weibull distribution is usually taken as the conventional model for fitting and analyzing wind speed observations. However, within the vanguard of statistical modeling, there exist other probability models which can also serve this purpose. In this paper, a study on ten probability models which can be used to fit wind speed observation is carried out. The models include: the Weibull, normal, Cauchy, gamma, power Lindley, generalized Lindley, generalized Rayleigh, generalized Maxwell, lognormal and the generalized exponential distributions. All ten models were used to fit monthly average wind speed observations from South-East Nigeria for the period (1987 – 2019). Results from the analysis showed that while the Weibull model remains a dominant model for the wind regime, the power Lindley and the normal distributions offered a very suitable alternative as well, as they offered very good fits to the wind speed observation which are indicative by their respective AIC, K-S and P-values.

Keywords Goodness-of-fit Measures, Maximum Likelihood, Parameters, Probability Distributions, Wind Speed

1. Introduction

Within the practice of wind speed modeling, the 2-parameter Weibull distribution with shape and scale parameters has been used more or less as the conventional wind speed model of which have been explored by several scholars. This has been due to the fact that the Weibull model covers a wide range of the shape of most wind regime because of the flexibility inherent in its density [10]. However, other models have also been suggested and used in many studies as alternative to the Weibull model depending on the wind regime that is being studied. Examples of such models include: the Rayleigh, Burr XII, gamma, inverse gamma, normal, inverse normal, exponential, log-normal, exponentiated Weibull, log-logistics, Pearson V, Pearson VI and the uniform distributions. The main reason for these studies has been to obtain the most appropriate theoretical wind speed distribution of a particular location or wind regime.

In this paper, a study of ten wind speed model is undertaken using data from South-East Nigeria which lies on a low wind speed zone. The Weibull distribution alongside the normal, Cauchy, gamma, power Lindley, generalized Lindley, generalized Rayleigh, generalized Maxwell, lognormal and the generalized exponential distributions are used to fit wind speed sample from the zone and the results compared using some statistical goodness-of-fit measures. These distributions are two variable models with some being special cases of some extended distributions [21]. The rest of the paper is organized as follows. In section 2 the ten wind speed models considered are specified by their cumulative distribution function (cdf) and their probability density function (pdf). A description of the data used for the study and the analytical tools of analysis are contained in section 3. Section 4 contains the results obtained from the study. The paper closes in section 5 with summary and conclusion.

* Corresponding author:
yinksko@yahoo.com (Opabisi Adeyinka Kosemoni)
Received: Nov. 16, 2022; Accepted: Dec. 19, 2022; Published: Dec. 28, 2022
Published online at http://journal.sapub.org/ijps

2. Wind Speed Models

Here we specify the cdfs and the pdfs of the various wind speed models considered in the study. The functional form of the Weibull model is used to begin the section.

2.1. Weibull Distribution (W)

The cdf and pdf of the Weibull distribution are given respectively by

$$F(x;c,k) = 1 - e^{-\left(\frac{x}{c}\right)^k} \tag{1}$$

$$f(x;c,k) = \frac{k}{c}\left(\frac{x}{c}\right)^{k-1} e^{-\left(\frac{x}{c}\right)^k} \tag{2}$$

$$x > 0, c > 0, k > 0$$

where *c* and *k* are scale and shape parameters respectively. For *k*=1, the distribution becomes the exponential distribution, for *k*=2, the distribution becomes the Rayleigh distribution and for *k*=4 the distribution becomes approximately normal [15].

2.2. Gamma Distribution (G)

The gamma distribution has cdf and pdf defines respectively as

$$F(x;c,k)=\gamma\left(k,\left(\frac{x}{c}\right)\right) \tag{3}$$

$$f(x;c,k)=\frac{\left(\frac{x}{c}\right)^{k}e^{-\frac{x}{c}}}{x\Gamma(k)}$$
$$x>0,c>0,k>0 \tag{4}$$

Where $\Gamma(.)$ and $\gamma(.,.)$ are the gamma and lower incomplete gamma functions respectively. The parameters *c* and *k* are scale and shape parameters respectively. For *k*=1, the distribution becomes the exponential distribution [16].

2.3. Normal Distribution (N)

The normal distribution is the most important distribution in statistical theory. It has cdf and pdf given respectively by

$$F(x;c,k)=\frac{1}{2}\pm\frac{1}{2}\gamma\left(\frac{1}{2},\frac{Z^2}{2}\right)/\sqrt{\pi}, \tag{5}$$

$$f(x;c,k)=\frac{1}{\sqrt{2\pi c^2}}e^{-\frac{z^2}{2}}$$
$$z=(x-k)/c,-\infty<x<\infty,-\infty<k<\infty,c>0, \tag{6}$$
$$\pi=3.141593$$

Where $\gamma(.,.)$ is the lower incomplete gamma function. The positive sign in $F(x;c,k)$ is valid for $z\geq 0$ and the negative sign for $z<0$. The parameters *k* and *c* are the mean (location parameter) and standard deviation (scale parameter) of the distribution respectively [17].

2.4. Lognormal Distribution (LN)

The lognormal distribution is the distribution of a random variable whose logarithm is normally distributed. It has cdf and pdf expressed respectively by

$$F(x;c,k)=\frac{1}{2}\pm\frac{1}{2}\gamma\left(\frac{1}{2},\frac{Z^2}{2}\right)/\sqrt{\pi}, \tag{7}$$

$$f(x;c,k)=\frac{1}{x\sqrt{2\pi c^2}}e^{-\frac{z^2}{2}} \tag{8}$$
$$z=(\ln x-k)/c,\ c>0,k>0,\ \pi=3.141593$$

Where $\gamma(.,.)$ is the lower incomplete gamma function. The positive sign in $f(x;c,k)$ is valid for $z\geq 0$ and the negative sign for $z<0$. The parameters k and *C* are the location and scale parameters of the distribution respectively [17].

2.5. Cauchy Distribution (C)

The Cauchy distribution also known as the Cauchy – Lorentz distribution, is the distribution of the ratio of two independent normally distributed random variables with zero means. It has cdf and pdf given respectively by

$$F(x;c,k)=\frac{1}{\pi}\arctan\left(\frac{x-k}{c}\right)+\frac{1}{2}, \tag{9}$$

$$f(x;c,k)=\frac{1}{\pi c\left[1+\left(\frac{x-k}{c}\right)^2\right]}$$
$$-\infty<x<\infty,-\infty<k<\infty,c>0,\pi=3.14159 \tag{10}$$

where the parameters k and *C* are the location and scale parameters of the distribution respectively [17].

2.6. Power Lindley Distribution (PL)

The power Lindley distribution was developed by Ghitany et al. [18]. It has cdf and pdf given by

$$F(x;c,k)=1-\left[1+\frac{cx^k}{c+1}\right]e^{-cx^k}, \tag{11}$$

$$f(x;c,k)=\frac{kc^2}{c+1}(1+x^k)x^{k-1}e_{-}{}^{cx^k} \tag{12}$$
$$x>0,c>0,k>0$$

where the parameters *c* and *k* are scale and shape parameters respectively and for *k* =1 the distribution becomes the Lindley distribution.

2.7. Generalized Lindley Distribution (GL)

The generalized Lindley distribution was proposed by Nadarajah et al. [19] with cdf and pdf expressed respectively as

$$F(x;c,k)=\left[1-\left[1+\frac{cx}{c+1}\right]e^{-cx}\right]^k, \tag{13}$$

$$f(x;c,k)=\frac{kc^2}{c+1}(1+x)e_{-}{}^{cx}\left[1-\left[1+\frac{cx}{c+1}\right]e^{-cx}\right]^{k-1} \tag{14}$$
$$x>0,c>0,k>0$$

where the parameters *c* and *k* are scale and shape parameters respectively and for *k=1* the distribution becomes the Lindley distribution.

2.8. Generalized Exponential Distribution (GE)

The generalized exponential distribution was defined by Gupta and Kundu [20] as an alternative distribution to the Weibull distribution. The cdf and pdf of the distribution is expressed respectively as

$$F(x;c,k)=(1-e^{-cx})^{k}, \quad (15)$$

$$f(x;c,k)=kce^{-cx}(1-e^{-cx})^{k-1} \quad (16)$$
$$x>0, c>0, k>0$$

where the parameters c and k are scale and shape parameters respectively and for k=1 the distribution becomes the exponential distribution.

2.9. Generalized Rayleigh Distribution (GR)

The generalized Rayleigh distribution is a generalization of the 1-parameter Rayleigh distribution through the addition of a shape parameter. The cdf and pdf of the distribution is given respectively by

$$F(x;c,k)=\left(1-e^{-\frac{x^2}{2c^2}}\right)^{k}, \quad (17)$$

$$f(x;c,k)=\frac{kx}{c^2}e^{-\frac{x^2}{2c^2}}\left(1-e^{-\frac{x^2}{2c^2}}\right)^{k-1} \quad (18)$$
$$x>0, c>0, k>0$$

Where c and k are scale and shape parameters respectively. For k=1, the distribution becomes the Rayleigh distribution [17].

2.10. Generalized Maxwell Distribution (GM)

The generalized Maxwell distribution is an extension of the 1-parameter Maxwell distribution through the addition of a scale parameter. The cdf and pdf of the distribution are given respectively by

$$F(x;c,k)=\left[2\gamma\left(\frac{3}{2},\frac{x^2}{2c^2}\right)/\sqrt{\pi}\right]^{k} \quad (19)$$

$$f(x;c,k)=k\sqrt{\frac{2}{\alpha^6\pi}}x^2e^{-\frac{x^2}{2\alpha^2}}\left[2\gamma\left(\frac{3}{2},\frac{x^2}{2c^2}\right)/\sqrt{\pi}\right]^{k-1} \quad (20)$$
$$x>0, \alpha c>0, k>0, \pi=3.141593$$

Where c is a scale parameter, k is shape parameter and $\gamma(.,.)$ is the lower incomplete gamma function [17]. For k=1 , the distribution becomes the Rayleigh distribution.

3. Data and Analytical Tools

The analysis is based on 33 years (1987- 2019) monthly average wind speed observations obtained from South-East Nigeria. The wind speed observations were obtained at a height of 10m with 384 sample observations. The highest and lowest wind speed observations are 5.8*m/s* and 0.7*m/s* respectively and this indicates that the region lies on the low wind speed zone in Nigeria. The mean wind speed is 3.59*m/s*, while the median wind speed is 3.60*m/s* The coefficient of skewness is -0.12 which clearly indicates that the distribution of the wind speed observation is skewed to the left. Also, the coefficient of excess kurtosis is -0.36 which implies that the distribution of the wind speed is light-tailed. Figure 1 shows the trend plot of the wind speed observation. The actual wind speed sample can be made available upon request from the corresponding author.

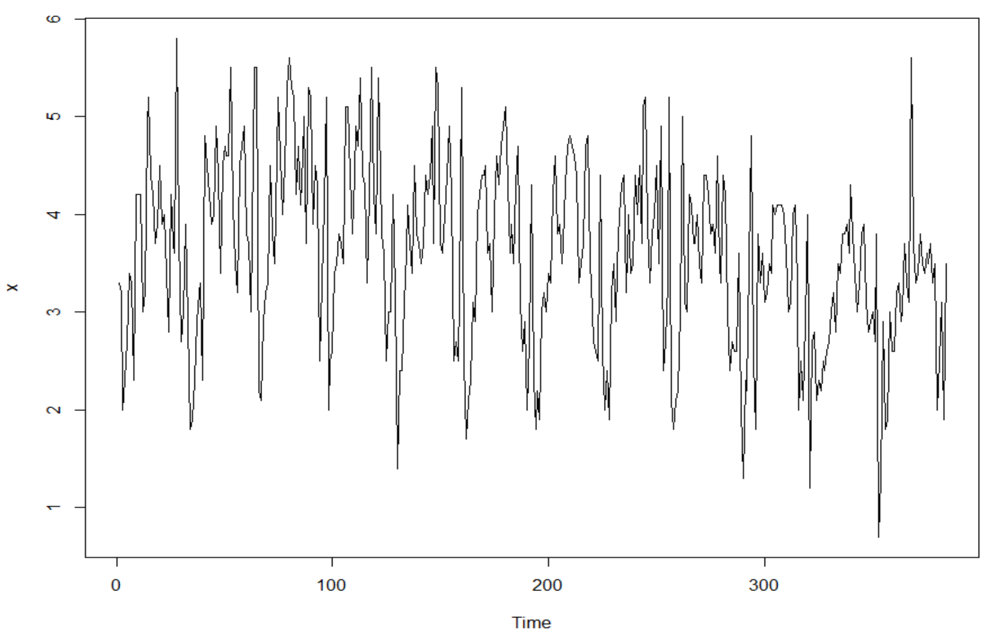

Figure 1. The monthly average wind speed observation from South-East Nigeria (1987-2019)

The ten parametric distribution presented in section 2 are used to fit the wind speed observations using the Maximum likelihood method of parameter estimation. For a random independent wind speed sample x_1, x_2,...x_n, the maximum likelihood approach involves the definition of a likelihood function given as

$$L=\prod_{i=1}^{n} f(x_i;\Theta) \tag{21}$$

The maximum likelihood estimate of the parameter vector Θ is usually obtained by maximizing the logarithm of the likelihood function *L* called the log-likelihood function. The log-likelihood function is defined as

$$L=\sum_{i=1}^{n} \log\left(f(x_i;\Theta)\right) \tag{22}$$

When carrying out the maximization of *L*, the solution of some of the systems of equations may not be analytically tractable. To resolve this, iterative numerical procedures were used to obtain the estimates of parameter(s) of respective distributions. The iterative scheme used in this paper is the Quasi-Newton Scheme which was implemented in the R programming language [21].

Two goodness-of-fit statistics were computed and used in the analysis to compare the performance of all ten distributions. These goodness-of-fit statistics include the Akaike Information Criterion (AIC) and the Kolmogorov – Smirnov (K-S) statistic and its attendant p-value. The AIC value is calculated using the relation

$$AIC\left(\tilde{\Theta}\right)=2p-2L \tag{23}$$

where $\tilde{\Theta}$ is the maximum likelihood estimator of the parameter vector Θ, *p* is the number of parameters in the parameter vector Θ and *L* is the value of the log-likelihood. Among several competing distributions, the one with the smallest AIC value is considered the best model for the data set. The Kolmogorov-Smirnov (K-S) statistic is one of the most useful non-parametric test statistics in statistical literature. The statistic is obtained by making use of ranks of the wind speed observations. 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) $\tilde{F}_n(x;\Theta)$ of a random sample of size *n*. The ECDF is based on the *order statistics*

$$X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)} \tag{24}$$

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

$$\hat{F}_n(x;\Theta)=\begin{cases} 0; & x<X_{(1)} \\ j/n; & X_{(j)} \le x < X_{(j+1)} \\ 1; & x \ge X_{(n)} \end{cases} \tag{25}$$

The K-S statistic is given as

$$D=\max[F(x;\Theta)-\hat{F}_n(x;\Theta)] \tag{26}$$

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 $F(x;\Theta)\ne\hat{F}_n(x;\Theta)$. 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 $D<D_\alpha$ where D_α is the tabulated critical value for the given $\alpha-$ 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. In this paper we took $\alpha=0.05$. Also, the distribution with the highest p-value of the K-S statistic value is adjudged as the best in fitting the wind speed observation. The density plots which compare the histograms of the data and the fitted densities, the cumulative distribution function plots which compare the ECDF of the data and the fitted cdfs, the Q-Q plots which compare the empirical quantiles and the theoretical quantiles of the fitted distributions and the P-P plots which compare the empirical probabilities and the theoretical probabilities of the fitted distributions for all the fitted distributions were also generated and used in assessing the performance of all the distributions in fitting the wind speed observations.

4. Results

The results of parameter estimates of the ten distributions are contained Table 1 with their respective *AIC*, *K-S* and *p*-values.

Results from the analysis clearly show that while the Weibull distribution remains a dominant wind speed model for all wind speed regimes, in this study, there is great evidence that the normal distribution offers a suitable alternative to the Weibull model. Also, the power Lindley distribution which has been limitedly used for wind speed analysis offers another great alternative as it provided a very good fit to the data also. The other distributions considered in this study were greatly out-performed by the Weibull, normal and the power Lindley distributions. This is supported by the AIC values, the K-S statistics values and its p-value. A look at the density plots, the cdf plots, the Q-Q plots and the P-P plots also reveal this position. The major result from this study is that the well-known Weibull and normal distributions can be convenient replaced in wind speed studies by the power Lindley distribution which possess similar shape characteristics to these two in low wind speed zones.

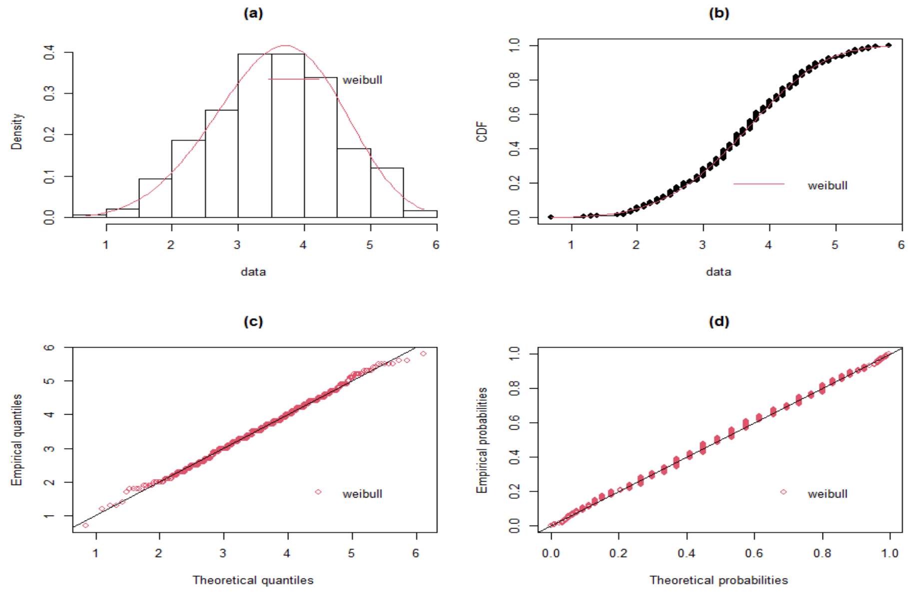

Figure 2 (a-d). Fitted Weibull distribution

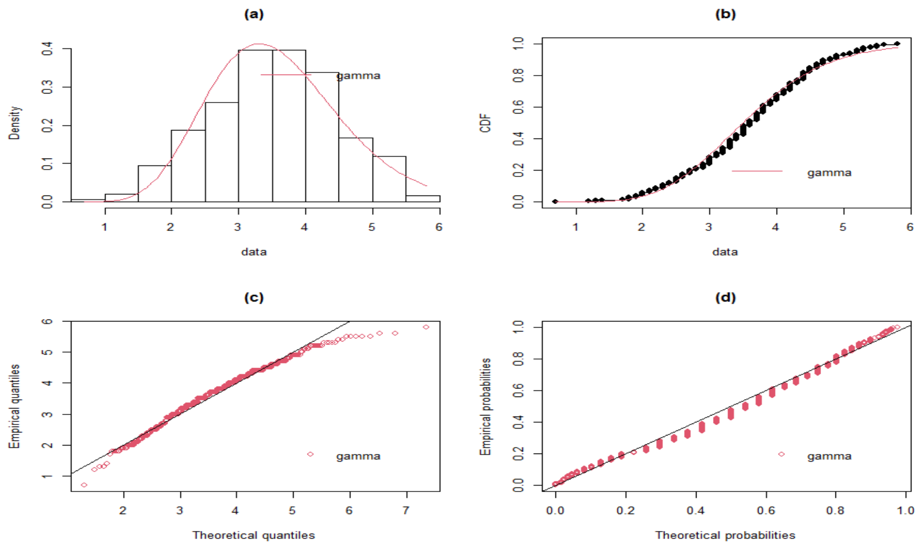

Figure 3 (a-d). Fitted gamma distribution

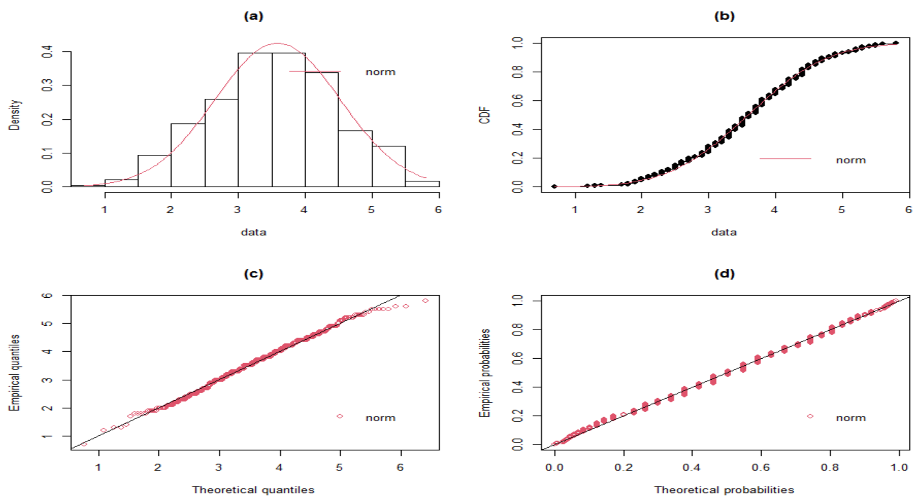

Figure 4 (a-d). Fitted normal distribution

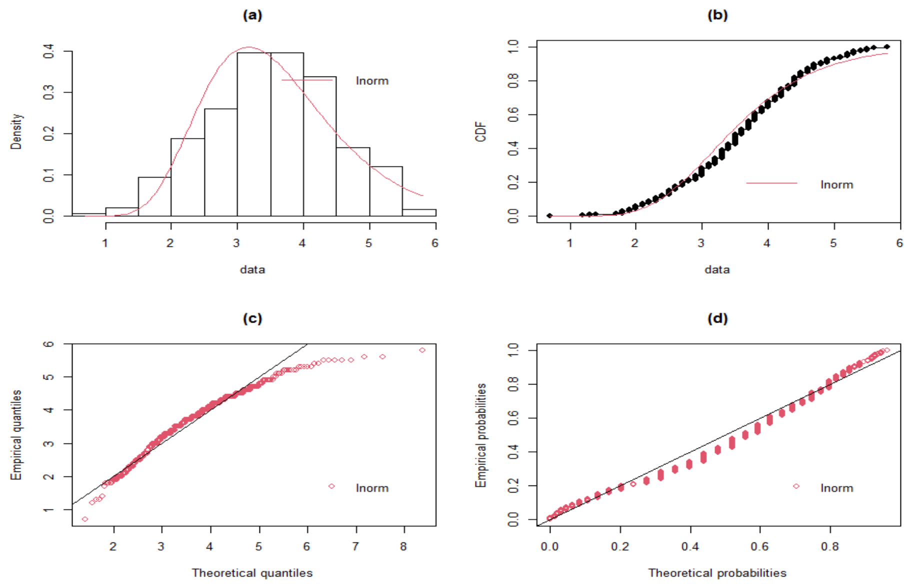

Figure 5 (a-d). Fitted lognormal distribution

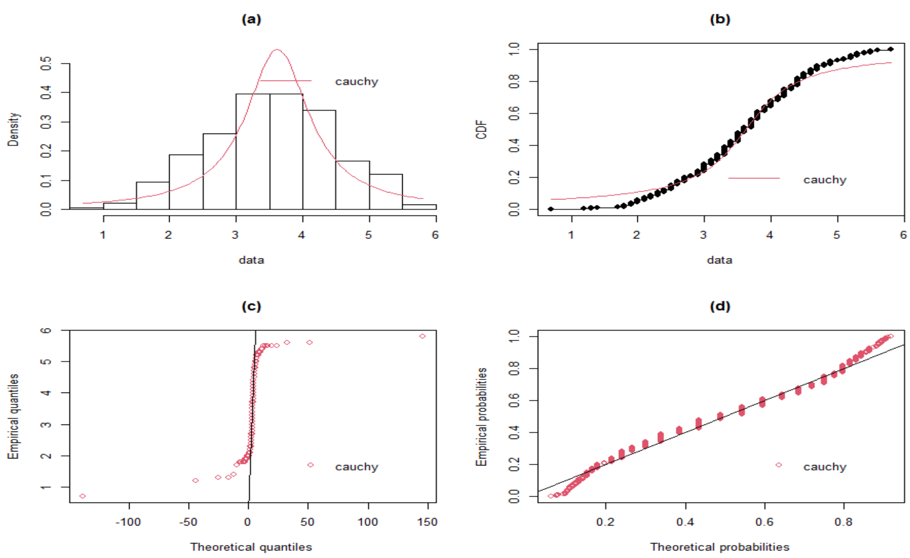

Figure 6 (a-d). Fitted Cauchy distribution

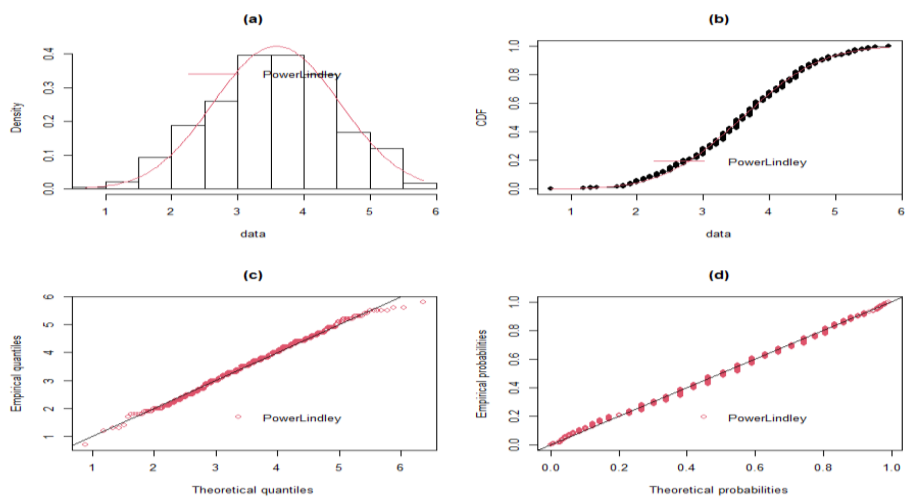

Figure 7 (a-d). Fitted power Lindley distribution

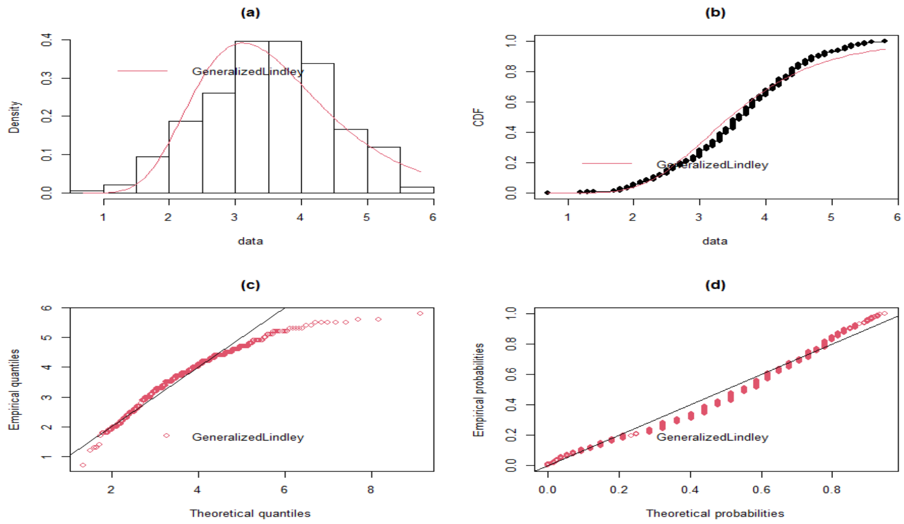

Figure 8 (a-d). Fitted generalized Lindley distribution

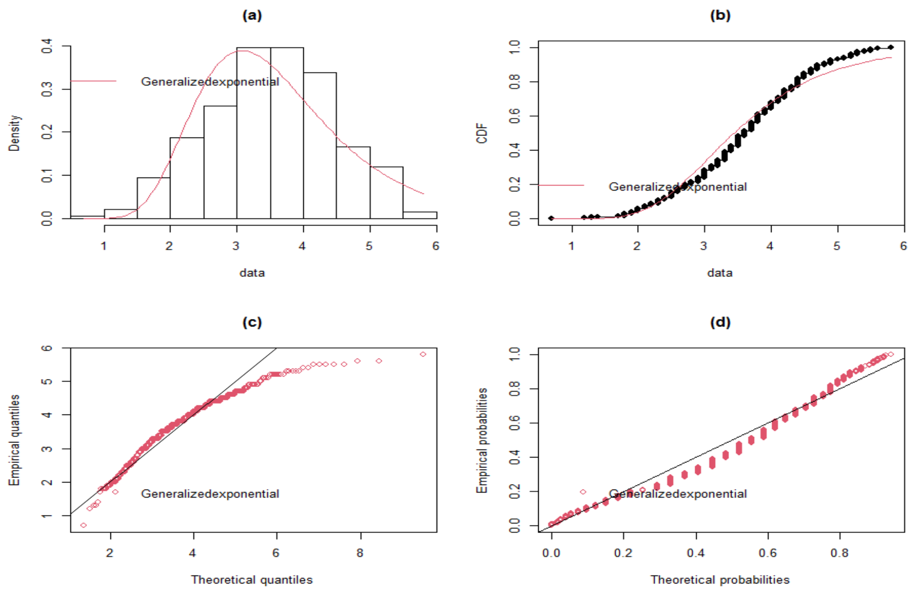

Figure 9 (a-d). Fitted generalized exponential distribution

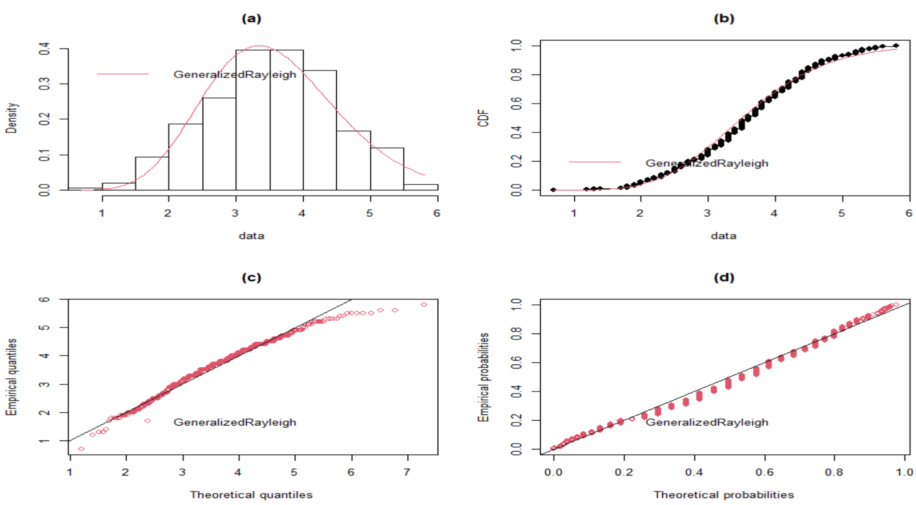

Figure 10 (a-d). Fitted generalized Rayleigh distribution

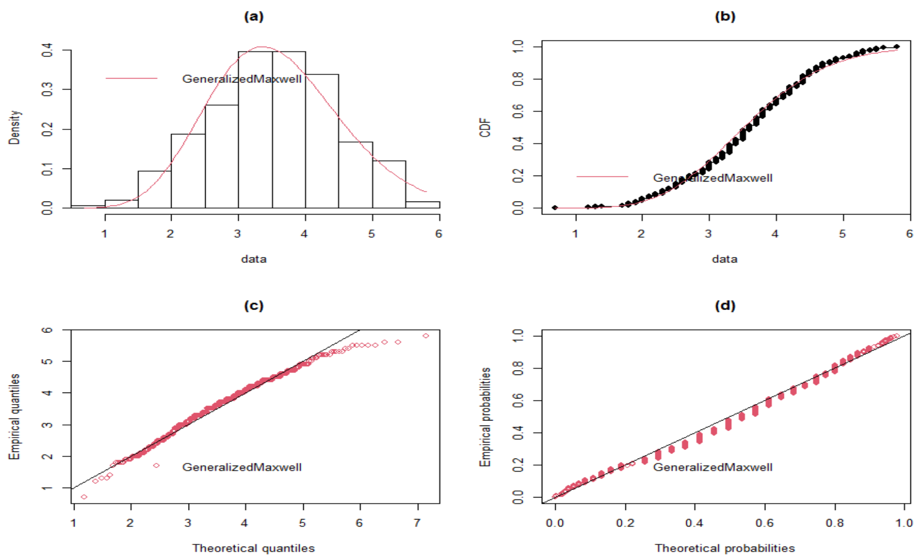

Figure 11 (a-d). Fitted generalized Maxwell distribution

Table 1. Maximum likelihood fit of wind speed distributions

Distribution\Parameter	*c*	*k*	*AIC*	*K-S*	*p-value*
W	3.9423 (0.0490)	4.3276 (0.3144)	1041.04	0.0382	0.6170
G	3.5952 (0.2612)	12.9014 (0.9193)	1072.76	0.0783	0.0170
N	0.9377 (0.0338)	3.5885 (0.0479)	1044.33	0.0380	0.6221
LN	0.2941 (0.0106)	1.2385 (0.0150)	1104.90	0.0986	0.0011
C	0.5806 (0.0378)	3.6202 (0.0.477)	1195.59	0.0882	0.0048
PL	0.0377 (0.0056)	2.9533 (0.1043)	1043.66	0.0388	0.6000
GL	1.2350 (0.0431)	16.6457 (2.0324)	1110.96	0.1002	0.0008
GE	1.0347 (0.0404)	24.0937 (2.9318)	1121.63	0.1055	0.0004
GR	1.8150 (0.0440)	4.1176 (0.3556)	1066.56	0.0756	0.0236
GM	0.3473 (0.0162)	2.6213 (0.2206)	1062.83	0.0724	0.0341

(Standard error of each parameter estimate in parenthesis)

5. Summary and Conclusions

The modeling of wind speed observations from South-East Nigeria has been carried out in this study. Ten probability distributions were used to fit the wind speed sample from the region for the period (1987 – 2019). The ten distributions include the well-known Weibull model which is the dominant model in wind speed studies, and nine other 2-parameters distributions. The maximum likelihood method was used to estimate the parameters of all the distributions. The results obtained from the study clearly showed that the normal and power Lindley distributions proved to be a very efficient alternative to the Weibull distribution. This has spurred our suggestion that these two models should also be considered when carrying out wind speed analysis for the region. We hope that this study will beckon the attention of other researchers working within the area of wind speed modeling.

ACKNOWLEDGEMENTS

The authors wishes to express our appreciation to TETFund for approving and releasing funds for this study through the Institution-Based Research programme for full sponsorship of this work; and sincerely thank the

Management of Federal Polytechnic Nekede, Owerri, Imo State, Nigeria, for graciously approving and positively recommending this proposal.

Appendix

Meaning of Terms Used

Term	Meaning
W	Weibull Distribution
G	Gamma Distribution
N	Normal Distribution
LN	Lognormal Distribution
C	Cauchy Distribution
PL	Power Lindley Distribution
GL	Generalized Lindley Distribution
GE	Generalized Exponential Distribution
GR	Generalized Rayleigh Distribution
GM	Generalized Maxwell Distribution
AIC	Alkaike Information Criteria
K-S	Kolgomorov Smirnov

REFERENCES

[1] 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(2); 1202-1208.

[2] 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.

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

[4] 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.

[5] 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.

[6] 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*, Vol. 26, No. 1, pp.6-12.

[7] 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.

[8] 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.

[9] 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*, Vol.2, pp. 22-31.

[10] Osatohanmwen, P., Oyegue, F.O., Ajibade, B., and Idemudia, R. (2017) Probability models for low wind speeds zones. Journal of the Nigerian Statistical Association. 29, 14 – 34.

[11] 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*, Vol. 6, Issue 10, pp. 1774-1779.

[12] 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,* Vol. 4, Issue 4.

[13] Yilmaz, V., and Celik, H.E. (2008). A statistical approach to estimate the wind speed distribution: The Case of Gelibolu region. Dogus University; *Dergisi* Vol. 9, Issue 1, pp. 122-132.

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

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

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

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

[18] Ghitany, M. E., Al-Mutairi, D. K., Balakrishnan, N., and Al-Enezi, L. J. (2013). Power Lindley distribution and associated inference. *Computational Statistics & Data Analysis*, 64, 20 - 33.

[19] Nadarajah, S, Bakouch, H.S., and and Tahmasbi, R. (2011). A generalized Lindley distribution. *Sankhya B*, 73, 331 – 359.

[20] Gupta, R. D. and Kundu, D. (1999). Generalized exponential distributions, *Australian and New Zealand Journal of Statistics*, 41, 173 - 188.

[21] Rather A.A. & Ozei G. (2020): The Weighted Power Lindley distribution with applications on the lifetime data, *Pakistan Journal of Statistics and Operations Research, Vol. 16 No 2 pp 225-237*

[22] Ekhosuehi, N., Nzei, L. C. & Opone F. (2020): A new mixture of exponential-Gamma distribution, *Gazi University Journal of Science, GUJ Sci 33(2): 548-564.*