Wind Speed Analysis for Coastal Regions of Pakistan using Extended Generalized Lindley Distribution

Rana Muhammad USMAN[1]   , Nurbanu BURSA[2]   , Muhammad AHSAN-UL-HAQ[1,*]

[1]College of Statistical & Actuarial Sciences, University of the Punjab, Lahore Pakistan.
[2]Department of Statistics, Faculty of Science, Hacettepe University, Turkey

Highlights

• This paper focuses on the modeling of wind speed data.
• Used extended generalized Lindley distribution as wind speed distribution.
• Comparison has been made with well-known models for wind speed analysis.

| Article Info	Abstract
	The wind energy potential of a specified area can be estimated using wind speed distribution. In this study, the selection of probability density functions is used to model wind speed data recorded at two stations in Pakistan. The suitability of fitted distributions is evaluated using the goodness of fit criterion, power density error, log-likelihood, root mean square error, coefficient of determination, AIC, and BIC. The wind speed data are obtained from two coastal regions of Pakistan at 10m/s average rate for session 2017-2018. Findings indicated that the extended generalized Lindley distribution provide generally the best fit to the wind speed data for both stations. However, it is also observed that power Lindley and extended generalized Lindley distributions have better performance based on power density error criteria in Gwadar and Haripur, respectively.

1. INTRODUCTION

Energy is considered one of the essential elements for the growth of societies. The increase in the world population, as well as the rising well-being of the developing world industry and economy, increases the need for energy day by day. Developed countries have focused on renewable energy resources due to a decrease in sources of fossil fuels i.e. coal, crude oil, and natural gas. Recently, wind energy is considered one of the most widely used renewable energy technologies because it does not cause air pollution and provides inexhaustible power [1, 2].

The wind consists of temperature and pressure differences due to the warming and cooling of the atmosphere. Wind energy is a derivative of solar energy. Approximately 2% of the solar energy is transformed into the kinetic energy of the wind. This amount is hundreds of times the total world energy consumption [3]. The chaotic nature of the wind is a major obstacle to predictability in energy production. For this reason, researchers prefer physical methods and statistical methods to minimize errors in wind energy estimates [4].

Since the energy crisis in Pakistan is enlarged over the past few decades due to the high prices of fossil fuels and severe environmental problems, it is the need of the hour to escalate the generation of energy from non-conventional sources. The energy demand has increased in Pakistan over the last few decades and is likely to increase more. The electricity generation from wind energy is the most operative use of

*Corresponding author, e-mail: ahsanshani36@gmail.com

wind energy. Moreover, the distribution of wind speed is significant for power generators. Therefore, it is preferred to use statistical distributions to evaluate the wind speed data in Pakistan.

Pakistan is a South Asian country with a population of approximately 200 million. Unfortunately, World Energy Outlook has provided a statistic that approximately 51 million (27% of the population) people of Pakistan did not have proper access to electricity. Fossil energy sources are still used intensively in Pakistan and renewable energy in the country is a less developed sector compared to Europe as seen in Figure 1.
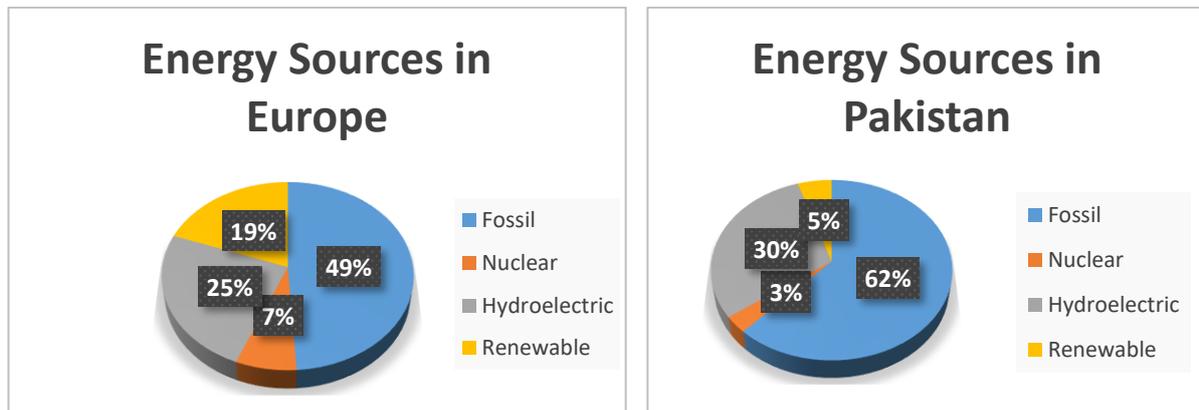
Figure 1. *Comparing percentages of energy sources in Europe and Pakistan in 2016*
(Source: World Energy Outlook Statistics Report, 2016)

Although it can produce 132 GW of wind energy nationwide, wind energy accounts for 0.1% of total electricity generation [1]. Lack of planned data related to wind energy and choice of proper distribution are the key reasons for this limited application of wind energy in Pakistan. Therefore, it is critical to model wind speed behavior in Pakistan according to a statistical perspective. In recent years, electricity crises that lived in the country and environmental groups have succeeded in attracting the attention of the authorities to renewable energy sources for energy production. The government has planted new projects about wind energy which will be completed soon.

Numerous works have been observed in the literature related to wind speed analysis for different regions of different countries [2]. It has been observed from the literature that the well-known Weibull distribution is believed as the most appropriate distribution to deal with wind speed data i.e. [5-9]. But it is also a fact that WD is not always apt to model wind speed, particularly in high and low, multi-modal, and skewed cases [10-15]. Therefore, it is a need of the hour to seek an alternate wind speed distribution for the analysis of wind data.

Numerous researchers had justified that other developed distributions can perform better than that of WD in modeling wind speed data. For example, [16] showed that the inverted WD is better than that of WD for the analysis of wind speed data in Turkey. [17] compared Log-Normal (LN), Gamma (GD), and Rayleigh (RD) distribution with WD for the wind speed analysis of five sites of Iran. They found that the LN distribution performs better as compared to Weibull as well as RD based on RMSE and adjusted R^2 criterion. [18] made a comparison of heterogeneous mixture distributions (HMD) with WD for bimodality skewed wind speed data of the United Arab Emirates. They observed based on Bayesian Information Criteria (BIC), R^2, and Chi-square statistics that the HMD is better as compared to WD for modeling. [19] found in their analysis that the multi-peak Gaussian model has a better fit as compared to WD, as well as RD[20], compared bimodal Weibull, Kappa (KAP), and Wakeby (WAK) distributions with two-parameter WD for the wind speed data of North America. Based on R^2, they observed that bimodal Weibull distribution performs better as compared to KAP, WAK, and two-parameter Weibull distribution. Moreover, [21] showed the comparison of other distributions with the WD to justify the superiority of generalized distributions.

It is observed from all given researches that WD is limiting to develop distribution frequency for all kinds of wind regimes. More appropriately, WD could not be used for the analysis of bimodal, multimodal, and highly skewed data sets in proficient manners. Therefore, we have to consider other probabilistic distributions to minimize the error of wind power estimation. From move on this point, the purpose of this study is to catalogue the goodness of fit of different generalized Lindley distributions for two monitoring stations in coastal regions of Pakistan (Gwadar and Haripur) and compare them with WD, the extensively used distribution for wind speed analysis.

The motivation of the study is to analyze the wind speed data for coastal areas of Pakistan by using some probability distributions. Parameters are being estimated through the ML estimation method. The evaluation of the goodness-of-fit of the pdfs to the data is carried out by the use of numerous goodness-of-fit criteria's including the log-likelihood (ln L) criteria.

The remaining study is structured in the following sections as Section 2 gave descriptive statistics about wind speed data for selected coastal areas of Pakistan to be analyzed. Section 3 took the review of extensions of Lindley's distribution and other distribution concerning their pdf. Section 4 provides a brief concept about the goodness of fit criteria. Section 5 presented the results and discussions about the analysis and then the final section concluded our paper with useful remarks.

	Nomenclature
WD	Weibull Distribution
LN	Log-Normal Distribution
GD	Gamma Distribution
RD	Rayleigh Distribution
HMD	Heterogeneous mixture distributions
KAP	Kappa Distribution
WAK	Wakeby Distribution
LD	Lindley Distribution
EGL	Extended Generalized Lindley
EQL	Exponentiated Quasi Lindley Distribution
PL	Power Lindley
GL	Generalized Lindley
RMSE	Root Mean Square Error
R^2	Coefficient of Determination
GW	Giga Watt
Pdf	Probability density function
Cdf	Cumulative distribution function
ρ	Density of Air (kg/m^3)
A	Blade speed area of wind turbine (m^2)
f_D	The density function of the given distribution
v_i	ith observation of the wind speed data
y_i	ith predicted probability from a special distribution
x_i	ith observed probability
z_i	the ith computed value from the correlated equation for the same value of x_i

2. WIND SPEED DATA FOR COASTAL AREAS

Wind speed data were collected from the World Bank via energy data under a project funded by ESMAP. The data was recorded at the height of 40m from 2016 to 2018. The observations in the data are 10 min. average wind speed observed two monitoring stations situated in Gwadar (situated at the shore of Arabian Sea) and Haripur (situated at the Shore of Indus River between Taxila and Abbottabad) in Pakistan. Table 1 shows geographical information for given stations and Table 2 shows descriptive statistics for both monitoring stations.

Table 1. *Data duration, geographical coordinates and elevation for Gwadar and Haripur regions*

Regions	Data Period	Latitude	Longitude	Elevation (m)	Height (m)
Gwadar	Nov 2016 – Aug 2018	25° 07 35′′ N	62° 19′ 21′′ E	08	40
Haripur	Oct 2016 – Aug 2018	33° 59′ 39′′ N	72° 56′ 00′′ E	520	40

According to Table 2, the average wind speed in Haripur is faster than that of Gwadar and also the maximum wind speed is observed in Haripur which is 520 m above sea level. Moreover, the variance provided in Table 2 also gives a better understanding of the measure of dispersion related to wind speed at both stations. The Haripur site has a minimum variation as compared to the Gwadar site while having maximum skewness as well kurtosis. Both of the sites have positive skewness and high kurtosis, especially data of Haripur is leptokurtic.

Table 2. *Descriptive measures for wind speed at Gwadar and Haripur regions*

Stations	n	Min.	Max.	Mean	Variance	Skewness	Kurtosis
Gwadar	94377	0.0024	15.8959	4.3467	6.07052	0.77809	0.32391
Haripur	138121	0.2877	37.4073	4.7172	5.15917	1.09196	4.32162

3. METHODOLOGY

Literature showed that the WD is the most appropriate and effective distribution for the analysis of wind speed due to its flexible behavior towards positively skewed data sets. But it has also been observed that the WD is not always appropriate for wind speed analysis for all regimes. Therefore, the present study has explored some useful extensions of Lindley distribution for wind speed analysis and summarized the performances of these extended distributions.

The Lindly distribution was proposed by [22] to make a compound of gamma and exponential distribution as it is very useful to deal with lifetime data sets as well as for reliability analysis. Although it has also been observed that the Lindley distribution has inadequate results theoretically. Therefore, many generalizations of Lindley distribution have been proposed in the literature. GL proposed by [23], PL introduced by [24], EGL developed by [25], EQL suggested by [26]. The pdf and cdf of Lindley distribution are

$$g(z) = \frac{\beta^2}{\beta+1}(1+z)e^{-\beta z} \ for \ \beta > 0 \qquad (1)$$

$$G(z) = 1 - \frac{(\beta+1+\beta z)}{\beta+1}e^{-\beta z} \ for \ \beta > 0 \qquad (2)$$

where β is the scale parameter. The density functions of the considered probability distributions are provided in Table 3 for wind speed analysis.

Table 3. *Competitor distributions for modeling wind speed data in coastal regions of Pakistan with pdf*

Distribution	Density function
WD	$g(z) = \frac{\alpha}{\beta}\left(\frac{z}{\beta}\right)^{\alpha-1} e^{-\left(\frac{z}{\beta}\right)^{\alpha}} \quad for \quad \alpha, \beta > 0$
EGL	$g(z) = \frac{\alpha\theta^2\beta(1+\beta z)^{2\alpha-1} e^{\theta-\theta(1+\beta z)^{\alpha}}}{\theta+1} \quad for \ \alpha, \beta, \theta > 0$
EQL	$g(z) = \frac{\beta\theta(\alpha+\theta z)}{\alpha+1}\left(1-\left(\frac{\theta z}{1+\alpha}\right)e^{-\theta z}\right)^{\beta-1} e^{-\theta z} \ for \ \alpha, \beta, \theta > 0$
GL	$g(z) = \frac{\alpha\beta^2(1+z)}{\beta+1}\left(1-\left(\frac{1+\beta+\beta z}{1+\beta}\right)e^{-\beta z}\right)^{\alpha-1} e^{-\beta z} \ for \ \alpha, \beta > 0$
PL	$g(z) = \frac{\alpha\beta^2(1+z^{\alpha})}{\beta+1} z^{\alpha-1} e^{-\beta z^{\alpha}} \ for \ \alpha, \beta > 0$
RD	$g(z) = \frac{2z}{\beta^2} e^{-\left(\frac{z}{\beta}\right)^2} \quad for \quad \beta > 0$

4. ANALYSIS AND RESULTS

This section evaluates the performance of different extensions of Lindley distributions and other competitor distributions based on different goodness of fit criteria i.e. AIC, BIC, HQIC, KS test, RMSE, and R^2. The formulation of these criteria is given in Table 4.

Table 4. *The formulation for the evaluation of model fit*

Criteria	Formulation
AIC	$2k - 2\ln L$
BIC	$k \ln n - 2 \ln L$
HQIC	$-2\ln L + 2k \ln(\ln n)$
KS	$max_{1 \leq i \leq N}\left(G(v_i) - \frac{(i-1)}{N}, \frac{i}{N} - G(v_i)\right)$
RMSE	$\sqrt{\frac{\sum_{i=1}^{n}(\hat{v}_i - v_i)}{n}}$
R^2	$1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(y_i - z_i)^2}$

where k represents the parameters count, n is the count of observations in data while lnL represents the log-likelihood values of the anticipated model.

It should be observed that the minimum values of AIC, BIC, HQIC, and RMSE show the superiority of the considered model on the other competitor models. Moreover, the least statistic value of the KS test shows the betterment of the model. The superiority of the model also depends upon R^2 and lnL; higher values of these criteria have a better model fitting. The potentiality of the wind speed model is also observed on other criteria known as power density error (PDE). The formulation of this criterion is given as

$$PDE = \left|\frac{P_{ows} - P_M}{P_{ows}}\right| \tag{3}$$

where P_{ows} represented the average power density of wind depends on the actual count of wind data and P_M represented the average power density depend on the considered model. The formulation for P_{ows} and P_M are given respectively

$$P_{ows} = \frac{1}{2n}\rho A \sum_{i=1}^{n} x_i^3 \tag{4}$$

$$P_M = \frac{1}{2}\rho A \int_0^\infty x^3 f_D(x)dx. \tag{5}$$

Table 5. *Results for the evaluation of appropriate model through the goodness of fit measures (Gwadar)*

Models Criteria	Gwadar						
	EGL	EQL	GL	PL	WD	LD	RD
α	1.2892	0.4973	2.1692	1.3945	1.8499	-	-
β	6.2595	2.1152	0.5618	0.2161	4.9036	0.3949	4.9964
θ	0.0256	0.5630	-	-	-	-	-
$\ln L$	-210874.7	-211234.4	-211234.9	-211220.6	-210939.1	-220842.5	-211436.6
AIC	421749.4	422474.9	422473.9	422445.2	421882.1	441687.0	422875.2
BIC	421783.8	422503.3	422492.8	422464.1	421901.0	441696.5	422884.7
$HQIC$	421764.0	422483.5	422479.7	422451.0	421887.9	441689.9	422878.1
KS	0.014851	0.021572	0.021541	0.019014	0.016075	0.110880	0.045156
$RMSE$	0.003150	0.008921	0.008123	0.004671	0.004286	0.012675	0.009451
R^2	0.972319	0.906217	0.903482	0.931260	0.946815	0.764915	0.873521

Tables 5 and 6 have provided the estimated value for extensions of Lindley distribution as well as for Weibull and Rayleigh distribution for Gwadar and Haripur regions respectively. It has been observed from the considered goodness of fit criteria for these two coastal regions Pakistan that the EGL has a better fit than that of the well-known WD and other considered models. As the $\ln L$, AIC, BIC, HQIC, KS, and RMSE have the least values for the EGL distribution, therefore, it has significant behavior towards model fit.

Table 6. *Results for the evaluation of appropriate model through the goodness of fit measures (Haripur)*

Models Criteria	Haripur						
	EGL	EQL	GL	PL	WD	LD	RD
α	1.5524	0.4921	3.2983	1.6209	2.1816	-	-
β	5.2954	3.1516	0.6105	0.1365	5.3282	0.3670	5.2357
θ	0.0117	0.6127	-	-	-	-	-
$\ln L$	-300731.7	-301505.7	-301511.5	-301223.4	-302084.1	-329390.8	-302994.7
AIC	601469.5	603017.5	603027.1	602450.9	604172.3	658783.7	605991.4
BIC	601499.0	603047.0	603046.7	602470.6	604192.0	658793.5	606001.2
$HQIC$	601478.3	603026.3	603033.0	602456.8	604178.2	658786.6	605994.3
KS	0.019669	0.043225	0.043254	0.021644	0.022443	0.166380	0.042746
$RMSE$	0.002186	0.008900	0.009110	0.003198	0.003450	0.039700	0.008610
R^2	0.973400	0.870830	0.851030	0.926100	0.906130	0.865102	0.861531
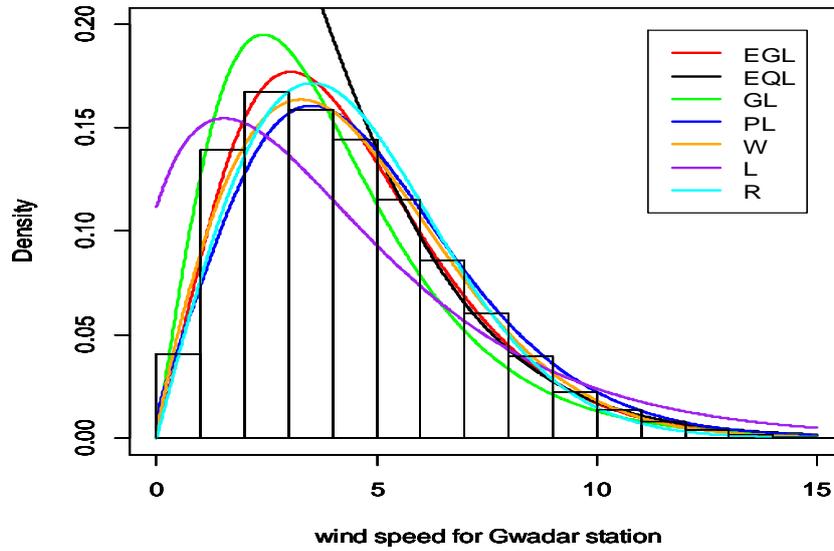
***Figure 2.** The fitted pdf of all distributions along with histogram for Gwadar region*

However, the fitting ability of all the considered models has presented in Figures 2 and 3 for better understandings. It has been seen for both coastal regions that the EGL distribution has the best fit as compared to the other observed wind speed distributions.
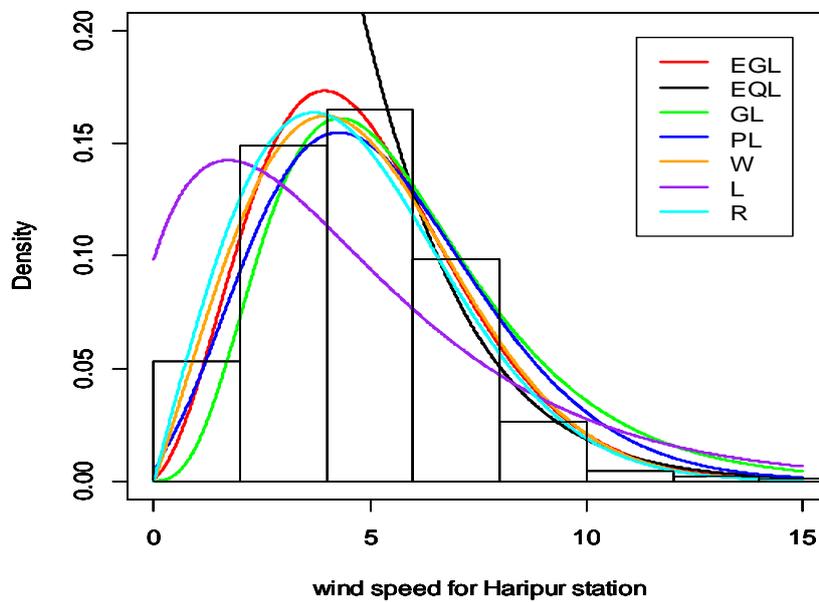
***Figure 3.** The fitted pdf of all distributions along with histogram for Haripur region*

Table 7. *Results for the PDE of both coastal regions (Gwadar and Haripur)*

	P_{ows}	P_{EGL}	P_{EQL}	P_{GL}	P_{PL}	P_{WD}	P_{LD}	P_{RD}
Gwadar	1330.312	1362.588	1643.298	203.1223	1328.555	1318.231	2361.519	3607.85
Haripur	1467.644	1423.123	1904.083	253.6326	1411.621	1422.452	2983.07	4151.463
	PDE_{EGL}	PDE_{EQL}	PDE_{GL}	PDE_{PL}	PDE_{W}	PDE_{L}	PDE_{R}
Gwadar	2.426	23.527	84.73	0.132	0.908	77.516	171.2033
Haripur	3.0335	29.7374	82.7184	3.8172	3.0792	107.2557	182.8658

The PDE values have demonstrated in Table 7 for all observed probability distributions. This criterion is used as another goodness of fit to observe the best suitable model of wind speed analysis. The model with the least PDE is considered the best-fitted model. The results indicated in Table 7 that the PL distribution has the least PDE as compared to the other considered distributions for the Gwadar region. However, EGL distribution is the most appropriate distribution for wind speed analysis of the Haripur region in terms of PDE values.

5.   CONCLUSION

It is seen that in the wind speed modeling literature, WD is not always the best distribution to model wind speed, especially in high and low, multi-modal, and skewed cases. For this reason, in this study, performances of some extensions of Lindley distribution that used earlier and not used in the modeling wind speed data are investigated in the two coastal cities of Pakistan for the first time. In this way, it is aimed to contribute to the renewable energy sector starting to develop new in Pakistan, in terms of statistical perspectives of wind energy with this paper.

The results of analyses in the study show that the EGL distribution is more appropriate than the WD and other competitor distributions to model wind speed data in Gwadar and Haripur due to the AIC, BIC, HQIC, KS, RMSE, and R^2 values. However, according to the PDE results, PL distribution is superior to the other distributions in estimating wind power in Gwadar.

As a result, it is concluded that extensions of Lindley distribution especially EGL and PL may use as alternatives to the WD to model and estimate wind energy in Haripur and Gwadar. In the wind energy projects which will be developed in these sites and other sites of Pakistan, the information presented in this study can be kept in front of the eye to assess the wind energy potential.

CONFLICTS OF INTEREST

No conflict of interest was declared by the authors.

REFERENCES

[1]   Khan, J. K., Shoaib, M., Uddin, Z., Siddiqui, I. A., Aijaz, A., Siddiqui, A. A., and Hussain, E., "Comparison of wind energy potential for coastal locations: Pasni and Gwadar", Journal of Basic & Applied Sciences, 11: 211–216, (2015).

[2]   Carta, J. A., Ramirez, P., and Velazquez, S., "A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands", Renewable and Sustainable Energy Reviews, 13(5): 933–955, (2009).

[3]   Akpinar, S., and Akpinar, E. K., "Estimation of wind energy potential using finite mixture distribution models", Energy Conversion and Management, 50(4): 877–884, (2009).

[4]   Keyhani, A., Ghasemi-Varnamkhasti, M., Khanali, M., and Abbaszadeh, R., "An assessment of wind energy potential as a power generation source in the capital of Iran, Tehran", Energy, 35(1): 188–201, (2010).

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

[6]   Chang, T. P., "Performance comparison of six numerical methods in estimating Weibull parameters for wind energy application", Applied Energy, 88(1): 272–282, (2011).

[7]   Usta, I., and Kantar, Y. M., "Analysis of some flexible families of distributions for estimation of wind speed distributions", Applied Energy, 89(1): 355–367, (2012).

[8]   Chellali, F., Khellaf, A., Belouchrani, A., and Khanniche, R., "A comparison between wind speed distributions derived from the maximum entropy principle and Weibull distribution. Case of study; six regions of Algeria", Renewable and Sustainable Energy Reviews, 16(1): 379–385, (2012).

[9]   Rocha, P. A. C., de Sousa, R. C., de Andrade, C. F., and da Silva, M. E. V., "Comparison of seven numerical methods for determining Weibull parameters for wind energy generation in the northeast region of Brazil", Applied Energy, 89(1): 395–400, (2012).

[10]   Kantar, Y. M., Usta, I., Arik, I., and Yenilmez, I., "Wind Speed Analysis Using the Extended Generalized Lindley Distribution", Renewable Energy, 30: 1-7, (2017).

[11]   Shoaib, M., Dar, I. S., Ahsan-ul-Haq, M., and Usman, R. M. "A sustainable generalization of inverse Lindley distribution for wind speed analysis in certain regions of Pakistan", Modeling Earth Systems and Environment, 1-13, (2021).

[12]   Ahsan-ul-Haq, M., Rao, G. S., Albassam, M., and Aslam, M. "Marshall–Olkin power Lomax distribution for modeling of wind speed data", Energy Reports, 6: 1118-1123, (2020).

[13]   De Andrade, C. F., Neto, H. F. M., Rocha, P. A. C., and da Silva, M. E. V., "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 Conversion and Management, 86: 801–808, (2014).

[14]   Bilir, L., Imir, M., Devrim, Y., and Albostan, A., "An investigation on wind energy potential and small scale wind turbine performance at İncek region–Ankara, Turkey", Energy Conversion and Management, 103: 910–923, (2015).

[15]   Arslan, T., Acitas, S., and Senoglu, B., "Generalized Lindley and Power Lindley distributions for modeling the wind speed data", Energy Conversion and Management, 152(8): 300–311, (2017).

[16]   Akgül, F. G., Şenoğlu, B., and Arslan, T., "An alternative distribution to Weibull for modeling the wind speed data: Inverse Weibull distribution", Energy Conversion and Management, 114: 234–240, (2016).

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

[18]   Shin, J. Y., Ouarda, T. B., and Lee, T., "Heterogeneous mixture distributions for modeling wind speed, application to the UAE", Renewable Energy, 91: 40–52, (2016).

[19]   Hossain, J., Sharma, S., and Kishore, V. V. N., "Multi-peak Gaussian fit applicability to wind speed distribution", Renewable and Sustainable Energy Reviews, 34: 483–490, (2014).

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

[21]   ul Haq, M. A., Chand, S., Sajjad, M. Z., and Usman, R. M., "Evaluating the suitability of two parametric wind speed distributions: a case study from Pakistan", Modeling Earth Systems and Environment, 1-9, (2020).

[22] Lindley, D. V., "Fiducial distributions and Bayes' theorem", Journal of the Royal Statistical Society: Series B, 102–107, (1958).

[23] Nadarajah, S., Bakouch, H. S., and Tahmasbi, R., "A generalized Lindley distribution", Sankhya B, 73(2): 331–359, (2011).

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

[25] Bhati, D., Malik, M., and Jose, K. K., "A new 3-parameter extension of generalized Lindley distribution", arXiv Prepr. arXiv1601.01045, (2016).

[26] Elbatal, I., Diab, L. S., and Elgarhy, M., "Exponentiated quasi Lindley distribution", International Journal of Reliability and Applications, 17(1): 1–19, (2016).