World Applied Sciences Journal 18 (11): 1533-1539, 2012
ISSN 1818-4952

DOI: 10.5829/idosi.wasj.2012.18.11.2040

Statistical Analysis of Wind Speed Data in Pakistan

[1]Kamran Abbas, [2]Alamgir, [3]Sajjad Ahmad Khan, [3]Amjad Ali, [3]Dost Muhammad Khan and [3]Umair Khalil

[1]Department of Statistics, University of Azad Jammu & Kashmir, Muzaffarabad, Pakistan
[2]Department of Statistics, University of Peshawar, Pakistan
[3]Department of Statistics, Islamia College Peshawar, Pakistan

Abstract: In the present study an effort has been made to find out the best fitting distribution of wind speed data recorded in Islamabad on a daily basis for the years 2001 to 2003. For this purpose two parameter Gamma, Weibull, Lognormal, Rayleigh, three parameter Burr and Frechet distributions are fitted to data and parameters for each distribution are estimated using the Maximum Likelihood (ML) method. The performance of the distributions are evaluated using three Goodness of Fit (GOF) tests namely Chi-Squared (CS), Kolmogorove-Smirnove (KS) and Anderson Darling (AD) test, Further, the fitted graphics of Cumulative Distribution Function (CDF), Probability Distribution Function (PDF) and Probability Probability (PP) plots are used to confirm the GOF for the above six distributions. Finally, graphical and GOF results are compared suggesting that Burr, Lognormal and Gamma distributions are found to be most appropriate as compared to the Weibull, Rayleigh and Frechet distributions.

Key words: Wind speed data . gamma distribution . weibull distribution . lognormal distribution . rayleigh distribution . burr distribution . frechet distribution . maximum likelihood method . goodness of fit tests

INTRODUCTION

Wind is renewable source of energy and is almost the fastest growing energy resource in the world, which is clean and offers many benefits to human beings. Wind energy has been getting immense attention because renewable energies has got tremendous focus. Due to increase in cost of fossil fuel and the various environmental problems, it is important to appreciate the potential of electricity generation from non-conventional sources. Recently, the demand of energy has increased in Pakistan and is likely to increase more in the near future. The effective use of wind energy is the conversion of wind power into valuable forms of electricity. The distribution of wind speed is important for power generators. Therefore, regionalized study according to statistical perspective is essential to model wind speed behavior.

Several studies have investigated the modeling of appropriate wind speed distribution. e. g. Zaharim *et al.* (2009) studied Weibull and Lognormal distributions to fit the wind speed data. They found that the Weibull distribution is more suitable as compared to Lognormal distribution to describe the behavior of wind speed data. Moreover, Akpinar and Akpinar (2004) discussed Weibull and Rayleigh distributions to describe the wind energy potential on the basis of 5 years hourly time series wind speed data. They concluded that the Weibull distribution provides better fit to probability distributions as compared to Rayleigh model. Further, Zaharim *et al.* (2009) analyzed the Malaysian wind speed data using statistical distributions. They applied Lognormal, Burr distribution and Frechet distributions to data collected from specific locality in Pahang, Malaysia. Based on the goodness of fit results and the graphical procedure, they concluded that Burr distribution provided best fit to the data. Moreover, Xiao *et al.* (2006) investigated various distributions to fit extreme wind speed data recorded in Hong Kong. They concluded that the Type-I and the three-parameter Weibull distributions are more suitable models as compared to the two-parameter Weibull distribution. Further, different researchers throughout the world provided the constructive applications of different distributions to wind speed data: see Celik (2004) for application in Turkey, Weisser (2003) for application in Grenada (West Indies), Akpinar and Akpinar (2005) for application in Keban-Elazig (Turkey), Ramirez and Carta (2005) for application in the Canarian Archipelago (Spain), Shata and Hanitsch (2005) for application in Egypt, Morgan *et al.* (2010) for application in United States. Wadagale *et al.* (2011) applied different distributions namely, Beta, Exponential, Gamma, Lognormal and Uniform. They compared Anderson Darling test with modified Anderson Darling test and found that Weibull

Corresponding Author: Alamgir, Department of Statistics, University of Peshawar, Pakistan

distribution is best fitted for wind speed data. The purpose of this paper is to find the most appropriate distribution(s) for describing the wind speed data. Therefore, we intend to compare Gamma, Weibull, Lognormal, Rayleigh, Burr and Frechet distributions by using different GOF criterion. Once a distribution function is assumed to be selected for present study, it remains to estimate its parameters from the sample data and to test the goodness of fit. In the present study we use the Maximum Likelihood (ML) method for estimating parameters of candidate distributions.

MATERIALS

About the data: Wind speed data for present study was obtained from Agricultural Meteorology Project, National Agricultural Research Council, Islamabad, starting 1st January, 2001 until 31th December 2003. The data was collected using a cup anemometer in km/day.

Methodology: In order to describe the behavior of wind speed at a particular area, it is required to identify the distribution (s), which best fit the data. In this study four two parameter distributions namely Gamma, Weibull, Lognormal, Rayleigh, two three parameter distributions, Burr and Frechet are used to model the distribution of wind speed along with different GOF tests namely, Chi-Squared (CS), Kolmogrove Simirnov (KS) and Anderson Darling (AD) test. The PDF and CDF for the candidate distributions are given as follows.

Gamma distribution: The PDF and CDF of two parameter Gamma distribution with shape parameter α and scale parameter β is given by:

$$f(x;\alpha,\beta)=\frac{x^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)}\exp[-x/\beta],\ \text{x>0, a, ß >0} \qquad (1)$$

and

$$F(x)=\frac{\Gamma(x/\beta)\alpha}{\Gamma(\alpha)} \qquad (2)$$

where $\Gamma(x)$ is the Incomplete Gamma function.

Weibull distribution: The PDF of Weibull distribution is give by

$$f(x;\alpha,\beta)=\frac{\alpha}{\beta}(\frac{x}{\beta})^{\alpha-1}\exp[-(\frac{x}{\beta})^{\alpha}],\ \text{x>0, a, ß >0} \qquad (3)$$

where α is the shape parameter and β is the scale parameter. Similarly, The CDF of Weibull distribution is given by

$$F(x)=1-\exp[-(\frac{x}{\beta})^{\alpha}] \qquad (4)$$

Lognormal distribution: The Lognormal distribution has the PDF and CDF given by

$$f(x;\mu,\sigma)=\frac{1}{x\sigma\sqrt{2\Pi}}\exp[-\frac{1}{2}(\frac{\ln x-\mu}{\sigma})^{2}].$$

x>0, s>0, -8 <μ<8 (5)

$$F(x)=\phi(\frac{\ln x-\mu}{\sigma}) \qquad (6)$$

where ϕ is the Laplace integral.

Rayleigh distribution: The Rayleigh density function, with parameters (σ) and (γ) is given by

$$f(x;\sigma,\gamma)=\frac{x-\gamma}{\sigma^{2}}\exp[-\frac{1}{2}(\frac{x-\gamma}{\sigma})^{2}]\ ,$$

x>0, s >0, -8 < ? <8 (7)

$$F(x)=1-\exp[-\frac{1}{2}(\frac{x-\gamma}{\sigma})^{2}] \qquad (8)$$

where (γ) is continuous location parameter and (σ) is the continuous scale parameter.

Burr distribution: The PDF and CDF of Burr distribution is given as:

$$f(x;\alpha,k,\beta)=\frac{\alpha k(\frac{x}{\beta})^{\alpha-1}}{\beta(1+(\frac{x}{\beta})^{\alpha})^{k+1}},\ \text{x>0, a, ß, k >0} \qquad (9)$$

$$F(x)=1-(1+(\frac{x}{\beta})^{\alpha})^{-k} \qquad (10)$$

where (α) and (k) are continuous shape parameters and (β) is the continuous scale parameter.

Frechet distribution: The PDF and CDF for Frechet distribution is

$$f(x;\alpha,\beta,\gamma)=\frac{\alpha}{\beta}(\frac{\beta}{x-\gamma})^{\alpha+1}\exp[-(\frac{\beta}{x-\gamma})^{\alpha}]$$

x>0, a, ß >0, -8 < ? <8 (11)

$$F(x)=\exp[-(\frac{\beta}{x-\gamma})^{\alpha}] \qquad (12)$$

Where α, β and γ are shape, scale and location parameters of Frechet distribution respectively.

Goodness of Fit (GOF) tests: In order to check how accurate a probability model fits the observed data, in

Table 1: Descriptive statistics of wind speed

Year	n	Mean	SD	Skewness	Kurtosis	Min.	Max.
2001	365	50.751	30.379	1.2246	1.7825	4.72	191.7
2002	365	58.537	35.787	1.4629	3.3492	5.18	252.6
2003	365	56.118	32.924	1.3916	2.9421	8.40	231.7

the present study, three GOF tests were applied at 5% level of significance to verify the GOF for the fitted distributions in which X represent the wind speed random variable and n sample size. The tests are described below.

Chi-Squared (CS) test: This test is used to determine whether the sample follows a particular probability distribution. The Chi-Squared statistic is defined as follows

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

where O_i denotes the observed frequency and E_i denotes the expected frequency and is calculated by

$$E_i = F(x_2) - F(x_1)$$

where F denotes the CDF of the distribution under consideration and x_1, x_2 are the lower and upper limits for bin i.

Kolmogorov-Smirnov (KS) test: This test is used to confirm the sample under consideration selected from a hypothesized continuous distribution. It is based on the largest vertical difference between the theoretical and the empirical CDF. Assume that we have a random sample x_1, x_2,…,x_n from some distribution with CDF $F(x_i)$. The KS test statistic,

$$D = \max(D^+, D^-)$$

where

$$D^+ = \max\left(\frac{i}{n} - F(x_i)\right)$$

and

$$D^- = \max\left(F(x_i) - \frac{i-1}{n}\right)$$

Anderson Darling (AD) test: AD test is generally used to compare the fit of an observed CDF to an expected CDF. This test gives higher weights to the tails than KS test and the test statistic is given by:

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1)[\ln F(x_i) + \ln(1 - F(X_{n-i+1}))]$$

RESULTS AND DISCUSSION

The descriptive statistics for wind speed data is presented in Table 1. It is revealed that the highest

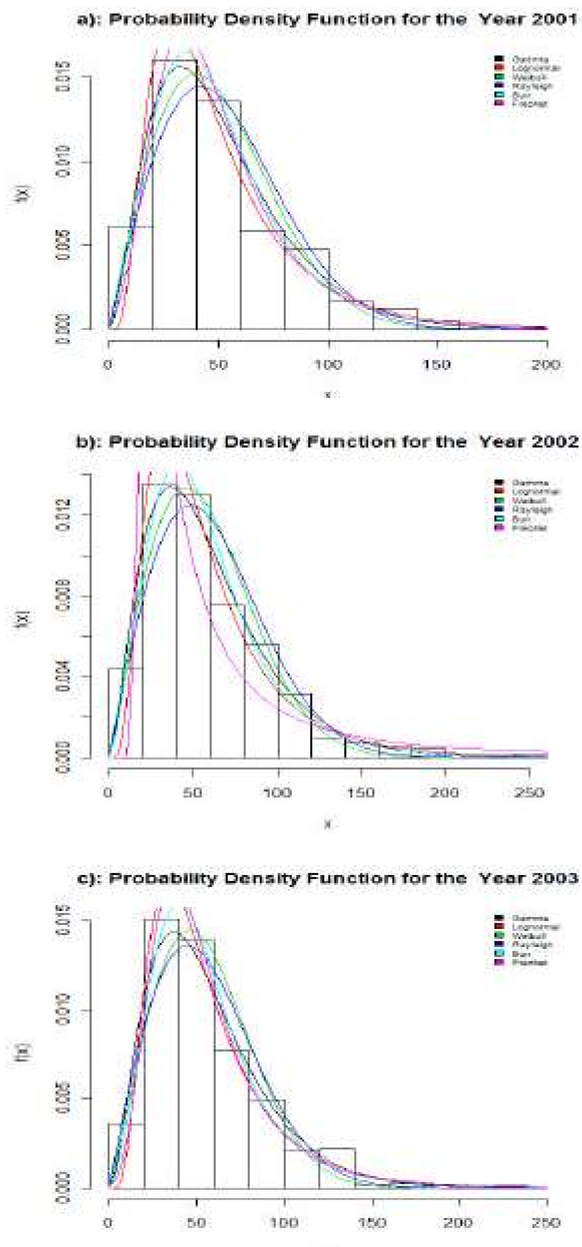


Fig. 1: Comparison of different probability densities of wind speed data

mean value is in 2002 with a standard deviation of 35.787 km/day. Coefficient of skewness is indicating that the distribution is itself an asymmetric; wind speed followed a positive extreme with a degree of skewness 1.2246, 1.4629 and 1.3916 for the year 2001, 2002 and

2003 respectively. This condition is also checked by coefficient of kurtosis, which gives a value of 1.7825, 3.3429 and 2.9421 for the year 2001, 2002 and 2003 respectively. The histograms give a graphical display of wind speed data which is presented in Fig. 1. These show that the distributions of wind speed is positively skewed i.e. the distribution curve is asymemetric being stretched out to the right. The other characteristic is that no wind speed can be less than zero km/day, thus the lower bound of distributions is zero. These results suggest that positively skewed distributions can be applied to model the wind speed data.

Fitting distributions: Gamma, Lognormal, Weibull, Rayleigh, Burr and Frechet distributions are fitted to wind speed data and parameters of these distributions are estimated by using ML method and results are presented in Table 2. To compare the probabilistic models, several goodness of fit: Kolmogrove-Smirnov test (1941), Simirnov (1944) Anderson Darling (AD) (1952) and Pearson's Chi-Squared (CS) test (1983) were used to quantify the best model. Table 3 shows the p-values of these GOF test. Results of KS, AD and CS test clearly indicated that Gamma, Lognormal and Burr distributions are well fitted whereas Frechet distribution is also a reasonable model for the year 2001 and 2003. Further p-values of GOF test are all less than 0.05 for Rayleigh distribution, indicating that data does not support the model. Moreover, KS, AD and CS p-values for Weibull distribution are also suggesting that there is no close agreement between wind speed data and Weibull distribution.

Table 2: Parameter estimates for six distribution models

Distribution		2001	2002	2003
Gamma	$\hat{\alpha}$	2.7909	2.6755	2.9051
	$\hat{\beta}$	18.1850	21.8790	19.3170
Lognormal	$\hat{\mu}$	3.7463	3.8903	3.8644
	$\hat{\sigma}$	0.6289	0.6167	0.5824
Weibull	$\hat{\alpha}$	1.9870	2.0177	2.1308
	$\hat{\beta}$	56.2350	64.6350	62.0520
Rayleigh	$\hat{\sigma}$	42.0210	48.5630	44.7410
	$\hat{\gamma}$	0.3482	0.1101	2.0583
Burr	$\hat{\alpha}$	2.2902	2.3319	2.5564
	$\hat{\beta}$	70.2080	76.7160	64.4350
	$\hat{k}$	2.2801	2.0977	1.7002
Frechet	$\hat{\alpha}$	8.9172	1.3409	6.6765
	$\hat{\beta}$	187.2800	29.9130	145.7400
	$\hat{\gamma}$	**0**- 151.19	4.9560	-105.7800

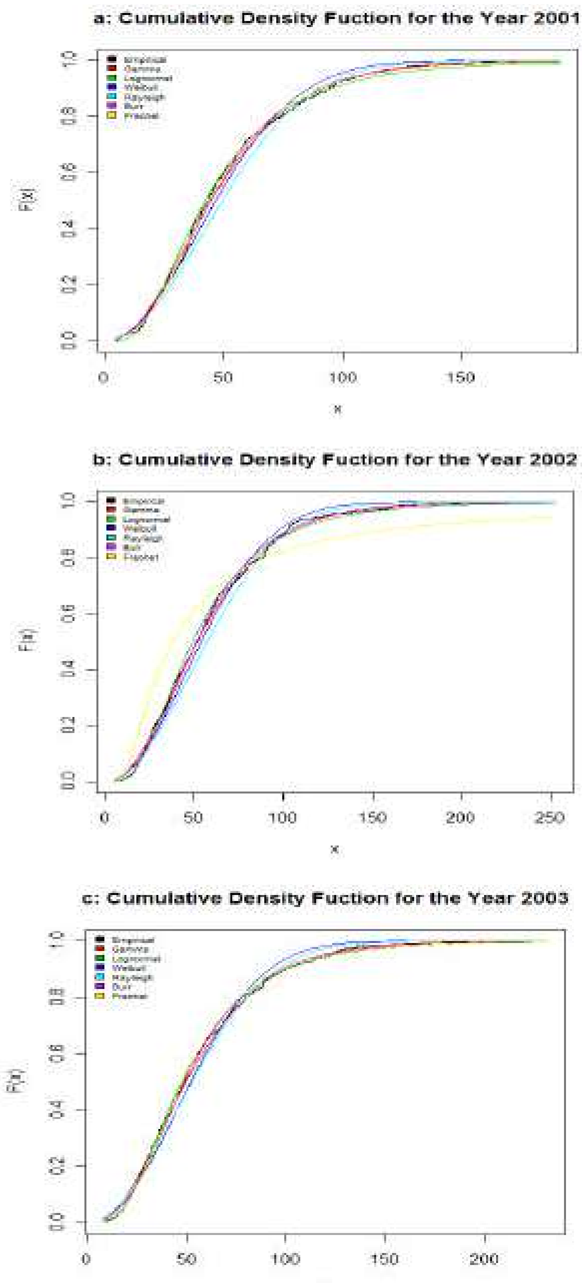


Fig. 2: Comparison of different cumulative densities of wind speed data

Figure 2 shows the CDF for the year 2001, 2002 and 2003 respectively. It can be seen from Fig. 2 that Gamma, Lognormal and Burr distributions are closer to the empirical distribution, indicating that these distributions are suitable for wind speed data, while Frechet distribution is not good only for the year 2002. In this paper that distribution is considered most suitable when at least two out of three GOF test best described the wind speed. Further, Rayleigh and Weibull distributions are departing from the wind speed empirical distribution. Finally, Probability Probability

Table 3: GOF test

	Gamma			Lognormal		
	KS	AD	CS	KS	AD	CS
Year	p-value	p-value	p-value	p-value	p-value	p-value
2001	0.7107	0.9110	0.1290	0.2121	0.7771	0.3573
2002	0.8176	0.5070	0.2783	0.5752	0.9000	0.1764
2003	0.7531	0.6072	0.7050	0.8648	0.1920	0.9670
	Weibull			Rayleigh		
	KS	AD	CS	KS	AD	CS
Year	p-value	p-value	p-value	p-value	p-value	p-value
2001	0.1061	0.0320	0.0484	0.0021	0.0047	0.0017
2002	0.1996	0.0083	0.1239	0.0061	0.0025	0.0070
2003	0.0544	0.0013	0.0089	0.0013	0.0043	0.0124
	Burr	Frechet				
	KS	AD	CS	KS	AD	CS
Year	p-value	p-value	p-value	p-value	p-value	p-value
2001	0.7634	0.9132	0.3048	0.7972	0.8470	0.1270
2002	0.5955	0.7640	0.4469	0.0598	0.0312	0.0071
2003	0.7766	0.9235	0.6620	0.7062	0.7743	0.9246

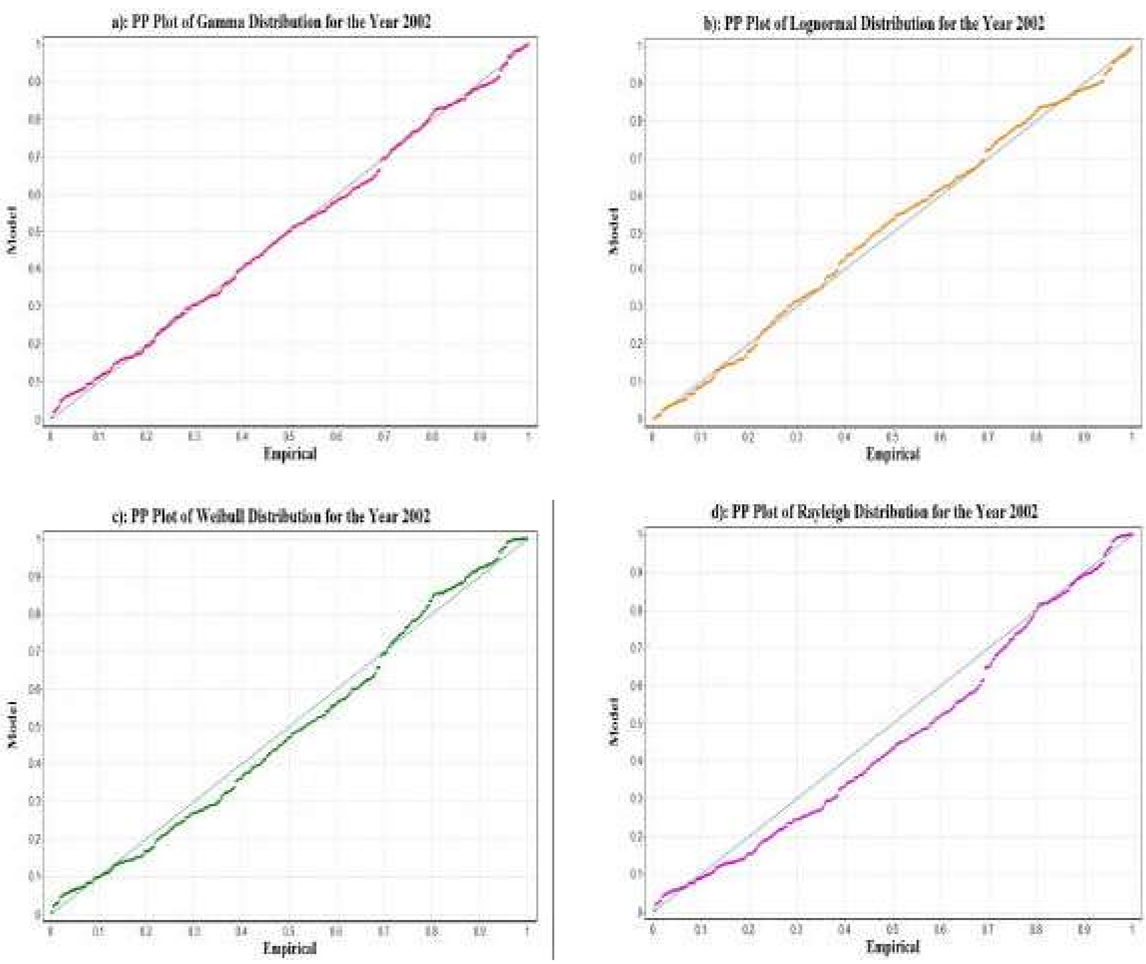
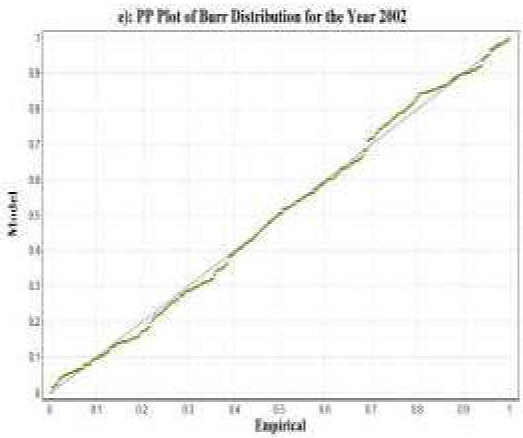


Fig. 3: Continued



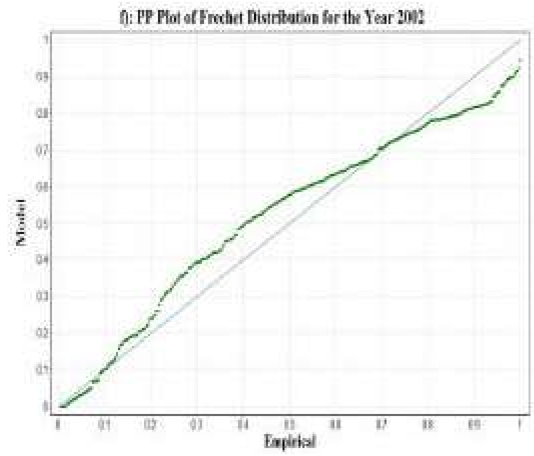


Fig. 3: Comparison of PP plots of wind speed data

(PP) plots are constructed in Fig. 3 for the year 2002 to see whether the hypothesized distributions are a suitable model for wind speed data and deviation from the observed data indicates incompatibility of the model. The plots indicate that Gamma, Lognormal and Burr distributions are most appropriate for wind speed data as compared to Frechet, Rayleigh and Weibull distributions. These plots and p-values of the three GOF test statistics are a final confirmation of the selected models.

CONCLUSION

The aim of this paper is to compare different probability models namely, Gamma, Lognormal, Weibull, Rayleigh, Burr and Frechet for wind speed data recorded in Pakistan. Three GOF tests were used to identify best fit at 5% level of significance. However, Gamma, Lognormal and Burr distributions are found to be more consistent in comparison to the other three distributions. These results are in agreement with the result obtained by Zaharim *et al*. (2009) and Morgan *et al*. (2010). According to Zaharim *et al*. (2009) Burr distribution is well fitted for wind speed in Peninsular Malaysia. Similarly, Morgan *et al*. (2010) concluded that 2-parameter Lognormal distribution is best for estimating extreme wind speeds in United States. Moreover, in the present study Weibull and Rayleigh distributions are found to be inappropriate for wind speed. These results differ from the result obtained by Zaharim *et al*. (2009), Al Buhairi (2006) and Akpinar and Akpinar (2004).

REFERENCES

1. Anderson, T.W. and D.A. Darling, 1952. Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. Annals of Mathematical Statistics, 23: 193-212.
2. Akpinar, E.K. and S. Akpinar, 2004. Statistical analysis of wind energy potential on the basis of the Weibull and Rayleigh distributions for Agin-Elazig, Turkey. Journal of Power and Energy, 218: 557-565.
3. Akpinar, E.K. and S. Akpinar, 2005. A statistical analysis of wind speed data used in installation of wind energy conversion systems. Energy Conversion and Management, 46 (4): 515-532.
4. 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 (4): 593-604.
5. Kolmogorov, A.N., 1941. Confidence limits for an unknown distribution function. Annals of Mathematical Statistics, 12: 461-463.
6. Morgan, E.C., M. Lackner, R.M. Vogel and L.G. Baise, 2010. Probability distributions for offshore wind speeds. Energy Conversion and Management, 52: 15-26.
7. Plackett, R.L., 1983. Karl Pearson and the Chi-Squared test. Int. stat. Rev., 51: 59-72.
8. Ramrez, P. and J.A. Carta, 2005. The use of wind probability distributions derived from the maximum entropy principle in the analysis of wind energy. A case study. Energy Conversion and Management, 47 (15-16): 2564-2577.
9. Shata, A.A.S. and R. Hanitsch, 2005. Evaluation of wind energy potential and electricity generation on the coast of Mediterranean sea in Egypt. Renewable Energy, 31 (8): 1183-1202.
10. Simirnov, N.V., 1944. Approximate laws of distribution of random variables from empirical data. Uspehi Matem. Nauk, 10: 179-206.
11. Wadagale, A.V., P.V. Thatkar, R.K. Dase and D.V. Tandale, 2011. Modified Anderson Darling test for wind speed data. International Journal of Computer Science and Emerging Technologies, 2 (2): 249-251.

12. Weisser, D., 2003. A wind energy analysis of Grenada: an estimation using the Weibull density function. Renewable Energy, 28 (11): 1803-1812.
13. Xiao, Y.Q., Q.S. Li, Z.N. Li, Y.W. Chow and G.Q. Li, 2006. Probability distribution of extreme wind speed and its occurrence interval. Engineering Structures, 28 (8): 1173-1181.
14. Zaharim, A., S.K. Najid, A.M. Razali and K. Sopian, 2009. Analyzing Malaysian wind speed data using statistical distribution. In Proceedings of the 4th IASME /WSEAS International Conference on Energy and Environment. University of Cambridge, February, pp: 24-26.
15. Zaharim, A., A.M. Razali, R.Z. Abidin and K. Sopian, 2009. Fitting of statistical distributions to wind speed data in Malaysia. European Journal of Scientific Research, 26 (1): 6-12.