Seasonal and continual wind speed modelling for the coastal urban city, Karachi

NAEEM SADIQ

Institute of Space and Planetary Astrophysics, University of Karachi, Karachi 75270, Pakistan

(*Received 10 October 2017, Accepted 22 January 2018*)

e mail : nsadiq@uok.edu.pk

सार – पवन गति में परिवर्तनशीलता से न सिर्फ दाब प्रणाली के मजबूत होने या कमजोर पडने का पता चलता है बल्कि तटीय क्षेत्र के इर्द-गिर्द पवन क्षेत्र में उसकी महत्वपूर्ण भूमिका का भी पता चलता है। पवन गति ऑकडा की समय श्रृंखला के द्वारा संभावित वितरण सबसे पहले कथित पैरामीटरों के लिए आधारभूत आवश्यकता को भी पूरा कर देती है। अनुसंधानात्मक ऑकड़ो के विश्लेषण से यह पता चलता है कि तटीय महानगर कराची में मानसून वर्षा ऋतु के दौरान अधिकतम पवन गति (~ 23 m/s) होती है और मानसूनोतर काल में अपने अधिकतम अंतर (~ 3.5 m/s) के साथ अपने चरम पर होती है। वसंत ऋतु और मानसूनोतर काल में मीन/ट्रिमड मीन (~ 11.5 m/s) होता है साथ ही साथ ये मानसून पूर्व और मानसून (~ 18.5 m/s) में लगभग एक सामान होते है जबकि शीत ऋतु और मानसूनोतर काल में जब पवन सबसे धीरे चलती है वह भी एक सामान (~ 6 m/s) ही पतझड़ और शीतऋतु में न्यूनतम मानक अंतर होते हैं। महत्वपूर्ण और सांख्यकीय मान की तुलना वितरण मॉडलिंग के साथ की गई है। जबकि विभिन्न मौसमी और निरंतर वितरण की पैरामीटरी मान का भी आकलन किया गया है। इस अध्ययन में संवहनी वितरण कार्य प्रणाली और संभाव्य संभावित प्लॉट की भी सहायता ली गई है। पवन गति मॉडलिंग के लिए वीबुल वितरण का उपयोग करना असामान्य नहीं है। तटीय स्टेशन कराची के लिए पवन गति के प्रतिदिन के डाटा समय श्रृंखला का उपयोग जैसे कि वेकबाए और सामान्यीकृत चरम मान वितरण की तुलना करते हुए यह पता चला है कि अन्य जटिल मॉडल ज्यादातर उपयोग में लाए जाने वाले वीबुल वितरण ने अपेक्षाकृत वितरण के खराब परिणाम दिए है।यह पाया गया है कि केवल मानसून पूर्व ग्रीष्म को छोड़कर, जिसके बाद शहर में सामान्यीकृत चरम मान वितरण (GEV) आता है, वार्षिक और मौसमी पवन वेकबाए वितरण के बाद आती है। कोई निरंतर और/ या मौसमी पवन गति अंततः तथा / या अधिक सही ढंग से वीबुल वितरण के बाद नही आती है। यह अध्ययन विमानन और पवन इंजीनियरिंग कार्यों को कुछ नए दृष्टिकोण देगा।

ABSTRACT. Variation in wind speed not only indicates the strengthening or weakening of pressure systems but its role in wind farm in the vicinity of coastal area is also crucial. Probability distributions through time series of wind speed data serves foremost basic need for the said parameters. Exploratory data analysis revealed that for coastal city Karachi, maximum wind speed (~23 m/s) occurred during monsoon with its peak during postmonsoon with maximum deviation (~3.5 m/s). Mean / trimmed mean during spring and postmonsoon (~11.5 m/s) as well as in premonsoon and monsoon (~18.5 m/s) remain almost identical while minimum wind blowing during winter and postmonsoon are also identical (~6 m/s). Autumn and winter exhibits least standard deviations. Critical and statistical values have been compared for distribution modelling, while parametric values of different seasonal and continual distributions are also estimated. The study is supported by cumulative distribution functions and probability-probability plots. It is not uncommon to use Weibull distribution for wind speed modelling. By using daily data time series of wind speed for the coastal station Karachi, it has been explored that widely accepted Weibull distribution provides comparatively poor distribution results when compared to other more complicated models (*i.e.*, Wakeby and generalized extreme value distributions]. It is found that annual and seasonal wind comes after the Wakeby distribution except premonsoon summer which follows the generalized extreme value distribution (GEV) for the city. No continual and / or seasonal wind speed follows the Weibull distribution, ultimately and / or more appropriately. The study may give some new insights for aviation and wind engineering purposes.

Key words – Exploratory data analysis; GEV distribution; Wakeby distribution; Weibull distribution; Wind modelling.

1. Introduction

Wind is considered as one of the inexpensive potential source of energy. Wind energy is renewable - in contrast to conventional power stations (Basumatary *et al.*, 2005). Alternate energy production or power generation is dependent on distribution of wind speed (Tuzuner and Yu, 2008). Hence, appropriate distribution modelling of wind
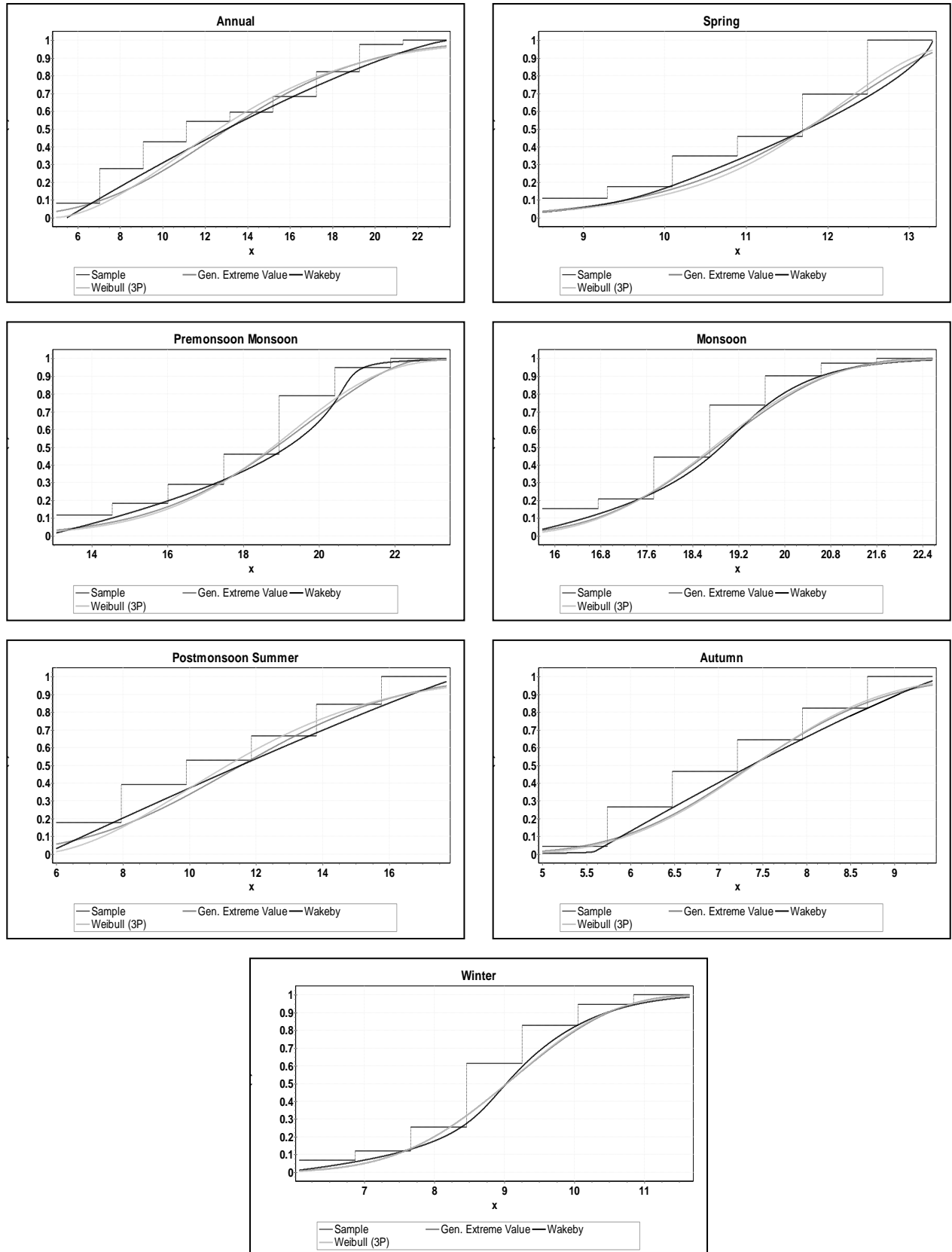
Fig. 1. Cumulative distribution function for different probabilities for Karachi

TABLE 1

Basic EDA statistical parameters

Season	Mean	Tr_{mean}	Maximum	Minimum	St. dev	Range
Winter	8.993	9.007	11.643	6.071	1.166	5.572
Spring	11.498	11.558	13.286	8.500	1.384	4.786
Pre-monsoon summer	18.497	18.532	23.357	13.071	2.513	10.286
Monsoon	18.760	18.748	22.571	15.786	1.538	6.785
Post-monsoon summer	11.646	11.622	17.714	6.000	3.540	11.714
Autumn	7.411	7.425	9.429	5.000	1.152	4.429

speed is the primary need for the close estimation of energy production. In general, Weibull (W3) distribution is the recognized modelling irrespective of the geographical location of station (Yu and Tuzuner, 2008; Edwards and Hurst, 2001; Archer and Jacobson 2003 and 2005).

In Pakistan, wind speed distribution plays a crucial role in forecasting storms and planning of coastal wind farms (Ahmed *et al*., 2006). Through distribution modelling of the wind data for the coastal urban city Karachi it is revealed that wind speed measurements may not inevitably W3.

This paper describes the wind speed distribution modelling as an alternative to estimate energy production which is dependent on the wind distribution. Therefore, accurate distribution modelling is the initial step to attain the estimation of the accurate alternate energy production.

The Shape factors have also been estimated. While, after fitting the close distribution set [*viz*., W3, Generalized Extreme Value (GEV) and Wake by (W5) distributions] over the coastal urban city it is shown that wind distributions may not be most appropriately follow the W3 distribution, for instance, in the case of Karachi.

2. Materials and method

2.1. *Data*

The available data set of daily wind speed from 1997 to 2014 has been utilized in this study; obtained from the archives of Pakistan Meteorological Department.

2.2. *GEV distribution model*

Extreme Value Theory (EVT) has been reckoned over Extremal Type Theorem (ETT) such that the rescaled sample maxima converge in distribution to a variable having distribution, possibly within any one of the Gumbel, Frechet and Weibull (also called Type I, Type II and Type III) families, respectively (Sadiq, 2015). The merger of Gumbel, Frechet and Weibull families comes out as

$$G(z) = \exp\left[-\left\{ 1 + \xi\left(\frac{z - \mu}{\sigma} \right) \right\}^{-1/\xi} \right]$$

defined on the set $\{z : 1 + \xi(z - \mu)/\sigma > 0\}$, where the parameters satisfy $-\infty < \mu < \infty, \sigma > 0$ and $-\infty < \xi < \infty$. This is the GEV distribution family comprises of three different parameters *viz*., μ, σ, and ξ known as location, scale and shape parameters, respectively. Gumbel, Frechet and Weibull corresponds to ξ = 0, ξ > 0 and ξ < 0, respectively.

2.3. *Wakeby (W5) distribution model*

If the probability density functions (pdf) exhibit a characteristic heavy tail then it can be better modelled by W5 distribution, as our results (Fig. 1) shown, that this distribution provides markedly good fits. The PDF of the distribution may be determined by Johnson (1994)

$$f(x) = \frac{[1 - F(x)]^{\delta+1}}{\gamma + \alpha[1 - F(x)]^{\beta+\delta}}$$

TABLE 2

Critical and statistic values for annual and different seasons regarding considered distributions

Status	Critical value	Wakeby (St. Val.)	GEV (St. Val.)	Weibull (St. Val.)
Annual	0.07108	0.07056	0.11589	0.11721
Spring	0.19625	0.06603	0.08497	0.11346
Pre-monsoon summer	0.15342	0.09804	0.08249	0.10044
Monsoon	0.15755	0.04786	0.06268	0.06964
Post-monsoon summer	0.18659	0.05466	0.09836	0.11046
Autumn	0.19837	0.08184	0.09792	0.10933
Winter	0.0563	0.04963	0.07941	0.08909

TABLE 3

Parametric values of different seasonal and continual distributions

Status	Wakeby	GEV	Weibull
Annual	$\alpha = 14.04$ $\beta = 0.77975$ $\gamma = 0$ $\delta = 0$ $\xi = 5.5013$	$\kappa = 0.18181$ $\sigma = 4.7201$ $\mu = 11.396$	$\alpha = 1.7621$ $\beta = 9.5953$ $\gamma = 4.8461$
Spring	$\alpha = 40.785$ $\beta = 25.3$ $\gamma = 6.2521$ $\delta = -1.4666$ $\xi = 7.413$	$\kappa = -0.59221$ $\sigma = 1.5527$ $\mu = 11.217$	$\alpha = 4.0419\text{E}+7$ $\beta = 4.3721\text{E}+7$ $\gamma = -4.3721\text{E}+7$
Pre-monsoon summer	$\alpha = 19.087$ $\beta = 2.4647$ $\gamma = 0.0674$ $\delta = 0.71141$ $\xi = 12.755$	$\kappa = -0.54852$ $\sigma = 2.7617$ $\mu = 17.937$	$\alpha = 8.9863$ $\beta = 19.522$ $\gamma = 0.03277$
Monsoon	$\alpha = 13.3$ $\beta = 4.5068$ $\gamma = 1.1752$ $\delta = -0.1109$ $\xi = 15.287$	$\kappa = -0.3469$ $\sigma = 1.5882$ $\mu = 18.263$	$\alpha = 3.4107$ $\beta = 5.1851$ $\gamma = 14.103$
Post-monsoon summer	$\alpha = 11.491$ $\beta = 0.91587$ $\gamma = 0$ $\delta = 0$ $\xi = 5.6475$	$\kappa = -0.24556$ $\sigma = 3.5571$ $\mu = 10.303$	$\alpha = 1.8462$ $\beta = 7.123$ $\gamma = 5.3072$
Autumn	$\alpha = 2620.3$ $\beta = 472.57$ $\gamma = 3.5075$ $\delta = -0.86749$ $\xi = 0$	$\kappa = -0.25468$ $\sigma = 1.1587$ $\mu = 6.981$	$\alpha = 3.1307$ $\beta = 3.5796$ $\gamma = 4.2146$
Winter	$\alpha = 20.217$ $\beta = 8.4475$ $\gamma = 1.2638$ $\delta = -0.23928$ $\xi = 5.8336$	$\kappa = -0.35669$ $\sigma = 1.2073$ $\mu = 8.6226$	$\alpha = 4.3261$ $\beta = 4.8855$ $\gamma = 4.5446$

where, $F(x)$ is the CDF with α, β, γ, δ shape parameters. The inverse CDF of the W5 may be given by

$$x(F) = \xi + \frac{\alpha}{\beta}[1 - (1 - F)]^\beta - \frac{\gamma}{\delta}[1 - (1 - F)]^{-\delta}$$

with the following conditions or restrictions that must apply among the various parameters

$\gamma \geq 0$ and $\alpha + \gamma \geq 0$

If $\alpha = 0$ then $\beta = 0$

If $\gamma = 0$ then $\delta = 0$

either $\alpha \neq 0$ or $\gamma \neq 0$

while parametric domain comprises of

$\xi \leq x < \infty$ if $\delta \geq 0$ and $\gamma > 0$

$\xi \leq x \leq \xi + \frac{\alpha}{\beta} - \frac{\gamma}{\delta}$ if $\delta < 0$ or $\gamma = 0$

The above parameterization is due to Hosking (1986) which is unlike that used by Landwehr *et al*. (1979). In point of fact, the parameterization by Sen (2009) presents the W5 distribution as an extension of the Generalized Pareto Distribution (GPD) that provides the guesstimates of the more stable parameters under small disorderly data (Landwehr *et al*., 1979). In order that $x(F)$ in the equation Marsh and Dale (2002) represents an inverse CDF, the conditions $\gamma \geq 0$ and $\gamma + \alpha \geq 0$ should be followed. As W5 is of supple nature, it can be utilized for the description of
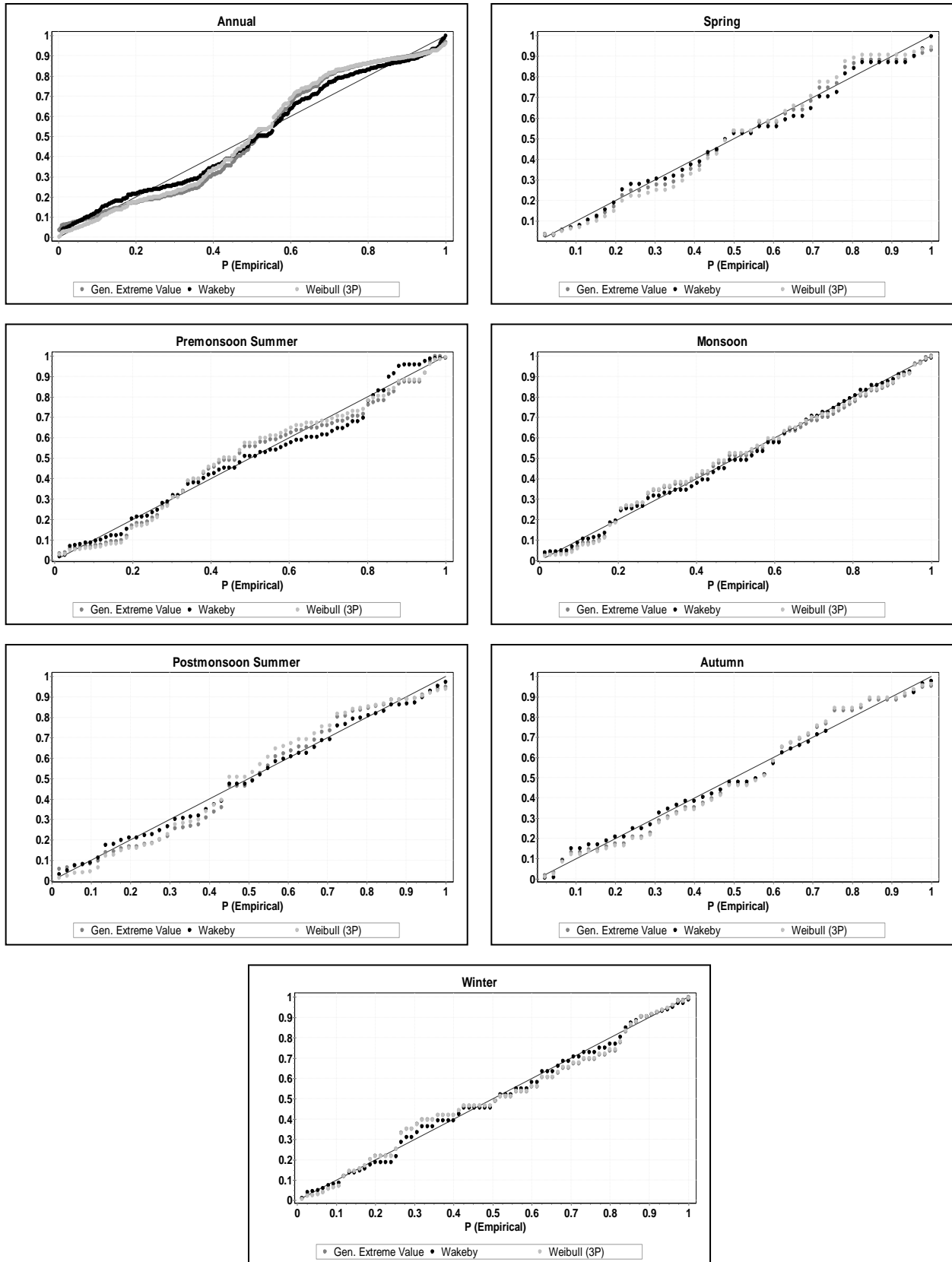
Fig. 2. PP plot for the different probabilities for Karachi

natural processes accompanied with multiple factors which should or else be modelled through the combination of more than a few distributions.

2.4. *Weibull (W3) distribution*

A random variable X is said to have a W3 with parameters α and β ($\alpha > 0, \beta > 0$) if the pdf of X is

$$f(x; \alpha, \beta) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-\left(\frac{x}{\beta}\right)^\alpha} \text{ if } x \geq 0$$

otherwise the function is 0 if $x < 0$

When, $\alpha = 1$, the pdf becomes

$$f(x; \beta) = \frac{1}{\beta^\alpha} e^{-x/B} \text{ if } x \geq 0$$

otherwise the function is 0 if $x < 0$

which is the pdf for an exponential distribtion with parameter $\lambda = \frac{1}{B}$ (Celik, 2004). Hence, it is concluded that the exponential distribution is the special case of Weibull distibutions.

2.5. *Testing hypothesis*

2.5.1. *Goodness of fit*

Test statistics using Kolmogorov-Smirnov technique has been applied to check that which distribution is statistically a good fit to the wind speed data. It is observed that for annual data other distributions like GEV and Weibull are unsuitable at our chosen significant level (α) of 0.05 level. Only W5 distribution shows that its critical value 0.07108 is greater than the static value 0.7056. Hence, for annual wind speed W5 distribution is also tested and verified through test statistics. All the goodness of fit test values and critical values are summarized in Table 2. Though, in seasonal cases, distributions may be drawn from other models but even then the statistic values suggests the W5 remains the best fitted distribution. Premonsoon is the exceptional case in which GEV distribution appears to be even better fitted than the W5.

3. Results and discussion

3.1. *Exploratory data analysis*

The trimmed mean is the specific method of averaging that remove the percentage of the largest and smallest values before calculating the mean. After removing the specified observation, the trimmed mean may be just by using a simple arithmetic average formula. The trimmed mean of wind is highest (18.748 m/s) in monsoon (which is also comparable to wind speed in Premonsoon summer) and lowest (7.425 m/s) in transitional autumn. Similarly, second summer (11.622 m/s) and transitional spring (11.558 m/s) have also identical values. All the values are summarized in Table 1.

Standard deviation (SD) is the measure of the dispersion of a set of data from its mean. The more spread apart the data the higher the deviation. While the low standard deviation shows the data values are nearer to the mean value. The maximum (*i.e.*, the most deviated observed value from the mean) and minimum SD has been observed during postmonsoon and autumn, respectively. The maximum and minimum Range has also been observed for postmonsoon and autumn, respectively.

3.2. *Best fitted distribution modelling*

As regards the seasonal distribution, seasonal classification is assorted as winter (Mid of December to February), spring (March to Mid April), premonsoon summer (mid April to June), monsoon (July to first denary of September), postmosnoon summer (second denary of September to October), and autumn (November to mid of December) as suggested by Sadiq (2009). It is evident that except winter annual analysis of wind speed, models may be drawn from other than the W5 distribution. The test statistics values decided that W5 distribution is the best for spring, monsoon, postmonsoon summer, autumn and winter seasons. However, for premonsoon summer, GEV distribution statistics is more appropriate (Table 2). Shape parameters have also been estimated for all the considered models for comparison and summarized in Table 3.

The cumulative distribution function (CDF) for W5, GEV and W3 distributions in respect of annual and seasonal winds have been plotted in Fig. 1. It is evident that W5 (*i.e.*, black) curve remain in consistent and close to the cumulative *steps* as compared to W3 (light grey) and GEV (dark grey). Though, other two distributions are also appears closed but visually it is clear that W5 is the closest, hence appears to be as the best fitted model.

3.3. *Validation of the model*

The plot of p (empirical) against p (model) has also been employed to check the suitability of the fitted models (Fig. 2). The graph should be ideally linear (or close to be

linear) if the chosen distribution is the correct choice. The supplementary graphs of p-p (*i.e.*, plot against empirical to theoretical probability) also show that W5 remain closer to the ideal line while W3 and GEV distributions fluctuate more than the W5. Therefore, it is confirmed that W5 is the most appropriate modelling for the considered data set.

4. Conclusions

The study undertook the EDA and modelling of continual and seasonal wind speed for the coastal urban city, Karachi. Premonsoon and monsoon seasons are found as the most wind blowing seasons for mean, maximum and minimum values while least values are generally observed in winter. The SD (*i.e.*, the observed value deviated with respect to mean value) and Range are maximum in premonsoon. There is no significant difference between Mean and Trimmed mean. As regards wind speed, Weibull distribution has not been found the best model when compared to more complicated models like GEV and Wakeby distributions. The continual wind at Karachi cannot be modelled at 95 per cent confidence but only with Wakeby distribution among the undertaken models. The seasonal wind has also been found as best fitted through W5 distribution with the exception of premonsoon when GEV appears to be more appropriate. The model has also been validated through p-p plots in addition to Kolmogorov-Smirnov test. Weibull distribution has failed to model the continual speed at the high chosen confidence level and also not found the most appropriate for seasonal wind modelling regarding the coastal urban city, Karachi. The study may be further extended by estimating the wind power generation in correspond to wind mill dimensions through wind potential at this coast.

5. Abbreviations

EDA : Exploratory Data Analysis

SD : Standard Deviation

GEV : Generalized Extreme Value

W5 : Wakeby

W3 : Weibull

EVT : Extreme Value Theory

ETT : Extremal Type Theorem

pdf : Probability distribution function

CDF : Cumulative Distribution Function

pp : (empirical) probability - (theoretical) probability

References

Ahmed, M. A., Ahmad, F. and Akhtar, W., 2006, "Assessment of wind power potential for coastal areas of Pakistan", *Turk. J. Phys.*, **30**, 127-135.

Archer, C. and Jacobson, M., 2003, "Spatial and temporal distributions of U.S. winds and wind power at 80 m derived from measurements", *Geophysical Research*, **108**, D9.

Archer, C. and Jacobson, M., 2005, "Evaluation of global wind power", *Geophysical Research*, **110**, D12110, doi:10.1029/2004JD005462.

Basumatary, H., Sreevalsan, E. and Sasi, K. K., 2005, "Weibull parameter estimation - A comparison of different methods", *Wind Engineering*, **29**, 3, 309-316.

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.

Edwards, P. and Hurst, R., 2001, "Level-crossing statistics of the horizontal wind speed in the planetary surface boundary layer", *Chaos*, **11**, 3, 611-618.

Hosking, J. R. M., 1986, "The Wakeby Distribution", *Research Report*, RC12302, IBM Research Division, York town Heights, New York.

Johnson, N. L., Kotz, S. and Balakrishnan, N., 1994, "Continuous Univariate Distributions", Wiley, New York.

Landwehr, J. M., Matalas, N. C. and Wallis, J. R., 1979, "Estimation of parameters and quantiles of Wakeby distributions", *Water Resources Research*, **15**, 6, 1361-1372.

Marsh, T. J. and Dale, M., 2002, "The U K floods of 2000/01 - A hydrometeorological Appraisal", *CIWEM Journal*, **16**, p180.

Sadiq, N., 2009, "Development of a Weather-Type Classification Scheme for Karachi by using Multivariate Techniques", *Pakistan journal of Meteorology*, **5**, 10, 65-79.

Sadiq, N., 2015, "Modeling of Rianfall Extremes through the application of Extremal Type Theorem (ETT)", *Mausam*, **66**, 4, 803-808.

Sen, A. K., 2009, "Complexity analysis of river-flow time series", *Stochastic Environmental Research and Risk Assessment*, **23**, p361.

Tuzuner, A. and Yu, Z., 2008, "A Theoretical Analysis on Parameter Estimation for the Weibull Wind Speed Distribution", in Proc. IEEE PES General Meeting, July 20-24, Pittsburgh.

Yu, Z. and Tuzuner, A., 2008, "Wind Speed Modeling and Energy Production Simulation with Weibull Sampling", PES General Meeting.