

Comparison of Weibull Distribution and Exponentiated Weibull Distribution Based Estimation of Mean and Variance of Wind Data

Debanshee Datta¹ and D. Datta²

¹Department of Instrumentation Engineering, A.C. Patil College of Engineering
Mumbai University, Navi Mumbai – 410210

²Health Physics Division, Bhabha Atomic Research Centre
Mumbai – 400085

debansheedatta@yahoo.co.in, ddatta@barc.gov.in

Abstract

Wind speed is the most important parameter in the design and study of the environmental impact assessment. The wind speed determines the concentration of any toxic pollutant released from any chemical industry. As the wind speed increases the concentration of the toxic pollutant at any location in the environment at any instant of time decreases. It is also known that as the wind speed increases the cost of the wind energy is reduced. Literature study reveals that the probability distribution of wind speed can be presented by Weibull distribution and it has been accepted so without any statistical investigations. In this study, it is being proposed that similar to Weibull distribution, a better fitting of wind speed data is possible by Exponentiated Weibull distribution. Statistical investigations as an evidence of this proposition via estimation of mean and variance of wind speed data by fitting through both Weibull and Exponentiated Weibull distributions are presented in this paper. A comparison of the representative estimates through respective probability distributions is also presented in this study. Estimation of parameters (shape and scale) of both the distributions is presented by using maximum likelihood algorithm.

Keywords: *Weibull distribution, Exponentiated Weibull distribution, Maximum likelihood*

1. Introduction

Environmental impact assessment of any industry is always carried out with respect to the estimated concentration of the toxic pollutant at any instant of time and at any spatial location in the environment. Impact of the toxic effluent on the atmospheric environment if released into the air depends not only on the quantity released but also on the wind speed. If wind speed increases then the concentration of the toxic effluent decreases at any spatio-temporal location and hence impact is reduced. Amount of material released into the environment can be obtained from the plant; however, knowing the exact wind speed (crisp value of the wind speed) for a span of time is very uncertain due to fluctuation of the wind speed for the specified period of observations. Therefore, knowledge of the wind speed is very important component for safety analysis of the specified industry. Evidence based on a large number of literature study indicates that wind speed is always characterized by a variety of standard parametric probability density functions (PDF) [1]. Weibull distribution is always selected as the best distribution for fitting wind speed data not only due to its greater flexibility and simplicity but also due to its highest value of goodness of fit to the experimental data [1, 2]. The basic requirement for safety analysis is to know the characteristics value of the wind speed (e.g., mean and variance of the wind speed). In order to obtain such values a prior

fitting of the wind speed data is essential and this is generally carried out mostly by a two parameter Weibull distribution [3, 4]. However, from the point of improvement in wind data fitting, similar to the Weibull family, a new family of distributions, namely Exponentiated Weibull distribution [5] is being proposed; in this context it is customary to say that a typical two parameter Weibull distribution does not accurately represent the entire wind regime in nature [6]. The Exponentiated Weibull distribution has been compared with the two parameter Weibull distribution. Parameters of the Exponentiated Weibull distribution and classical Weibull distribution are estimated using maximum likelihood method algorithm [8].

In this paper, we have described the Exponentiated Weibull distribution with its many properties. Maximum likelihood based estimation of the parameters of this new distribution and that of Weibull distribution has been presented in detail. Finally, a comparison of the estimates of the mean and variance of the wind data fitted through Weibull and Exponentiated Weibull distributions are presented.

2. Weibull and Exponentiated Weibull Distribution

In this section, we mainly introduce the Weibull distribution based model and Exponentiated Weibull distribution based model for the analysis of wind speed.

2.1 Weibull distribution

The probability density function (PDF) and cumulative distribution function (CDF) of a random variable v representing wind speed and having the Weibull distribution is defined by

$$f_W(v; \gamma, \beta) = \frac{\gamma}{\beta} \left(\frac{v}{\beta}\right)^{\gamma-1} \exp\left[-\left(\frac{v}{\beta}\right)^\gamma\right], \quad v > 0 \quad (1)$$

and

$$F_W(v; \gamma, \beta) = 1 - \exp\left[-\left(\frac{v}{\beta}\right)^\gamma\right] \quad (2)$$

respectively; where $\gamma > 0$ is a shape parameter and $\beta > 0$ is a scale parameter, that depends on γ , and is related to the mean value of wind speed. The two Weibull parameters and the mean wind speed are related by

$$\bar{v} = \beta \Gamma\left(1 + \frac{1}{\gamma}\right) \quad (3)$$

where $\Gamma(\cdot)$ is the gamma function.

It is easily proved that the n^{th} wind speed moment of the Weibull PDF is given by

$$\langle v^n \rangle = \beta^n \Gamma\left(1 + \frac{n}{\gamma}\right) \quad (4)$$

where the angular bracket $\langle \cdot \rangle$ denotes the expectation and $\Gamma(\cdot)$ represents the gamma function.

Based on Eq.(1) and Eq.(4), the variance of the wind speed is given by

$$\sigma_v^2 = \frac{\Gamma(1+2\gamma)}{\Gamma(1+1/\beta)^2} - 1 \approx \gamma^{-11/6} \quad (5)$$

For the derivation of the scale parameter β , without loss of generality, it is assumed that the $\langle v \rangle = 1$ and setting $n = 1$ in Eq. (4), yields

$$\beta = \frac{1}{\Gamma(1+1/\gamma)} \quad (6)$$

2.2 Exponentiated Weibull distribution

Exponentiated Weibull distribution (EW) has a scale parameter and two shape parameters. The PDF and CDF of a random variable v described by the EW distribution in [7] is given by

$$f_{EW}(v; \gamma, \beta, \alpha) = \frac{\alpha\gamma}{\beta} \left(\frac{v}{\beta}\right)^{\gamma-1} \exp\left[-\left(\frac{v}{\beta}\right)^\gamma\right] \left\{1 - \exp\left[-\left(\frac{v}{\beta}\right)^\gamma\right]\right\}^{\alpha-1} \quad (7)$$

and

$$F_{EW}(v; \gamma, \beta, \alpha) = \left\{1 - \exp\left[-\left(\frac{v}{\beta}\right)^\gamma\right]\right\}^\alpha \quad (8)$$

respectively; where $\gamma > 0$ is a shape parameter related to the wind speed, and $\beta > 0$ is a scale parameter, that depends on γ and is related to the mean wind speed as in the precedent case. The additional parameter, compared with Eq. (2), $\alpha > 0$ is an extra shape parameter that gives more versatility to the EW distribution in the shape of the tails. Figure 1 shows that the density function of EW is unimodal and for fixed value of β and γ it becomes more and more symmetric as α increases. It can be easily noted that for $\alpha = 1$ Eq.(7) reduces to the Weibull distribution (*e.g.*, Eq.(2)). If we calculate the n th moment of the EW PDF, it can be easily mention that the analytical derivation of the EW parameters is rather a complete task. Therefore, a heuristic approach, based on simulation data was used to obtain first approximation to the EW parameters.

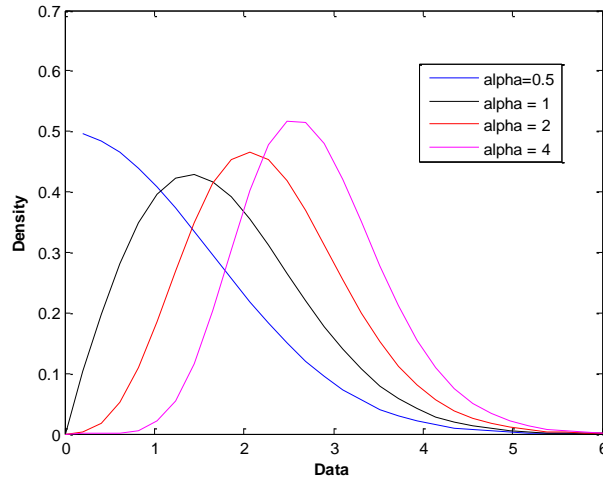


Figure 1. PDF of Exponentiated Weibull distribution with $\beta=2$, $\gamma=2$ when $\alpha=0.5$, 1, 2, 4

The shape parameter γ can be related with the random variable v as

$$\gamma \cong (\alpha \sigma_v^2)^{-6/11} \quad (9)$$

where σ_v^2 denotes the variance of the random variable v . The scale parameter β can be written as

$$\beta = \frac{1}{\alpha \Gamma(1+1/\gamma) \sum_{p=0}^{\infty} \frac{(-1)^p (p+1)^{-(1+\gamma)/\gamma} \Gamma(\alpha)}{p! (\alpha-p)}} \quad (10)$$

It can be easily verified that for fixed values of the shape parameter γ and the scale parameter β , the shape parameter α controls the lower-tail steepness – when data is visualized in a logarithmic scale. This is an attractive property of the EW distribution; it is precisely the lower-tail of maximum importance because it defines the error rate and fades probability. Let us now define the moments of the EW distribution. The k -th moment of the exponentiated Weibull variable, v , with distribution function given in Eq. (8) [7] is written as

$$E(v^k) = \alpha \beta^k \Gamma\left(\frac{k}{\gamma} + 1\right) \sum_{i=0}^{\alpha-1} \binom{\alpha-1}{i} (-1)^i (i+1)^{\frac{k}{\gamma}-1}, \quad \text{if } \alpha \in N \quad (11 \text{ a})$$

$$E(v^k) = \alpha \beta^k \Gamma\left(\frac{k}{\gamma} + 1\right) \sum_{i=0}^{\alpha-1} \frac{{}_{\alpha-1}P_i}{i!} (-1)^i (i+1)^{\frac{k}{\gamma}-1}, \quad \text{if } \alpha \notin N, \text{ for } k=0,1,2,\dots \quad (11 \text{ b})$$

where ${}_aP_i = a(a-1)(a-2) \dots (a-i+1)$ and N is the set of natural numbers. Since Eqs. (11a and 11b) are a convergent series for all $k \geq 0$, all moments exist.

The expectation value of the EW random variable v can be written as by using Eqs. (11a) and (11b) as

$$E(v) = \begin{cases} \alpha \beta \Gamma\left(\frac{1}{\gamma} + 1\right) \sum_{i=0}^{\alpha-1} \binom{\alpha-1}{i} (-1)^i (i+1)^{-\frac{1}{\gamma}-1}, & \text{if } \alpha \in N \end{cases} \quad (12a)$$

$$E(v) = \alpha \beta \Gamma\left(\frac{1}{\gamma} + 1\right) \sum_{i=0}^{\alpha-1} \frac{\alpha-1 P_i}{i!} (-1)^i (i+1)^{-\frac{1}{\gamma}-1}, \quad \text{if } \alpha \notin N \quad (12 b)$$

In a similar way, the variance of the EW random variable v can be written as

$$\sigma_v^2 = E(v^2) - [E(v)]^2 \quad (13)$$

where,

$$E(v^2) = \begin{cases} \alpha \beta^2 \Gamma\left(\frac{2}{\gamma} + 1\right) \sum_{i=0}^{\alpha-1} \binom{\alpha-1}{i} (-1)^i (i+1)^{-\frac{2}{\gamma}-1}, & \text{if } \alpha \in N \end{cases} \quad (14a)$$

$$E(v^2) = \alpha \beta^2 \Gamma\left(\frac{2}{\gamma} + 1\right) \sum_{i=0}^{\alpha-1} \frac{\alpha-1 P_i}{i!} (-1)^i (i+1)^{-\frac{2}{\gamma}-1}, \quad \text{if } \alpha \notin N \quad (14b)$$

3. Methodology of Maximum Likelihood Estimation

3.1 Weibull Distribution Parameters

Maximum likelihood is a method of estimation of parameters of a distribution. Maximum likelihood technique [9-11] with many required features is the most widely used technique among parameter estimation techniques. The abbreviation MLE may refer to maximum likelihood estimation (the method), to the estimate, or to the estimator. The method finds a value of the parameter that maximizes the likelihood function. The MLE method has many large sample properties that make it attractive for use. It is asymptotically consistent, which means that as the sample size gets larger, the estimates converge to the true values. It is asymptotically efficient, which means that for large samples, it produces the most precise estimates. It is asymptotically unbiased, which means that for large samples, one expects to get the true value on average. The estimate themselves are normally distributed if the sample is large enough. These are all excellent large sample properties.

Let x_1, x_2, \dots, x_n be a random sample of size n drawn from a probability density function $f_x(x; \theta)$ where θ is an unknown parameter. The likelihood function of this random sample is the joint density of the n random variables and is a function of the unknown parameter. Thus,

$$L = \prod_{i=1}^n f_{x_i}(x; \theta) \quad (15)$$

is the likelihood function. The maximum likelihood estimator (MLE) of θ , say $\hat{\theta}$, is the value of θ that maximizes L or, equivalently, the logarithm of L . Often, but not always, the MLE of θ is a solution of

$$\frac{d \log L}{d \theta} = 0 \quad (16)$$

where solutions those are not functions of the sample values x_1, x_2, \dots, x_n are not admissible, nor are solutions which are not in the parameter space. Now, we are going to apply MLE to estimate the Weibull parameters, namely the shape and the scale parameters. Consider the Weibull PDF given in (1), then the likelihood function will be

$$L(x_1, \dots, x_n; \gamma, \beta) = \prod_{i=1}^n \left(\frac{\gamma}{\beta}\right) \left(\frac{x_i}{\beta}\right)^{\gamma-1} e^{-\left(\frac{x_i}{\beta}\right)^\gamma} \quad (17)$$

On taking the logarithms of Eq. (17), differentiating with respect to γ and β in turn and equating to zero, we obtain the estimating equations

$$\frac{\partial \ln L}{\partial \gamma} = \frac{n}{\gamma} + \sum_{i=1}^n \ln x_i - \frac{1}{\beta} \sum_{i=1}^n x_i^\gamma \ln x_i = 0 \quad (18)$$

$$\frac{\partial \ln L}{\partial \beta} = -\frac{n}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^n x_i^\gamma = 0 \quad (19)$$

On eliminating β between these two equations and simplifying, we have

$$\frac{\sum_{i=1}^n x_i^\gamma \ln x_i}{\sum_{i=1}^n x_i^\gamma} - \frac{1}{\gamma} - \frac{1}{\beta} \sum_{i=1}^n \ln x_i = 0 \quad (20)$$

which may be solved to get the estimates of $\widehat{\mu}_k = \gamma$. This can be accomplished by the use of standard iterative procedures (i.e. Newton-Raphson method). Once γ is determined, β can be estimated using Eq. (21) as

$$\beta = \frac{\sum_{i=1}^n x_i^\gamma}{n} = 0 \quad (21)$$

3.2 Estimates of EW Parameters

In order to estimate the parameters of the EW distribution, we adopt the maximum likelihood method. Let v_1, v_2, \dots, v_n be a random sample of size n from the EW distribution given by Eq.(8). Now the log likelihood function can be written as

$$L(\alpha, 1/\beta, \gamma) = n \ln \alpha + n \ln \gamma + n\gamma \ln(1/\beta) + (\gamma - 1) \sum_{j=1}^n \ln v_j + (\alpha - 1) \sum_{j=1}^n \ln[1 - \exp\{-(v_j/\beta)^\gamma\}] - \sum_{j=1}^n (v_j/\beta)^\gamma \quad (22)$$

Now, by taking the derivative of L , MLE estimates of $\alpha, 1/\beta, \gamma$ that maximizes Eq. (22), normal equations can be written as

$$\frac{d}{d\alpha} L(\alpha, 1/\beta, \gamma) = \frac{n}{\alpha} + \sum_{j=1}^n \ln[1 - \exp\{-(v_j/\beta)^\gamma\}] = 0 \quad (23)$$

$$\frac{d}{d(1/\beta)} L(\alpha, 1/\beta, \gamma) = n\gamma\beta + (\alpha - 1)\gamma(1/\beta)^{\gamma-1} \sum_{j=1}^n \frac{e^{-(v_j/\beta)^\gamma}}{1 - e^{-(v_j/\beta)^\gamma}} v_j^\gamma - \gamma(1/\beta)^{\gamma-1} \sum_{j=1}^n v_j^\gamma = 0 \quad (24)$$

$$\begin{aligned} \frac{d}{d\gamma} L(\alpha, 1/\beta, \gamma) &= \frac{n}{\gamma} + n \ln(1/\beta) + \sum_{j=1}^n \ln v_j \\ &+ (\alpha - 1)(1/\beta)^\gamma \sum_{j=1}^n \frac{\exp\{-(v_j/\beta)^\gamma\}}{1 - \exp\{-(v_j/\beta)^\gamma\}} v_j^\gamma \ln(v_j/\beta) - (1/\beta)^{\gamma-1} \sum_{j=1}^n v_j^\gamma \ln(v_j/\beta) = 0 \end{aligned} \quad (25)$$

From Eq. (23) we obtain the MLE of α as a function of $(1/\beta, \gamma)$, say $\hat{\alpha}(1/\beta, \gamma)$ given by

$$\hat{\alpha} = \hat{\alpha}(1/\beta, \gamma) = -\frac{n}{\sum_{j=1}^n \ln[1 - \exp(-(v_j/\beta)^\gamma)]} \quad (26)$$

Multiplying Eq. (24) by $(1/\beta\gamma)$ we get

$$n + (1/\beta)^\gamma [(\alpha - 1) \sum_{j=1}^n \frac{\exp\{-(v_j/\beta)^\gamma\}}{1 - \exp\{-(v_j/\beta)^\gamma\}} v_j^\gamma - \sum_{j=1}^n v_j^\gamma] = 0 \quad (27)$$

Subtracting $\ln(1/\beta)$ times Eq. (27) from Eq. (25) we have

$$\frac{n}{\gamma} + \sum_{j=1}^n \ln v_j + (1/\beta)^\gamma [(\alpha - 1) \sum_{j=1}^n \frac{\exp\{-(v_j/\beta)^\gamma\}}{1 - \exp\{-(v_j/\beta)^\gamma\}} v_j^\gamma \ln v_j - \sum_{j=1}^n v_j^\gamma \ln v_j] = 0 \quad (28)$$

Using Eq.(26) in Eqs.(27) and (28) we get two equations, which are satisfied by the MLEs $(1/\hat{\beta})$ and $\hat{\gamma}$ of $(1/\beta)$ and γ , respectively.

4. Results and Discussion

In this study mean and variance of the wind speed data has been estimated. The yearly maximum wind speed data, in miles/hour, used in this study has been quoted from Castillo (1988) [12] and is presented in Table 1. Before the estimation, Weibull and Exponentiated Weibull distributions have been fitted independently. Fitted probability distributions are shown in Figures 2 and 3. Since fitting a probability distribution through a set of relevant data is nothing but the estimation of the parameters of the respective probability distributions, maximum likelihood based estimated value of the parameters of the corresponding probability distributions with their standard error are presented in Table 2. Mean value of the wind speed and the corresponding variance of the wind speed on the basis of Weibull distribution are estimated as 33.82 and 198.64 respectively. Further, the mean and variance of the wind speed based on the exponentiated Weibull distribution are estimated as 31.16 and 32.92 respectively.

Table 1. Yearly maximum wind speed data

| Serial No | Wind Data (miles per hour) | | | | |
|-----------|----------------------------|-------|-------|-------|-------|
| 1 | 22.64 | 22.80 | 23.75 | 24.01 | 24.04 |
| 2 | 24.24 | 24.74 | 25.45 | 25.55 | 25.66 |
| 3 | 25.99 | 26.63 | 26.69 | 26.88 | 26.89 |
| 4 | 27.12 | 27.43 | 27.69 | 27.71 | 28.12 |
| 5 | 28.58 | 28.88 | 29.12 | 29.45 | 29.48 |
| 6 | 30.18 | 31.31 | 31.55 | 31.57 | 32.54 |
| 7 | 32.98 | 33.83 | 33.86 | 34.64 | 35.21 |
| 8 | 36.82 | 37.23 | 38.09 | 38.26 | 38.82 |
| 9 | 38.96 | 38.90 | 42.99 | 43.66 | 44.61 |
| 10 | 45.24 | 47.91 | 54.75 | 69.40 | 98.16 |

Table 2. Parameters of the fitted distributions

| Distribution | Parameters | | |
|--|---|---------------------|-----------------|
| | Shape (std error) | Scale (std error) | -Log Likelihood |
| Weibull (beta, alpha) | Beta = 2.57 (0.23) | Alpha = 38.09 (2.2) | 394.7 |
| Exponentiated Weibull (alpha, beta, gamma) | Alpha = 91.27 (71.69), Gamma = 0.90 (0.15) | Beta = 5.50 (2.46) | 357.4 |

The third sub-column heading “-Log likelihood” under column heading “Parameters” as presented in Table 1 denotes the negative logarithm of the maximum likelihood. Thus it follows by the standard likelihood ratio test that the exponentiated Weibull distribution is a much better fitted distribution than the classical Weibull distribution for fitting of the wind speed data. This observation has been confirmed by the probability plots corresponding to the two fits as shown in Figure 4 and Figure 5.

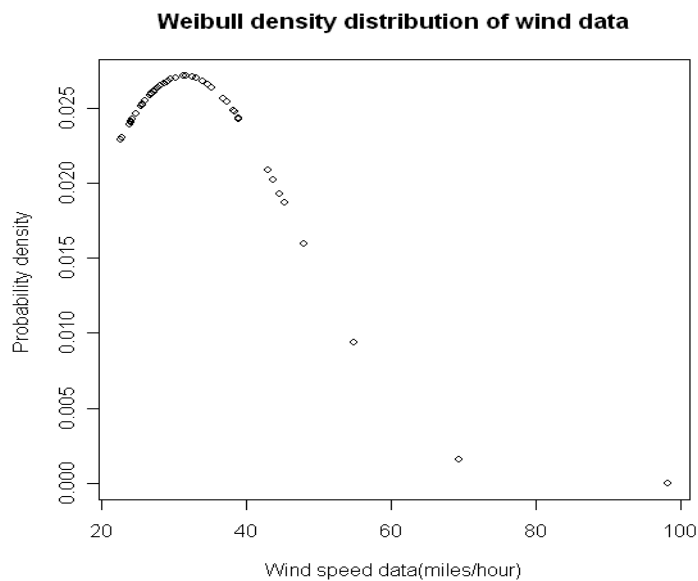


Figure 2. Fitted Weibull density Distribution of wind speed data

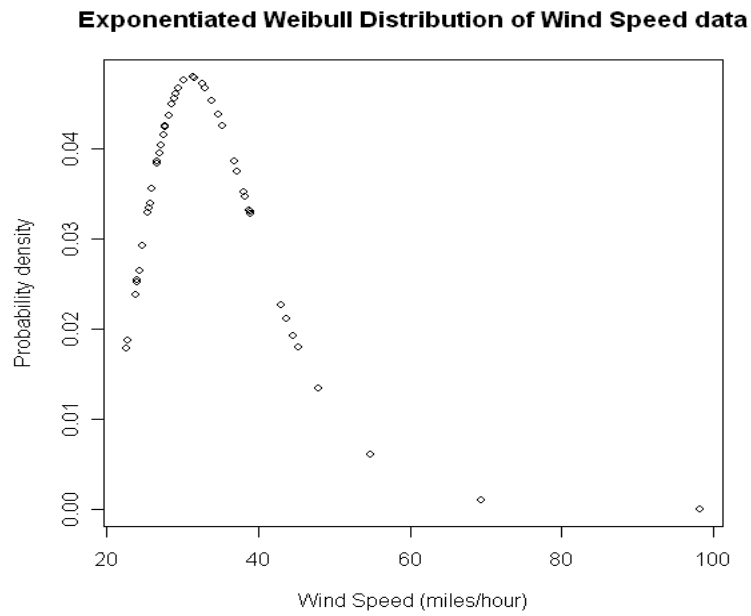


Figure 3. Fitted Exponentiated Weibull density distribution of wind speed data

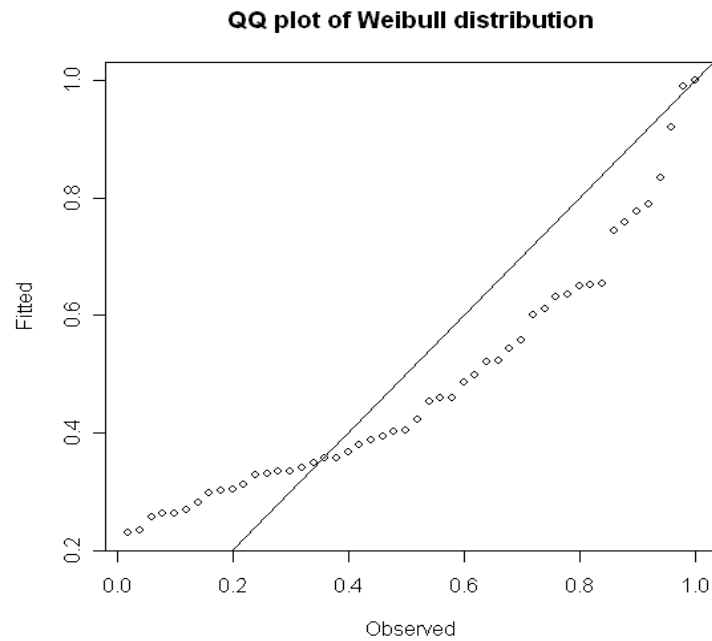


Figure 4. QQ plot of Weibull Distribution of wind speed data

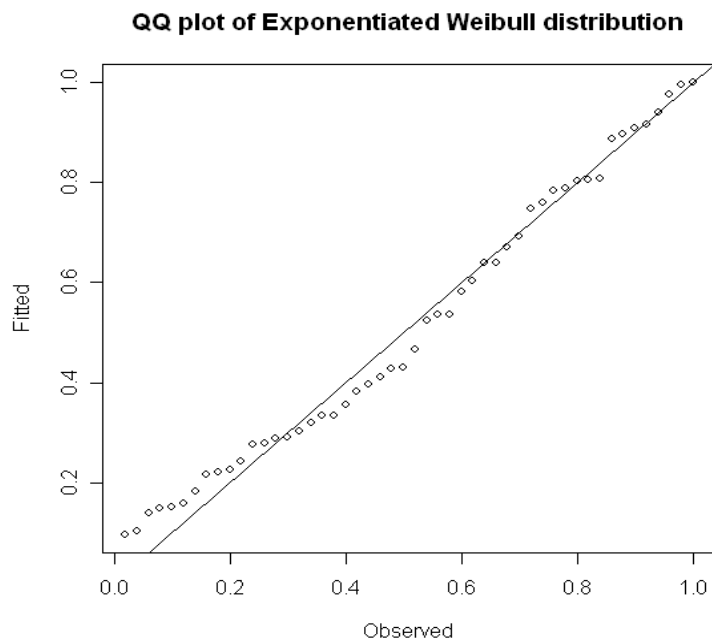


Figure 5. QQ plot of Exponentiated Weibull Distribution of wind speed data

References

- [1] J. A. Carta, P. Ramlrez and S. Velquez, "A Review of wind speed probability distributions used in wind energy analysis case studies in the Canary Islands", *Energy Conversion and Management*, vol. 13, (2009), pp. 933-955.
- [2] A. S. S. Dorvlo, "Estimating wind speed distribution", *Energy Conversion and management*, vol. 43, (2007), pp. 2311-2318.
- [3] J. A. Carta and P. Ramrez, "Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions", *Renewable energy*, vol. 32, (2007), pp. 518-531.
- [4] J. A. Carta and P. Ramrez, "Use of finite mixture distribution models in the analysis of wind energy in the canarian Archipelago", *Energy Conversion & Management*, vol. 48, (2007), pp. 281-291.
- [5] R. D. Gupta and D. Kundu, "Exponentiated exponential family: an alternative to gamma and Weibull", *Biometrical Journal*, vol. 43, no. 1, (2001), pp. 117-130.
- [6] Q. Meng and M. Tang, "On asymptotic distributions of yearly maximum values of surface temperature and wind speed over Chengdu and the estimation of their parameters", *Journal of Chengdu Institute of Meteorology*, vol. 12, (2007), pp. 284-291.
- [7] G. S. Mudholkar and D. K. Srivastava, "Exponentiated Weibull family for analyzing bathtub failure rate data", *IEEE Transactions on Reliability*, vol. 42, (1993), pp. 299-302.
- [8] A. Dempster, N. Laird and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *Jornal of the Royal Statistical Society Series B*, vol. 39, no. 1, (1977), pp. 1-38.
- [9] H. L. Harter and A. H. Moore, "Point and interval estimators based on order statistics, for the scale parameter of a Weibull population with known shape parameter", *Technometrics*, vol. 7, no. 3, (1965) August.
- [10] H. L. Harter and A. H. Moore, "Maximum likelihood estimation of the parameters of Gamma and Weibull populations from complete and from censored samples", *Technometrics*, vol. 7, no. 4, (1965) November.
- [11] A. C. Cohen, "Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples", *Technometrics*, vol. 7, no. 4, (1965) November.
- [12] E. Castillo, "Extreme Value Theory in Engineering", Academic Press, INC.(London), Harcourt Brace Jovanovich Publishers, ISBN 0-12-163475-2, (1988), pp. 334.

Authors



Debanshee Datta received the BE degree in Instrumentation Engineering from Mumbai University, India in the year 2011. She was associated with the Department of Instrumentation Engineering in Smt. Indira Gandhi College of Engineering, affiliated under Mumbai University in the capacity of a Lecturer and gained her teaching experience in various fields of control systems and data communications. At present she is pursuing her PhD from Indian Institute of Technology, Banaras Hindu University, Varanasi. Her research interest is on control theory.

E-mail: : debansheedatta@yahoo.co.in



D. Datta received the M.Sc degree in Nuclear Physics and M.Phil degree in high energy nuclear physics from University of Kolkata in 1983 and in 1984 respectively. He has graduated in nuclear engineering from 28th batch of BARC training school in 1984-85. He has received his Ph.D. degree in Computer Science from Mumbai University in 2000. At present he is working in the Health Physics Division of Bhabha Atomic Research Centre, Mumbai, Department of Atomic Energy, India in the capacity of senior scientist, Head, Computational Radiation Physics Section and the Professor of Homi Bhabha National Institute, Deemed University under the Department of Atomic Energy, India. He is the recipient of Eminent Scientist Award “Millennium Plaques of Honour” from the Indian Science Congress Association, in the year 2010. He has got more than 100 publications to his credit. He is a professional software developer and developed many software for various use in radiological safety, science and engineering. The science interests and competence field is soft computing, statistical and mathematical modeling, reliability, uncertainty and sensitivity analysis of systems.

E-mail: ddatta@barc.gov.in

