Modified Anderson Darling Test for Wind Speed Data

Atul Viraj Wadagale[1], P.V. Thatkar[2], R.K. Dase[3] and D.V. Tandale[4]

[1] Assistant Professor (Statistics), MIMSR Medical College, Latur (MS), INDIA
[2] Lecturer, MIT College of Arts & Science, Alandi (D), Pune (MS), INDIA
[3] Assistant Professor (Statistics), MGM Medical College, Aurangabad (MS), INDIA
[4] Assistant Professor, Padmashree Dr. D.Y. Patil University, Navi Mumbai (MS), INDIA

Corresponding authors Email: atulvw@gmail.com, thatkar@indiatimes.com

Abstract: In this article we collect daily wind speed of three stations in Marathwada region for one year duration. By analyzing collected data it seems that Weibull distribution is best fitted for the wind speed data. The given data is compared by other continues distribution like Beta, Exponential, Gamma, Lognormal, Uniform. Maximum likelihood method is used for estimating parameters of weibull distribution. . Maximum likelihood method is used for estimating parameters of weibull distribution. We calculate Goodness of fit of these continuous distributions by comparing Anderson Darling test with modified Anderson Darling test.

Keywords: weibull distribution, beta distribution, Modified Anderson Darling normality test, maximum likelihood method, wind speed data.

Introduction:

India is fastest growing and developing country. Population of India is nearly equal to China. So, need of basic things demanded more like energy but the present energy sources are limited for used.[7] In Marathwada region (situated in middle of Maharashtra state with eight districts) people face 8-12 hours load shading in rural & urban area due to shortage of regular energy resources. This effect on their day-to-day life, agriculture and yield product and also on progress. As wind energy is renewable, environmental, cost effective energy so it may help for the people situated in this region. Wind speed is most important parameter of wind energy[2,5]. For evaluating the values of wind speed, we need an accurate probability distribution. As the values of wind speed are continuous one so we come to know that we must used continuous type of probability distribution such as, in literature, the Weibull distribution is used for wind energy. We compare Anderson-Darling normality test with Modified Anderson-Darling test. It can be shown that Weibull distribution is best fitted for wind speed data but, sometimes probability density function of wind speed data is not statistically accepted as weibull pdf so we used other continuous distribution like Beta, Exponential, Gamma, Lognormal and Uniform with Weibull pdf.

Material and Method:

We collected wind speed data from 3 different places from Marathwada region. We daily collect 1 sample from each station from January 2009 to Dec 2009. Therefore, there are 1035 total samples collected from all the stations. Table [1] shows the station, district and duration of records from 3 places.

Station	District	Latitude		Longitude		Annual Wind Speed	Duration
		Deg	Min	Deg	Min		
Kankara	A'bad	19	59	75	27	5.40	Jan'09-Dec'09
Sautada	Beed	18	48	75	20	5.72	
Rohina	Latur	18	27	76	57	5.57	

Table [1]: Study data from three stations in Marathwada region[1]

Statistical Distributions:

Here, we deal with different statistical continuous probability distributions with their probability density function.

Beta Distribution:

Probability density function for Beta distribution is,

$$f(x) = \frac{1}{B(\mu, v)} x^{\mu-1}(1-x)^{(v-1)}\ ;\ (\mu, v) > 0,\quad 0 < x < 1$$

Exponential Distribution:

Probability density function for Exponential distribution is,

$$f(x,\theta) = \theta e^{-\theta x}, \; x \geq 0$$

Gamma Distribution:

Probability density function for Gamma distribution is,

$$f(x) = \frac{e^{x} x^{\lambda-1}}{\Gamma(\lambda)} \; ; \lambda > 0, 0 < x < \infty$$

Uniform (Rectangle) Distribution:

Probability density function for Uniform distribution is,

$$f(x) = \frac{1}{b-a} \; ; if \; a < x < b$$

Lognormal Distribution:

The pdf of two parametric Lognormal distribution is given as follow:

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} exp\left(-\frac{1}{2\sigma^2}(log(x) - \mu\right)^2$$

Weibull Distribution:

The analysis of wind speed data can be computed by weibull function. The formula for the probability density function of the general weibull distribution is,

$$f(x) = \frac{\gamma}{\alpha}\left(\frac{x-\mu}{\alpha}\right)^{(\gamma-1)} exp\left(-\left(\frac{(x-\mu)}{\alpha}\right)^{\gamma}\right) \; wherex \geq 0, \gamma > 0$$

Annual mean wind speed at 10m Ht	Index value of wind resource
Below 4.5(m/s)	Poor
4.5-5.4(m/s)	Marginal
5.4 – 6.7(m/s)	Good to very good
Above 6.7(m/s)	Exceptional

Table 2 : Significance of wind energy according to speed[3]

Maximum likelihood method:

Maximum likelihood estimation begins with the mathematical expression known as a likelihood function of the sample data. The likelihood of a set of data is the probability of obtaining that particular set of data given the chosen probability model. This expression contains the unknown parameters. Those values of the parameter that maximize the sample likelihood are known as the maximum likelihood estimates.[4,6]

MLE for Weibull Distribution:

The scale and shape parameters were estimated using the method of maximum likelihood estimation (MLE) as follows:

$$\prod_{i=1}^{n} P(t_i|\gamma,\alpha)$$

Taking log of this gives,

$$ln\left(\prod_{i=1}^{n}\left(P(t_i|\alpha,\beta)\right)\right)$$

Hence we derive the following expression describing the likelihood function,

$$\sum_{i=1}^{n} ln\left(P(t_i|\alpha,\beta)\right)$$

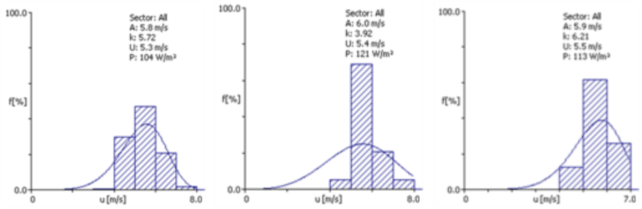
where γ is the shape parameter , μ is location parameter and α is the scale parameter. The case where μ= 0 & α=1 is called standard weibull distribution.
We put μ = 0, it is called the two-parameter weibull distribution.

The equation for the standard weibull distribution reduces to,

$$f(x) = \gamma x^{(\gamma-1)} exp\left(-(x^{\gamma})\right) \; where \; x \geq 0; \gamma > 0$$

Analysis showed that the result of wind speed could be represented by cumulative distribution function. The formula for cumulative distribution function of Weibull distribution is,

$$F(x) = 1 - e^{-(x^{\gamma})} \; where \; x \geq 0; \gamma > 0$$



Figure(1): Histogram for three stations A'bad, Beed, Latur respectively
Report produced by WAsP OWC Wizard (version 2.1.2)

Method	Parameter	
MLE	Shape(γ)	Scale(α)
	9.9658	5.8488

Table 3: Estimation values of shape and scale parameters of weibull obtained with MLE

Goodness-of-fit test:

1) **Anderson-Darling Normality test:**

The A-D test for normality can be computed as follow:

H_0:The data follow the specified distribution.
H_A: The data do not follow the specified distribution.
The hypothesis regarding the distributional form is rejected at the chosen significance level (*alpha*) if the test statistic, A^2, is greater than the critical value obtained from a table.

AD = -N-S

where,

$$S = \sum_{i=1}^{N} \frac{2i-1}{N} \left[\ln F(Y_i) + ln\left(1 - F(Y_{N+1-i})\right)\right]$$

therefore,

$$AD = -N - \frac{2i-1}{N}\left(ln\left(F(Y_i)\right) + ln\left(1 - F(Y_N + 1 - i)\right)\right)$$

The critical values for the Anderson-Darling test are dependent on the specific distribution that is being tested. Tabulated values and formulas have been published for a few specific distributions (normal, lognormal, exponential, Weibull, logistic, extreme value type 1)[8,9]

Distribution	Anderson-Darling Test				Modified Anderson-Darling Test			
	AD_c	AD_α	p value	Result	AD_c^*	AD_α^*	p value	Result
Beta	2.76	2.49	0.0364	Reject	3.06	2.49	0	Reject
Exponential	111	2.49	0	Reject	123	2.49	0	Reject
Gamma	6.54	2.49	0.0005	Reject	7.26	2.49	0	Reject
Lognormal	14.5	2.49	0	Reject	16.11	2.49	0	Reject
Uniform	49.4	2.49	0	Reject	54.88	2.49	0	Reject
Weibull	1.36	2.49	0.213	Accept	1.51	2.49	0.0072	Accept

Table 4 : Test of Goodness of fit

Modified Anderson-Darling Normality Test:

The Modified A-D test is nothing but the imposed form of A-D test. It can state as follow:

$$AD^* = AD\left(1 + \frac{0.75}{N} + \frac{2.25}{N^2}\right)$$

Station	Measured	Weibull fit	Discrepancy
A'bad	5.40	5.34	1.14%
Beed	5.72	5.42	5.24%
Latur	5.57	5.52	0.99%

Table 5: Site Description for three stations
Report produced by WAsP OWC Wizard (version 2.1.2)

Result:

The results shown in present study are statistically significant. The two parametric weibull distribution is best fitted for wind speed data. Annual average wind speed of three stations is more than 5 m/s while it is observed that the monthly mean wind speed data of three stations in Marathwada are fitted to the Weibull distribution. Weibull is best fitted distribution as considering to MLE. Weibull distribution is accepted in both the normality tests. It seems that Modified Anderson Darling test is better than Anderson Darling test because the outcome of modified A-D test is better than A-D test.

Acknowladgement:

Energy department, Govt. of Maharashtra, India.

References:

1. Energy department, Govt. of Maharashtra, India.
2. Stevens, M. and Smulders, P. [1979], The estimation of the parameters of the Weibull wind speed distribution for wind energy utilization purposes, *Wind Engng* 3, 132-145.
3. Ramachandra, T.V.; Subramanian, D.K. and Joshi, N.V. [1997], Wind Energy Potential Assessment in UttaraKanada, District of Karnataka, India, *Renewable Energy* 10(4), 585-611.
4. Jaynes, E. T. [1983] Papers on Probability, Statistics, and Statistical Physics. Netherlands: *D. Reidel, Dordrecht*.
5. Li M., Li X.[2005] Investigation of wind characteristics and assessment of wind energy potential for Waterloo region, *Energy conversion management*, 46, 3014-3033.
6. Gracia, A., Torres, J.L., Prieto, E., and DeFrancisco, A. [1998]. Fitting wind speed distribution: A case study, *Solar Energy* 2, 139-144.
7. Veysel YILMAZ, H. Eray CELIK [2008]. A Statistical approach to estimate the wind speed distribution: The case of Gelibolu region, *Doğuş Üniversitesi Dergisi* 9(1),122-132.
8. Anderson, T. W. and Darling, D. A. [1954]. A test of goodness of fit, *J. Amer. Statist. Assoc.*, 49, 765–769.
9. Chen, C. [2006]. Test of fit for the three-parameter lognormal distribution, *Computational Statistics & Data Analysis*, 50, 1418-1440.