**Chapter 1**

**Statistical Analysis of Wind Speed Data**

*In this chapter detailed statistical analysis for three selected sites namely Bagalkot, Vijayapura and Bengaluru is presented. Further statistical tests results are presented with historical wind speed data.*

Wind is stochastic in nature, speed and direction of wind at a location vary randomly with time. Knowledge of the statistical properties of the wind speed is essential for predicting the energy output of a wind energy conversion system. Thus, a proper statistical analysis of wind speed data is a very important step, not only for structural and environmental analysis, but also for the assessment of the wind energy potential and the performance of wind energy conversion system as well. The probability distribution of wind speeds is the key information needed to estimate wind energy output at a given site. In this study, statistical methods are used to analyze the wind speed data.

* 1. **Sites Selected for Research**

Three sites Bagalkot, Vijayapura and Bengaluru located in Karnataka State, India are selected for developing wind speed forecasting models. Bagalkot and Vijayapur are having low regime wind regimes. Bengaluru site having medium wind regime. To check and validate wind forecasting model two types of wind regime sites are selected. Fig. 2.1 presents Wind Power Density Map at 50 m level and also indicates location of study sites.

* + 1. **Site Representation**

**Bagalkot:** Site is located at Basaveshwar Engineering College (Autonomous), Bagalkot, Karnataka, India at Latitude of 16o 10’34.3"N and Longitude of 75o39’044.7"E . Data is measured using a 50m wind mast procured under R&D grant sanctioned by Technical Education Quality Improvement Program (TEQIP) of World Bank. Wind Data is recorded using NRG Symphonie Data logger at 10 min interval. Ten year historical wind speed data from 2007-17 is used for forecasting. Research facility at BEC Bagalkot is shown in Fig. 2.2. Wind speed is measured at 50m height. Cup type anemometer with direction sensor is used to measure wind speed. There were no obstacles for wind mast during the observation period.

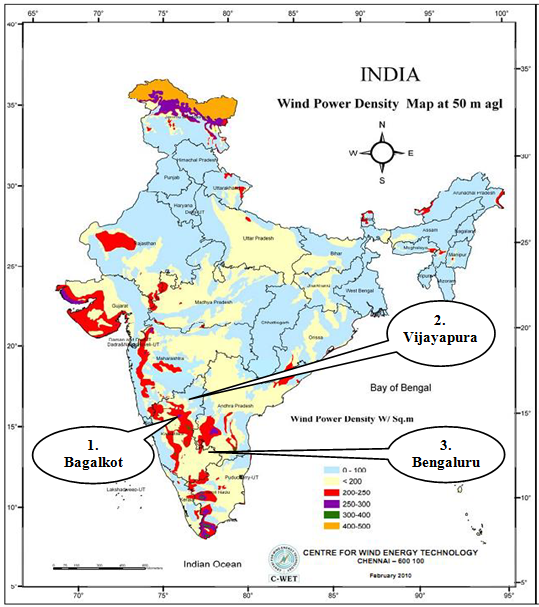


Fig.1.1:Indian Wind Energy Potential at 50 m Level (Source: CWET-Chennai)

**Vijaypura:** Site is located at Latitude of 16º49'48.61" N and Longitude of 75º42'36.11" E. Wind speed location is collected from automatic meteorological station operating in Regional Agricultural Research Station (RARS) at Vijayapur, Karnataka India. Time-series Wind Speed data measured at 10 min interval at 50 m height, for year 2009-2016 is used in the present study.

|  |  |  |
| --- | --- | --- |
|  |  | **http://www.rkwindmast.com/images/products/metsensors/windvanes/200p-vane.jpg** |
| **Anemometer** | **Wind Vane** |
| **http://www.rkwindmast.com/images/products/metsensors/other%20metsensors/110S-temp.jpg** |  |
| **Temperature Transudezer** | **NRG Symphonie Data Logger** |
|  | |
|  | |
| **50 meter wind mast** | **Weather Station** | |

**Fig.1.2: Research Facility at BEC(A), Bagalkot**

|  |  |
| --- | --- |
| http://kptclsldc.com/LDC.jpg | http://kptclsldc.com/MCC.jpg |
| **Fig. 1.3: KPTCL SLDC Bengaluru (Source: http://kptclsldc.com)** | |

**Bengaluru:** Site is located at Latitude 12.972358 N Longitude E 77.594619 E. Wind speed for the location is collected from SCADA centre at State Load Dispatch Center(SLDC) under Karnataka Power Transmission Corporation Limited at Race Course Road, Anand Rao Circle, Bengaluru. SLDC is an Apex body to ensure integrated operation of the power system in Karnataka and responsible for the Real time Load Dispatch functions, Operation and Maintenance of the SCADA System and Energy Accounting. Seven year historical wind data from 2009-16 is collected for study. Fig. 2.3 shows SCADA centre situated at Race course road, Anand Rao circle, Bengaluru.

**2.1.2 Statistical Analysis**

Goal of statistical analysis applied to forecasting is to identify pattern, seasonality, outliers, stationarity and statistics involved in historical data. There are various methods available for statistical analysis. Autocorrelation, partial autocorrelation Function, spectrum, frequency distribution and box plots are used in this study.

**Frequency Distribution of Wind Speed:** Two commonly used functions for fitting a wind speed probability distribution over a certain period of time are Weibull and Rayleigh distributions. Probability density function of the Weibull distribution is given by Jangamshetti S. H [64]:

|  |  |  |
| --- | --- | --- |
|  | k>0, c>1,v>0 | (1.1) |

Where,

k is dimensionless shape parameter

c is scale parameter in m/s

*v*i is wind velocity in m/s

Approximation for estimating Weibull shape parameter k is given as [64]:

|  |  |  |
| --- | --- | --- |
|  |  | (1.2) |

Where,

is mean wind velocity in m/s

is standard deviation given by:

Once k is known, then Scale factor, c is computed from:

|  |  |  |
| --- | --- | --- |
|  |  | (1.3) |

Where, is Gamma function

**Plot of the Data:** A good starting point for statistical analysis is a graphical plot of data to identify reference of trends, seasonality, missing data and outliers.

**Autocorrelation Coefficient:** Key statistic in statistical analysis is autocorrelation coefficient (or correlation of time series with itself, lagged by 1 or 2, or , more periods). Simple autocorrelation coefficient between Yt and Yt-1 is given by Box Jenkins [46]:

|  |  |  |
| --- | --- | --- |
|  |  | (1.4) |

Where, Ytis Present value ,Yt-1is Previous value and is Mean of sereis

Similarly, (1.4) extended for autocorrelations at 1,2,3,4,…..,k time lags is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (1.5) |

Auto-Correlation coefficient is a valuable tool for investigating properties of time series.

**Periodogram and Spectral Analysis:** Value of examining the set of amplitudes of various waves is fold to identify

* Randomness and seasonality in a time series
* Predominance of positive or negative autocorrelation

**Partial Autocorrelation Function (PACF):**For a time series, PACF is the partial correlation of Yt, Yt-h  is given by Box Jenkins [46]:

|  |  |  |
| --- | --- | --- |
|  |  | (1.6) |

PACF is used to measure the degree of association between Xt and Xt-k , when the effects of other time lags-1,2,3,…., up to k-1 are somehow partialled out. Their singular purpose in time series analysis is to help identify an appropriate order of model for forecasting.

**2.1.3 Statistical tests**

Objective of statistical analysis is to depict inferences about a population by tentative examination of a sample from that population. A statistical hypothesis is an assertion about distribution or more random variables. A test statistical hypothesis is a decision rule leading to either acceptance or rejection of hypothesis [64]. In this thesis paired t-test, F-test and regression test are conducted on model output forecasted wind speed data and error measured after prediction.

**Paired t-test**: If vartype is equal, test computes a sample standard deviation using:

|  |  |  |
| --- | --- | --- |
|  |  | (1.7) |

Where *Sx* and *Sy* are sample standard deviations of X and Y, respectively, and *N* and *M* are sample sizes of X and Y, respectively. It returns *P* value a probability, under null hypothesis,t-statistic is given by [64]:

|  |  |  |
| --- | --- | --- |
|  |  | (1.8) |

Where *Sx* and *Sy* are sample standard deviations,and are sample means, and *N* and *M* are sample sizes. Test statistic, has Student's *t* distribution with *N* + *M*– 2 degrees of freedom under null hypothesis.

**Regression Analysis** is a statistical technique commonly used to find a relationship between two variables. The regression can be curved or linear. In linear regression, a relationship of the form y=ax+b is constructed. Below the two linear regression methods used in this report are outlined.

Null hypothesis and an alternative hypothis is for each statistical test performed.

Two tailed paired sample hypothesis are [64]:

|  |  |  |
| --- | --- | --- |
|  |  | (1.9) |

Where and are mean of two variables population, denotes hypothesized value for which population mean is compared. Subscripts 1 and 2 denote two variable populations. Distribution ‘t’ statistic and test statistic for null hypothesis is given for paired sample by:

|  |  |  |
| --- | --- | --- |
|  |  | (1.10) |

Where is mean of difference of two variables, is standard error of mean is given by:

|  |  |  |
| --- | --- | --- |
|  |  | (1.11) |

Where SD is standard deviation of sample difference given by:

|  |  |  |
| --- | --- | --- |
|  |  | (1.12) |

Where n is number of pairs in paired variables, df is degree of freedom equal to n-1.

* 1. **Results of Statistical Analysis of sites**

Data plot, ACF, PACF, frequency distribution, frequency spectram and power spectral density are used to analyze wind speed data of three sites. Same period data for three sites from Sep-2008 to Aug-2009 is selected for statistical analysis and testing.

**Bagalkot:** Month-wise analysis from Sep-2008 to Aug-2009 are presented in Fig. 2.4(a) to (b) and Fig. 2.5. Data plot indicates that every month has different pattern and seasonality. Mean and variance are changing throughout data. Near zero wind speed values are at higher level as observed in frequency distribution plot and indicate higher uncertainty level. ACF indicates lags decaying exponentially with positive lags in Aug to Jan and negative lags in Feb to Jul. PACF indicates that very few lags namely intial two lags are significantly different from others. It depicts that present value depends only on last few data points. Frequency distribution plots indicate some wind speed values are close to zero wind speed and range from 2m/s to 10m/s. Spectrum and frequency indicates dominance of low frequency **indicating existence of non-stationarity**. Results reveal that clearly there is a need for processing for processing data.

**Vijayapura:** Wind speed data analysis plots for Vijayapur site are shown in Fig. 2.6. Data plot and frequency distribution plot indicate that, near zero wind speed values are very less and uniform distribution throughout wind speed range is observed. ACF plot indicate exponentially decaying positive lags in Oct to Nov, exponentially decaying positive lags with 24hour seasonality in Sep to Oct, Jan, Aug months and exponentially decaying alternative positive and negative lags in Feb to Jul months. PACF plots indicate two initial lags are significantly higher than remaining in all months. Frequency spectrum plots indicate lower frequency dominance in data. Wind speed values range from 2 m/s to 12 m/s.

**Bengaluru:** Wind speed data analysis plots for Bengaluru site are shown in Fig. 2.7. Data plot and frequency distribution plot indicate that, near zero values are lesser. Uniform distribution of wind speed values in Sep to Dec and Mar to Aug. Non uniform distribution observed in Jan and Mar. Wind speed values range from 2 m/s to 8 m/s. ACF plots indicate exponentially decaying positive lags in Oct-Jun months and exponentially decaying positive cum negative lags in Jul to Aug months. PACF plots indicate two initial lags significantly higher than remaining lags. Frequency spectrum plot show low frequency dominance in wind speed.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | | |  | |  | |
|  | | |  | |  | |
| **Fig. 1.4(a): Wind Speed at 50 m level (data Plot, histogram, ACF, PACF, Spectrum, Frequency) from Sep-2008 to Feb-2009 of Bagalkot BEC Data** | | | | | | |
|  | |  | |  | | |
|  | |  | |  | | |
|  | |  | |  | | |
| **Fig. 1.4(b): Wind Speed at 50 m level (data Plot, histogram, ACF, PACF, Spectrum, Frequency) from Mar-2009 to Aug-2009 of Bagalkot BEC Data** | | | | | | |
|  |  | | | | |  |
|  |  | | | | |  |
|  |  | | | | |  |
|  |  | | | | |  |
| **Fig. 1.5: Wind Speed Frequency Distribution of Bagalkot Site for Yr.2007 to 2017** | | | | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | |  | |  | |  | |  | |  | |
|  | |  | |  | |  | |  | |  | |
|  | |  | |  | |  | |  | |  | |
|  | |  | |  | |  | |  | |  | |
|  | |  | |  | |  | |  | |  | |
|  | |  | |  | |  | |  | |  | |
| **Fig. 1.6: Wind Speed at 50 m level (data Plot, histogram, ACF, PACF, Spectrum, Frequency) from Sep-2008 to Aug-2009 of Vijayapura Wind Mast** | | | | | | | | | | | |
|  | |  | |  | |  | |  | |
|  | |  | |  | |  | |  | |
|  | |  | |  | |  | |  | |
|  | |  | |  | |  | |  | |
| **Fig. 1.7: Wind Speed at 50 m level (data Plot, histogram, ACF, PACF, Spectrum, Frequency) from Sep-2008 to Aug-2009 of Bengaluru Wind Mast** | | | | | | | | | |

* 1. **Comparison of Three Sites**

Fig. 2.8 presents compares monthly average wind speed of three site data for year 2017.X-axis represents site name with month and Y-axis presents wind speed in m/s. Annual averaged cubic mean wind speed observed for Bagalkot, Vijayapura and Bengaluru sites are 3.81m/s, 6.14m/s and 6.18m/s respectively. Thick dotted line with circles represents annual average wind speed curve. This curve indicates that average wind speed curve for all three sites have different patterns. Thin lines present diurnal hourly average wind speed. It is observed that, peak average wind speed occurs in Jun-Aug at Bagalkot, Nov-Dec at Vijayapura and Mar-May at Bengaluru sites respectively.

**Fig. 1.8: Monthly Average Wind Speed variation for 3 sites for the year 2017**

Wind rose plots for three sites given in Fig. 2.9 indicate higher wind velocity arriving from north direction for Bagalkot and Vijayapura. Wind velocity for Bengaluru site arriving from all directions. Weibull parameters and power densities using Windographer Software are given in Table 2.1 and Fig. 2.10. It is found that, Weibull parameters calculated with three methods are nearly same. Power density estimated for Bagalkot site is 72.7kW/m2, 222.8kW/m2 for Vijayapur and 251.8kW/m2. Table 2.2 presents comparison of statistics of hourly wind speed data for year 2017 of three study sites. It is seen that confidence interval, Kurtosis and Skewness for Bagalkot site are 0.043, 1.449 and 1.099 respectively which indicates higher level of uncertainty.

|  |  |
| --- | --- |
| **10 min data** | **Hourly data** |
|  |  |
| **(a). Bagalkot** | |
|  | |
|  | |
|  | |
|  |  |
| **(b). Vijayapura** | |
|  | |
|  | |
|  | |
|  | |
|  |  |
| **(c). Bengaluru** | |
| **Fig. 1.9: Wind Rose Plot for 10 min and Hourly Wind Speed Data** | |
|  |  |
| **(a). Bagalkot** | **(b). Vijayapur** |
|  | |
| **(c). Bengaluru** | |
| **Fig. 1.10: Frequency distribution of study sites using Windographer Software** | |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1.1: Comparison Weibull Parameter and Power density of sites** | | | | | | | |
| **Site** | **Method** | **Weibull** | | **Mean**  **m/s** | **Properties**  **above** | **Power**  **Density**  **kW/m2** | **R2** |
| **k** | **C**  **m/s** |
| Bagalkot | Max. Likelihood | 1.845 | 4.292 | 3.812 | 0.447 | 70.6 | 0.8607 |
| Least Squares | 1.948 | 4.302 | 3.815 | 0.453 | 66.7 | 0.8716 |
| WAsP | 1.629 | 4.088 | 3.659 | 0.409 | 72.7 | 0.8121 |
| Vijayapura | Max. Likelihood | 2.664 | 6.899 | 6.132 | 0.480 | 213.2 | 0.9925 |
| Least Squares | 2.566 | 6.950 | 6.171 | 0.483 | 222.8 | 0.9865 |
| WAsP | 2.739 | 6.934 | 6.169 | 0.488 | 213.2 | 0.9941 |
| Bengaluru | Max. Likelihood | 2.221 | 6.986 | 6.188 | 0.467 | 251.4 | 0.9811 |
| Least Squares | 2.298 | 6.949 | 6.156 | 0.466 | 240.6 | 0.9742 |
| WAsP | 2.161 | 6.935 | 6.141 | 0.459 | 251.8 | 0.9853 |

**Table 1.2: Descriptive statistics of hourly Wind Speed**

|  |  |  |  |
| --- | --- | --- | --- |
| **Parameter** | **Bagalkot** | **Vijayapura** | **Bengaluru** |
| Mean | **3.817** | **6.142** | **6.179** |
| Standard Error | 0.022 | 0.025 | 0.030 |
| Median | 3.450 | 6.050 | 5.800 |
| Mode | 3.683 | 6.583 | 3.700 |
| Standard Deviation | 2.061 | 2.371 | 2.886 |
| Sample Variance | 4.248 | 5.625 | 8.333 |
| Kurtosis | 1.449 | 0.356 | 0.262 |
| Skewness | **1.099** | **0.308** | **0.536** |
| Minimum | 0.400 | 0.400 | 0.433 |
| Maximum | 13.833 | 15.566 | 16.950 |
| Confidence Level  (95.0%) | **0.043** | **0.049** | **0.060** |

* 1. **Results of Statistical tests on Wind Speed Data**

In order to carryout hypothesis test, following tests are conducted on original wind speed data of three sites. Table 2.4 gives summary of p-test, Anova single factor, Kurtosis, skewness, F-test and t-tests. In this table t-stat for each sample pair, and t-critical for 95% confidence intervals for two tailed and one-tailed test is given. Pearson correlation and covariance are given for each site.

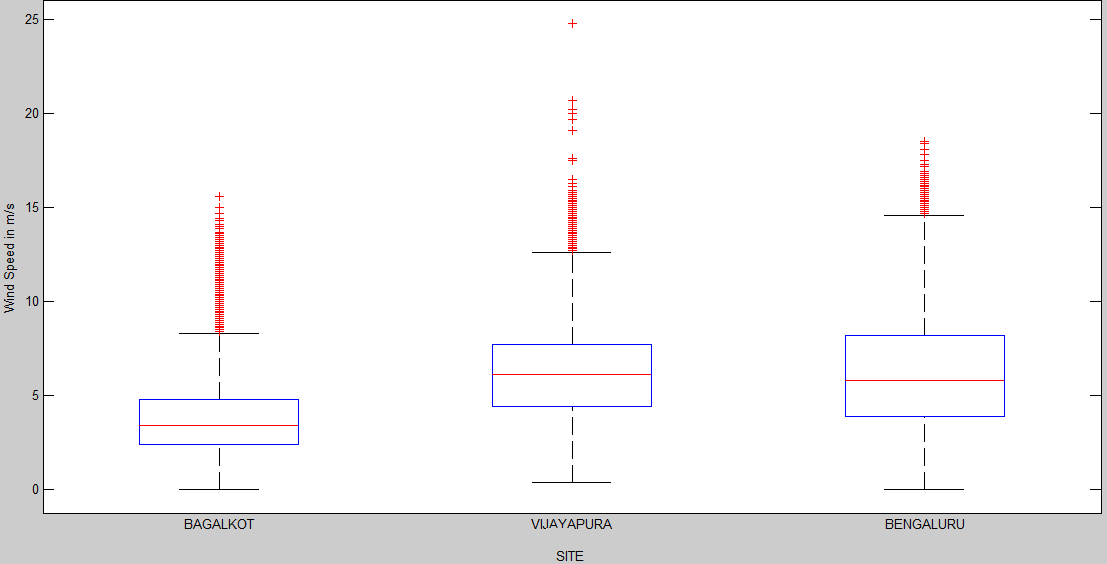
**Table 1.3: Test results of original wind speed data**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Test** | | **Bagalkot** | **Vijayapura** | **Bengaluru** |
| **Anova Single Factor** | F | 712081.8 | 1030514.0 | 908509.9 |
| MSE | 17.831 | 11.171 | 11.555 |
| **Pearson Correlation** | | -0.029 | -0.231 | -0.291 |
| **Covariance** | | -0.357 | -2.292 | -3.262 |
| **Kurtosis** | Velocity | 1.540 | 0.157 | 0.207 |
| Temperature | 6.652 | -0.101 | 0.117 |
| **Skewness** | Velocity | 1.080 | 0.299 | 0.514 |
| Temperature | -1.403 | 0.627 | 0.718 |
| **Confidence Level** | Velocity | 0.018 | 0.021 | 0.025 |
| Temperature | 0.047 | 0.034 | 0.032 |
| **F-test** | F | 0.147 | 0.372 | 0.607 |
| F-cri | 0.985 | 0.985 | 0.985 |
| **t-test Paired Two Sample for Means** | Pearson  Correlation | -0.029 | 0.230 | -0.291 |
| t-stat | -835.521 | -924.668 | -841.722 |
| t-cri 1 tail | 1.644 | 1.6443 | 1.644 |
| t cri 2 tail | 1.960 | 1.960 | 1.960 |
| **Z-test** | Z | -843.850 | -803.468 | -767.300 |

Test results reveal that, negative Pearson correlation between wind velocity and temperatures identified for Bagalkot and Bengaluru sites. Lesser MSE found for Vijayapura and Bengaluru sites in ANOVAs Single Factor test.

**Outliers Detection in Wind Speed**

Existence of outliers in wind speed data effect performance of models. Hence detection of outliers is necessary to improve efficiency of forecasting models. Box plots are used here to detect outliers for three selected sites as shown in Fig. 2.10. There are very less outliers present in Bagalkot and Bengaluru data. But for Vijayapura wind speed data many outliers exist between wind speeds 15m/s to 25m/s. Hence these outliers need to be considered while estimating parameters of models.

****

**Fig. 1.11: Outlier detection using Box plots for three sites**

* 1. **Salient Outcome of Wind Data Analysis**

From perusal of results and graphical representation of results, following salient observations are deducted:

* Detailed analysis from statistics and graphs presented shows different patterns in wind speed time series.
* ACF and PACF plots describe seasanility, non-staitionarity, outlier’s existence, and autocorrelation among present and previous data. Raw wind data is to be made stationary and outliers are to be filtered to improve forecast accuracy.
* Analysis of ACF, PACF, Spectral density, Cumulative frequency distribution, and box plots, it is found that Bagalkot and Vijayapura show lesser annual average wind speed than Bengaluru site.
* Bengaluru wind speed data has highest annual mean wind speed of 6.986m/s.
* Statistical tests carried out for wind speed data for three sites.
* Comparing results of three sites it is found that, uncertainty level is higher.

Salient observations presented above after a detailed investigation by using ACF, PACF, Spectrum plots that, each season have different patterns and seasonal models must be used in order to obtain accurate forecasting of wind speed. In addition, non-stationarity is to be converted to stationary series by normalizing data in order to avoid mis specification in estimating model parameters.