



A Statistical Analysis and Datamining Approach for Wind Speed Prediction

Mahesh K^{1*}, M V Vijayakumar ^{2.}, Gangadharaiyah.Y.H³

^{1*}Researcher Scholar JAIN University CSE dept and Faculty at
New Horizon College of Engineering Bangalore-560103, India.

E-mail: maheshkgowda@gmail.com

² Department of Computer Science Engineering, Dr AIT, India.

E-mail: dr.vijay.research@gmail.com

³ Department of Mathematics, New Horizon College of Engineering, India.

E-mail:gangu.honnappa@gmail.com

ABSTRACT

The wind power industry has seen an unprecedented growth in last few years. The surge in orders for wind turbines has resulted in a producer's market. This market imbalance, the relative immaturity of the wind industry, and rapid developments in data processing technology have created an opportunity to improve the performance of wind farms and change misconceptions surrounding their operations. This research offers a new paradigm for the wind power industry, data-driven modeling. Each wind Mast generates extensive data for many parameters, registered as frequently as every minute. As the predictive performance approach is novel to wind industry, it is essential to establish a viable research road map. This paper proposes a data-mining-based methodology for long term wind forecasting (ANN), which is suitable to deal with large real databases. The paper includes a case study based on a real database of five years of wind speed data for a site and discusses results of wind power density was determined by using the Weibull and Rayleigh probability density functions. Wind speed predicted using wind speed data with Datamining methodology using intelligent technology as Artificial Neural Networks (ANN) and a PROLOG program designed to calculate the monthly mean wind speed.

Indexing terms/Keywords:

Wind Speed prediction; Datamining; ANN; Weibull; Rayleigh; PROLOG.

Academic Discipline And Sub-Disciplines

Computer science engineering And Mathematics

SUBJECT CLASSIFICATION

Renewable Energy and Data Mining

TYPE (METHOD/APPROACH)

Statistical Methods, Artificial neural network methods, PROLOG.

Council for Innovative Research

Peer Review Research Publishing System

Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY

Vol. 14, No. 2

www.ijctonline.com , editorijctonline@gmail.com



1. INTRODUCTION

The increased use of energy and the depletion of the fossil fuel reserves combined with the increase of the environmental pollution have encouraged the search for clean and pollution-free sources of energy. One of these is wind energy. This is a clean, inexhaustible and a “free” source of energy that has served the mankind for many centuries by propelling ships, driving wind turbines to grind grains and for pumping water. Despite the high cost of wind power this may become a major source of energy in the years to come. This is so because the severe pollution of the planet originating from the burning of the fossil fuels and the nuclear energy risks cannot continue forever. The predicted variations of meteorological parameters such as wind speed, relative humidity, solar radiation, air temperature, etc. are needed in the renewable industry for design, performance analysis, and running cost estimation of these systems [1, 8]. For proper and efficient utilization of wind power, it is important to know the statistical characteristics, persistence, availability, diurnal variation, and forecasting of wind speed. The wind characteristics are needed for site selection, performance forecasting and planning of wind turbines. Of these characteristics, the forecasting of mean monthly and daily wind speed is very important. During recent years according to global environmental pollutions, trends towards the sustainable energy and green power sources such as wind energy was largely increased.

Wind energy is one of the economic renewable sources and a valuable supplement to conventional energy sources. One of the important problems in wide uses of wind power is difficulties of accurate wind power forecast [4, 5, 6]. Fluctuations in wind power production, also makes it difficult for owners of wind power plants to compete in electricity markets. An accurate forecast allows grid operators to schedule economically efficient generation to meet the demand of electrical customers. One of the most important facts of wind energy production is related to the rapid changes in wind speed and direction, so the output power of wind turbines vary with these changes. The theoretical calculation shows that the output power of wind turbines varies with the cubic power of wind speed, wind direction, and air density. Therefore, the forecasting of energy generation must be, based on wind speed and direction forecasting. Another Challenge directly related to the nature of wind power is maintaining ancillary services. Since wind Power is intermittent, variable, and largely uncontrollable, it requires additional reserve be allocated to hedge against uncertainty in the availability of wind power. Due to its variability, intermittency, and controllability characteristics, wind power integration presents unique challenges. Generally, they can be grouped in the following: challenges in wind speed forecasting and forecasting wind power. Often wind and wind power rely on wind data (wind speed as well as wind direction) recorded over sufficiently large period of time. Such data varies to a large degree over time. In order to make an accurate forecast, one needs to find meaningful patterns in data collection. So, Data mining provides us with the tools for discovering patterns in data automatically. Data mining and analysis techniques approaches can be used to resolve these challenges. Whereas the other wind resource models such as the simple persistent model and numerical weather forecasting at various wind regimes could not give accuracy [2, 3, 9]. Data mining can be defined as the extraction of knowledge form data, the exploration and analysis of large amounts of data in order to discover meaningful patterns. Although statistical analysis is useful in numerical data analysis, it does not solve data mining problems, such as discovering meaningful patterns in large quantities of data which are very much essential with the wind power systems. Data mining uses intelligent technologies and often applies such tools as Artificial Neural Networks (ANN), Fuzzy systems, Neuro- fuzzy systems, Artificial intelligence and other soft computing techniques.

In this paper, we have considered wind speed data of the given site, of a database with 5 years of historical data [5], [6]. a model is developed to estimate monthly mean wind speed for a given period using PROLOG [3] (AI Tool) Weibull probability distribution of wind speed and forecast the wind speed using artificial neural network. The Weibull distribution is commonly used to describe the probability distribution of wind speed at a given location [10, 11, 23]. Two parameters, the scale factor (c, also sometimes referred to as A) and the shape factor (k) are sufficient to describe a curve which approximates the probability distribution of the wind speed. The Rayleigh distribution (defined as a Weibull distribution with the shape factor k = 2), is considered frequently as a reference frequency distribution. Using an artificial neural network (ANN) [7] in order to forecast the wind speed taking into account different input attributes, namely the number of previous hours used to predict the wind speed at a given site.

2. WEIBULL AND RAYLEIGH WIND SPEED STATISTICS

In order to describe the wind speed frequency distribution, there are several probability density functions. The probability density functions point out the frequency distribution of wind speed, and which is the interspace of the most frequent wind speed, and how long a wind turbine is out and on of action. The Weibull and the Rayleigh functions are the two most known functions [14– 16]. The Weibull distribution is a special case of generalized gamma distribution, while the Rayleigh distribution is a subset of the Weibull. The Weibull is a two parameter distribution while the Rayleigh has only one parameter and this makes the Weibull somewhat more versatile and the Rayleigh somewhat simpler to use. The Weibull distribution function is expressed as follows where v is the wind speed, c the Weibull scale parameter in m/s, and k the dimensionless Weibull shape parameter. These parameters can be determined by the mean wind speed-standard deviation method [15 –19] using equations 1 to 4.

$$P_v = \frac{k}{v} \left(\frac{v}{c} \right)^{k-1} \text{Exp} \left\{ - \left(\frac{v}{c} \right)^k \right\} \text{----- (1)}$$



$$P_v = \text{Exp} \left\{ - \left(\frac{v}{c} \right)^k \right\} \text{-----(2)}$$

$$k = \left(\frac{\sigma}{\bar{v}} \right)^{-1.086} \text{-----(3)}$$

$$c = \frac{\bar{v}}{\Gamma \left(1 + \frac{1}{|k|} \right)} \text{-----(4)}$$

$$\bar{v} = \frac{1}{n} \left(\sum_{i=1}^n v_i \right) \text{-----(5)}$$

$$\sigma = \left[\frac{1}{n-1} \sum_{i=1}^n v_i - \bar{v}^2 \right]^{0.5} \text{-----(6)}$$

Where \bar{v} is the mean wind speed, is calculated using equation 5 and σ is the standard deviation, is calculated using equation 6 and n is the number of hours in the period of the time considered such as month, season or year. The dimensionless shape parameter, k, of Weibull distribution is assumed as 2 in Rayleigh distribution functions [19]. The probability density function of the Rayleigh distribution is expressed as

$$f_R v = \frac{\pi v}{2v^2} \exp \left[- \left(\frac{\pi}{4} \right) \left(\frac{v}{\bar{v}} \right)^2 \right] \text{-----(7)}$$

The wind power density of any windy site per unit area based on any probability density function to estimate the wind power can be expressed as

$$P_m = \frac{1}{2} \rho \int_0^\infty v^3 f_v dv \text{-----(8)}$$

Where ρ is the standard air density, 1.225 kg/m [15, 17, 21]. Assuming that there is no variation in air density [20, 22]. When the Weibull function is chosen as distribution function $f(v)$, the average wind power density is calculated as

$$P_{mw} = \frac{1}{2} \rho v^{-3} \frac{\Gamma(1+3/k)}{[\Gamma(1+1/k)]^3} \text{-----(9)}$$

3. PROLOG

PROLOG is a simple, yet powerful programming language, based on the principles of first order predict logic. The name of the language is an acronym for the French 'PROgramma-tion en LOGique'. About 1970, PROLOG was designed by A. Colmerauer and P. Roussel at the University of Marseille, influenced by the ideas of R.A. Kowalski concerning programming in the Horn clause subset of first order predict logic. The name of PROLOG has since then been connected with a new programming style, known as logic programming [23]. Figure 1 show the basic prolog flow chart used to estimate the mean annual or monthly wind speed of wind data of the given site.

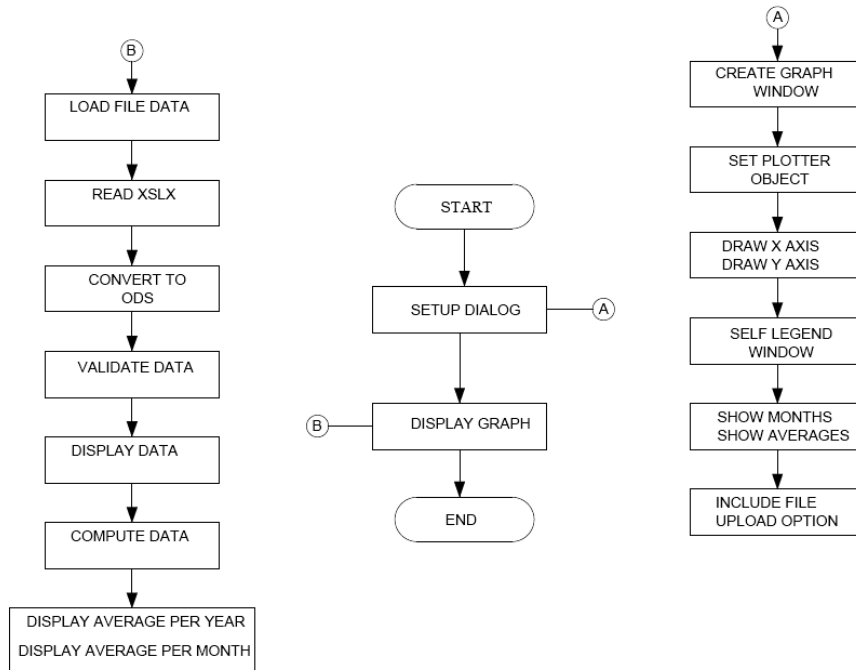


Figure 1: Flow Chart

Until the end of the seventies, the use of PROLOG was limited to the academic world. Only after the development of ancient PROLOG interpreter and compiler by D.H.D. Warren and F.C.N. Pereira at the University of Edinburgh, the language entered the world outside the research institutes. The interest in the language has increased steadily. However, PROLOG is still mainly used by researchers, even though it allows for the development of serious and extensive programs in a fraction of the time needed to develop a C or Java program with similar functionality. The only explanation is that people like wasting their precious time. Nevertheless, there are a large number of fields in which PROLOG has been applied successfully. The main applications of the language can be found in the area of Artificial Intelligence; but PROLOG is being used in other areas in which symbol manipulation is of prime importance as well. Some application areas are: Natural-language processing, Compiler construction, the development of experts systems, Work in the area of computer algebra, the development of (parallel) computer architectures and Database systems [12].

4. ARTIFICIAL NEURAL NETWORK (ANN)

ANN is a massively parallel computational model that simulates the structure and the functional aspects of biological nervous system. The brain which is central to the nervous system is made up of discrete units called neurons (the Greek word for nerves). Neurons are polarized cells that receive signals via highly branched extensions, called dendrites and send information along un-branched extensions, called axons. In total, the human brain contains approximately 1014 to 1015 interconnections of neurons. All neurons process information in much the same way and information within neurons are transmitted in the form of electrical impulses called action potentials via the axons from other neuron cells. When the action potential arrives at the axon terminal, the neuron releases chemical neurotransmitter which effects the interneuron communication at specialized connections called synapses [2, 3, 9].

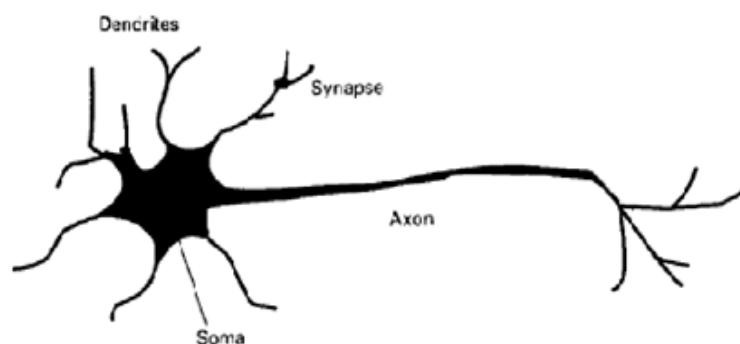


Figure 2. Anatomy of a biological neuron

In biological systems, learning involves adjustments to the synaptic connections between these neurons. Figure 2 shows the anatomy of a biological neuron. Artificial neuron was inspired from the structure and functions of the biological

neuron and figure 3 shows the structure and components of an artificial neuron. Artificial neuron functionality and properties such as its flexibility and ability to approximate functions to be learned depend on its activation function. Some important activation functions are linear, sigmoid, threshold and hyperbolic tangent activations. A neuron learns through an iterative process of adjustment of its synaptic weights and a neuron become more knowledgeable after each iteration of the learning process.

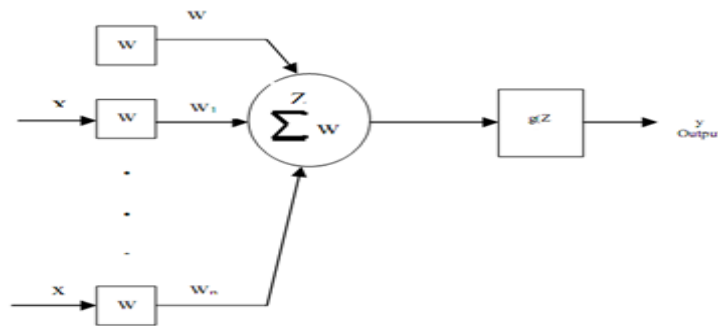


Figure 3: The structure of an Artificial Neuron

The functionality of a single artificial neuron which is the basic unit of an artificial neural system is very limited. To learn how to solve a problem that cannot be learned by a single neuron, an interconnection of multiple neurons called Neural Network (NN) or Artificial Neural Network (ANN) must be employed. Figure 4 shows the simplest Artificial Neural Network (ANN).

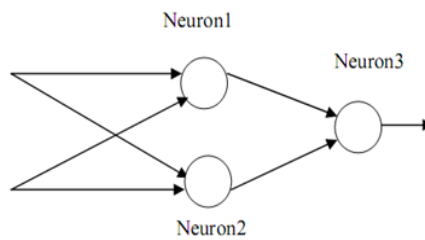


Figure 4: The simplest Artificial Neural Network

MATLAB R2008a Neural Network Toolbox is utilized in this work because it contains various functions and backpropagation training algorithms for implementing feed forward neural networks. There are various backpropagation training algorithms with diverse capability based on the nature of problem the network is designed to solve. Generally, ANN can be trained for tasks such as function approximation (regression) or pattern recognition (discriminant analysis). Our problem falls under approximation (regression) in which the network is trained with the available wind data and predict wind speed data of the site. This is aimed at obtaining the optimal algorithms for the site out of the available backpropagation training algorithms in MATLAB R2008a. It is actually very difficult to know which training algorithm will be the fastest for a given problem. It depends on many factors, including the complexity of the problem; the number of data points in the training set, the number of weights and biases in the network, the error goal and the task the network is to perform (Howard et.al, 2010). The Figure 5 shows, the network architecture for our ANN based model as created by the MATLAB R2008a wizard.

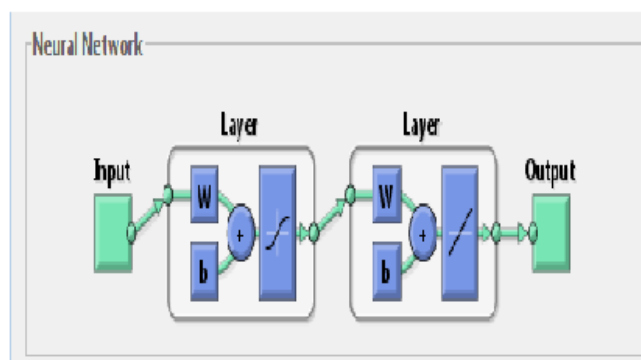


Figure 5: ANN Model



5. CASE STUDY

Step 1: Site Description

A 50 m lattice met tower was installed near the proposed site in districts of Karnataka State in Southern India as in table 1. The met tower was equipped with wind speed and direction sensors, a data logger, an ambient air temperature sensor, and a lightning arrester for protecting the measuring equipment from a severe thunder storm. Wind speeds and directions at 10 m, 30 m, and 50 m as well as ambient temperature at 3 m above ground level (AGL) were measured between January 2006 to April 2010 with a 1 min sampling interval and a 10 min recording interval. Wind speeds were measured using 3-cup anemometers and wind directions were measured using standard wind vanes both of which were attached to the end of booms mounted on the met tower. The instruments were connected to a NRG data logger which was powered by a PV-battery system.

Table 1: Site Details

Sl.No.	Location	Latitude	Longitude	Elevation	Districts	Date of Commission
1	Ingaleshwar	N 16° 38' 45"	E 76° 7' 35"	677m	Bijapur	10.01.2006

The resolution of wind speed measurements were 0.19 m/s and the resolution of wind direction measurements were 1.4°. Measurement accuracy as specified by the instrument manufacturer was ±3% for wind speeds of 17-30 m/s. The measured wind speeds at 10 m, 30 m, and 50 m AGL were analyzed using the Weibull distribution technique. Weibull parameters were investigated based on the analytical method. A sample data obtained is has shown in the Table 2.

Step 2: Prolog plots

The data obtained from the anemometer was a raw set of wind speed data. We input this data to the prolog code, which calculated mean values and gave us the prolog plots as shown in Figures (6a-6e).

Table 2: Sample data of the site recorded by NRG data logger

Site Information:			Sensor on channel 1:			January 2006																				
Project:	wind energy survey		NRG #40 Anem. m/s			Hourly Averages Table Ch 1																				
Location:	Ingaleshwar,Basavanabagewadi Tq. (56 D/2)		Height:	50 m	Units:	SITE 5006																				
Elevation:	660m		Serial #:	SN:		Lat:N 16°38'45" Long:E 76°7'35"																				
	Hour																									
Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	AVG	
1	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
2	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
3	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
4	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
5	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
6	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
7	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
8	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
9	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
10	5.3	4.4	5.1	5.3	5.4	4.7	5.0	4.4	3.9	3.3	5.1	6.2	6.8	5.6	4.3	4.9	5.4	5.7	6.1	5.6	4.8	6.2	5.8	5.1	5.2	
11	4.9	5.0	4.4	4.0	3.9	4.7	5.6	5.3	4.8	4.7	5.9	7.1	7.6	7.2	6.5	7.5	8.2	8.0	7.7	6.2	4.6	4.2	3.7	4.5	5.7	
12	4.5	3.7	4.3	4.0	4.4	5.1	6.3	6.7	5.6	6.1	7.5	7.2	7.4	8.4	8.3	7.1	7.2	8.2	8.4	8.6	6.1	4.4	3.9	4.2	6.1	
13	3.3	3.8	5.2	4.3	4.1	5.6	6.8	6.2	6.2	7.0	8.3	7.2	6.6	6.8	6.7	6.6	8.0	8.3	7.3	6.9	5.6	5.4	6.0	4.9	6.1	
14	5.1	4.9	4.6	4.4	5.0	5.4	5.8	5.7	6.0	7.1	5.0	6.1	7.1	8.1	7.7	6.3	6.4	6.0	5.4	4.5	5.8	6.3	6.6	6.4	5.9	
15	6.0	5.8	5.2	5.2	5.6	5.9	7.2	6.8	7.3	8.4	8.4	8.2	7.9	8.6	8.6	9.7	8.6	7.8	7.8	7.8	7.0	6.5	6.1	5.5	7.2	
16	5.3	3.9	2.7	3.4	3.6	4.8	6.4	6.7	6.3	5.4	6.5	7.0	7.4	8.4	8.8	8.5	8.5	9.6	9.3	9.1	7.4	6.3	5.0	3.7	6.4	
17	2.5	3.1	3.2	3.1	2.9	4.5	5.9	5.3	4.5	4.1	2.7	1.7	2.4	3.1	1.4	1.5	3.5	1.4	1.3	3.6	4.3	3.9	3.1	3.4	3.2	
18	4.0	3.6	3.8	4.1	4.7	4.8	5.5	5.2	4.9	3.1	2.2	3.7	3.8	4.7	4.2	5.1	5.9	5.2	4.7	4.1	3.0	2.8	4.5	4.3	4.3	
19	5.7	4.9	5.5	5.0	5.2	4.5	3.8	5.2	4.7	3.9	5.1	5.7	6.0	5.4	5.7	5.7	5.5	4.9	5.4	5.4	3.2	3.8	4.9	5.2	5.0	
20	5.0	5.2	4.7	4.1	3.5	2.9	2.3	3.8	4.0	3.7	3.5	4.8	4.5	4.2	5.1	7.3	7.1	6.2	4.7	4.7	3.8	5.3	5.9	3.8	4.6	
21	3.9	3.5	3.0	3.2	3.9	5.3	6.3	7.3	7.0	5.6	5.3	5.6	6.3	6.4	5.5	5.9	4.3	1.6	2.0	2.0	1.6	2.3	4.4	4.2	4.4	
22	3.4	4.2	3.9	4.4	3.7	4.1	6.3	7.4	7.7	7.6	7.6	6.8	6.3	5.3	4.4	6.6	7.7	6.6	7.0	6.6	4.8	6.0	5.7	6.0	5.8	
23	5.6	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
24	6.1	5.9	7.4	6.7	8.2	8.3	8.6	8.5	6.5	7.7	7.7	7.1	6.3	4.6	3.0	2.8	2.9	4.4	5.6	5.9	8.5	8.0	7.8	7.9	6.5	
25	7.8	8.0	8.2	7.7	7.2	7.1	6.8	7.5	7.8	8.2	6.6	5.6	4.3	3.8	3.7	3.4	4.0	5.0	6.2	5.7	5.7	6.6	6.8	6.8	6.3	
26	6.7	7.3	7.4	7.2	7.6	7.3	7.5	7.7	6.4	5.7	5.5	4.6	3.7	3.8	3.6	3.3	3.4	4.0	4.0	5.4	5.7	4.6	6.1	8.1	5.7	
27	8.1	7.9	7.6	7.3	7.7	6.9	7.1	6.7	5.0	5.4	6.8	6.0	5.2	4.7	5.3	4.9	4.5	4.0	4.5	5.0	4.7	4.0	3.9	6.1	5.8	
28	6.2	7.2	6.6	6.6	6.9	6.2	4.9	5.2	3.8	5.2	4.3	4.5	4.5	4.8	4.9	4.7	4.5	4.6	4.2	3.8	3.8	3.9	4.4	6.1	5.1	
29	8.5	6.9	7.8	6.6	5.1	4.2	4.9	7.6	5.9	5.5	4.5	4.4	3.4	3.6	4.0	3.8	3.5	4.2	5.8	6.2	5.9	4.7	3.8	4.1	5.2	
30	4.2	5.3	4.9	7.2	8.3	8.6	8.9	8.1	7.3	5.8	5.0	4.9	5.2	5.7	5.0	4.5	4.9	5.2	6.6	6.3	6.5	6.7	6.6	6.8	6.2	
31	7.4	6.8	8.0	9.6	10.3	9.2	8.2	7.6	6.9	6.5	6.0	6.5	6.7	6.7	6.5	6.6	6.8	7.0	8.0	8.2	8.0	9.0	8.2	8.9	7.6	
AVG	5.4	5.3	5.4	5.4	5.6	5.7	6.2	6.4	5.8	5.7	5.7	5.8	5.6	5.7	5.5	5.5	5.7	5.7	5.8	5.8	5.4	5.3	5.3	5.5	5.6	

Generated Wednesday, January 02, 2008 Total 10-minute intervals: 4464 Intervals used in calculations: 3094 Percent data used: 69.3 NRG Systems SDR Version 5.04

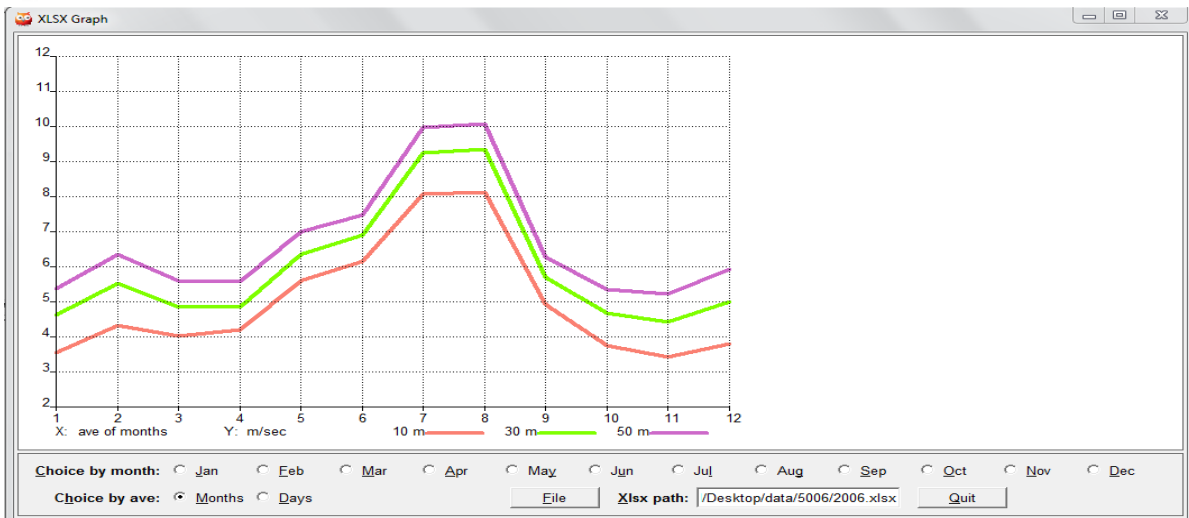


Figure 6a: Annual mean wind speed for the year 2006

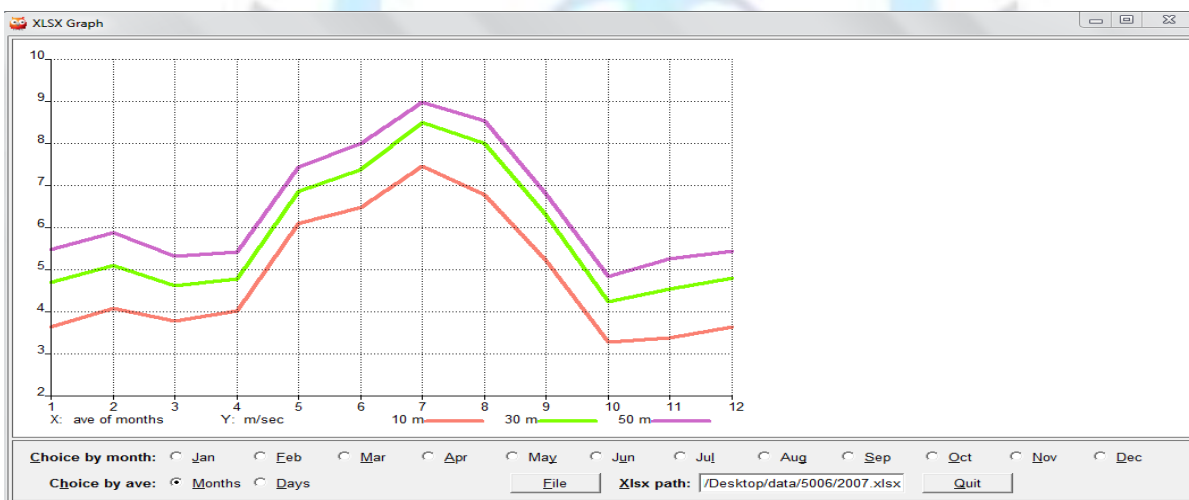


Figure 6b: Annual mean wind speed for the year 2007

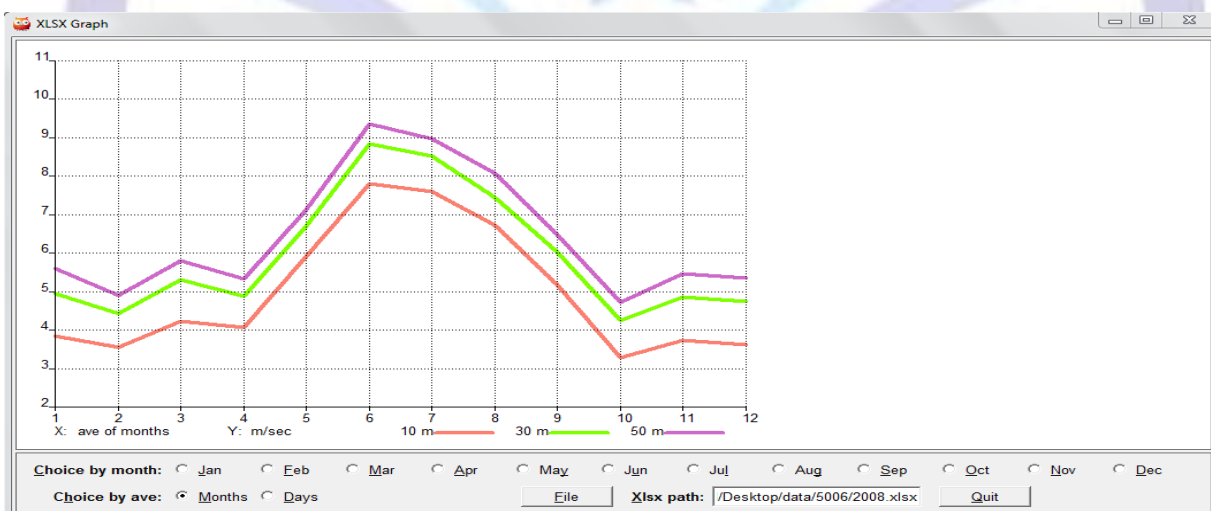


Figure 6c: Annual mean wind speed for the year 2008

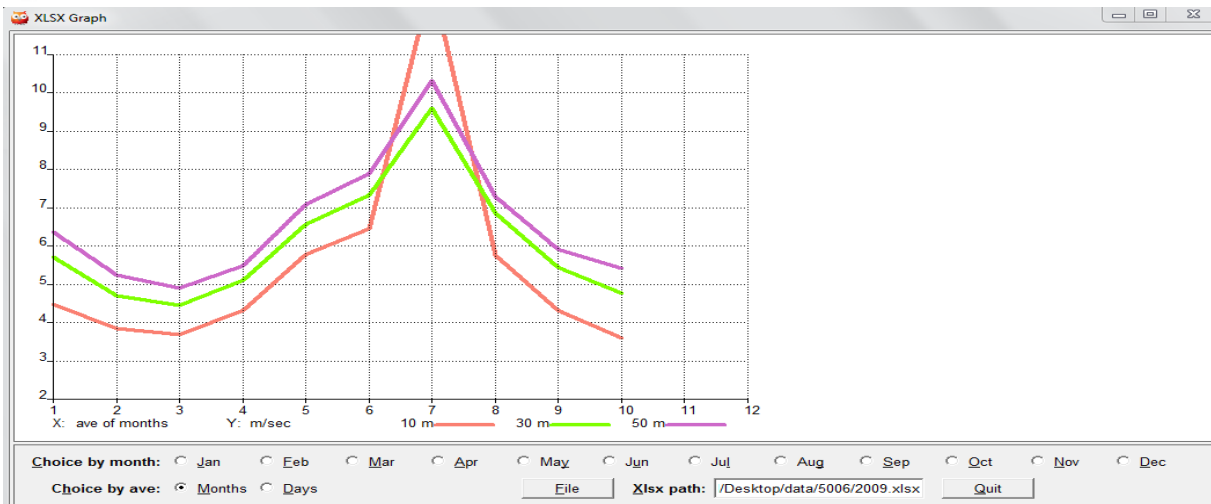


Figure 6d: Annual mean wind speed for the year 2009

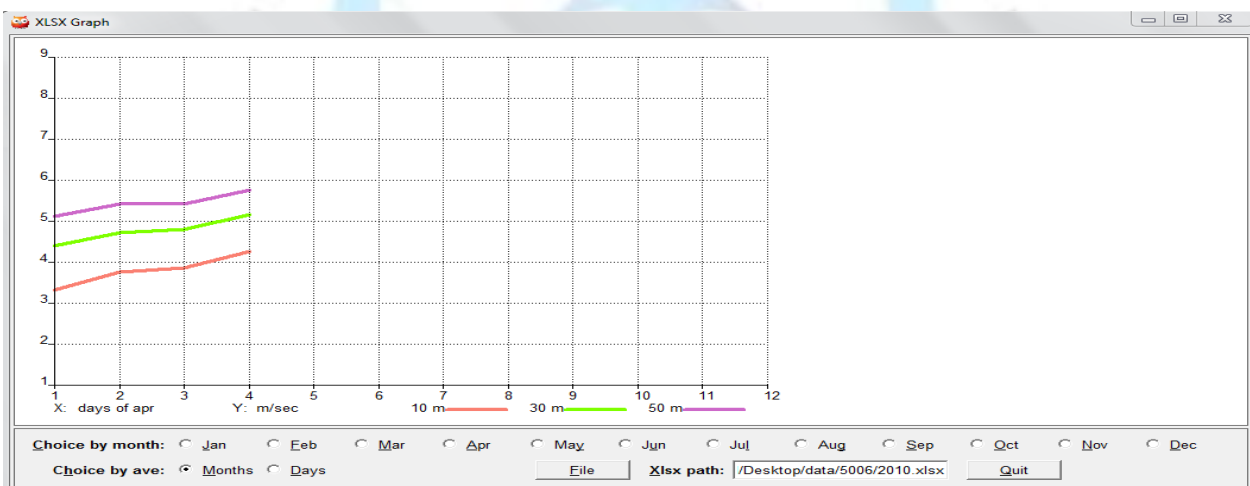


Figure 6e: Annual mean wind speed for the year 2010

Table 3: Annual Mean Wind speed (Vm) and Standard Deviation (SD) for the give site

Height	Years									
	2006		2007		2008		2009		2010	
	Vm	SD	Vm	SD	Vm	SD	Vm	SD	Vm	SD
10mts	4.5282	2.2722	4.8263	1.9144	4.9755	2.0312	5.2373	5.7790	1.1307	1.7790
30mts	5.4140	2.5558	5.8290	2.0844	5.9258	2.1674	5.8934	1.9966	1.4247	2.2371
50mts	6.0836	2.7071	6.4518	2.1262	6.4466	2.2054	6.4712	2.0661	1.6279	2.5448

Step 3: Weibull and Rayleigh plots

The mean wind speed value and standard deviation calculated were used to calculate the Weibull parameters k , c and power.

The plots below display the probability distribution function, cumulative distribution function and power for the given site.

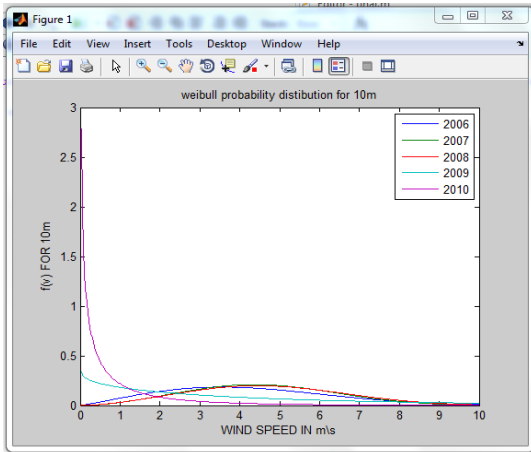


Figure 7a

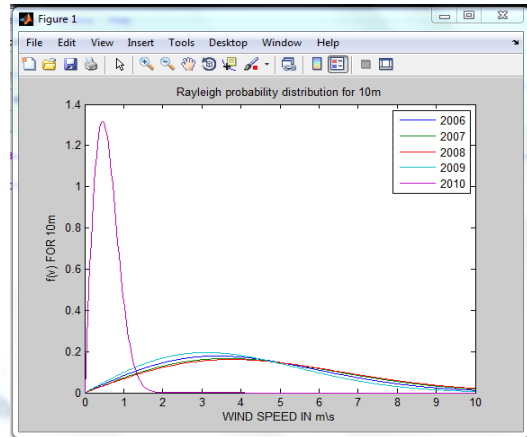


Figure 7d

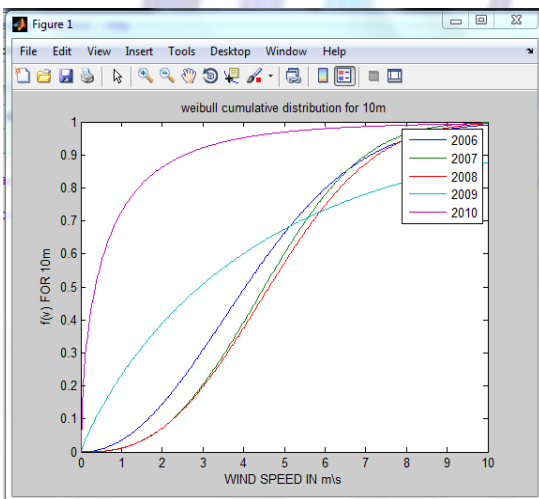


Figure 7b

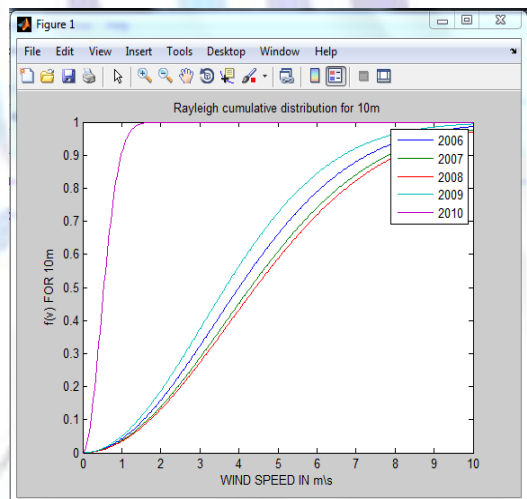


Figure 7e

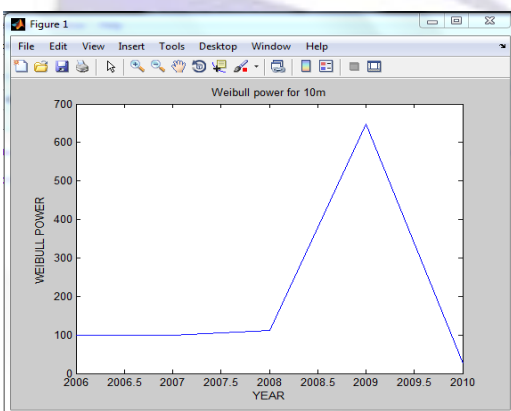


Figure 7c

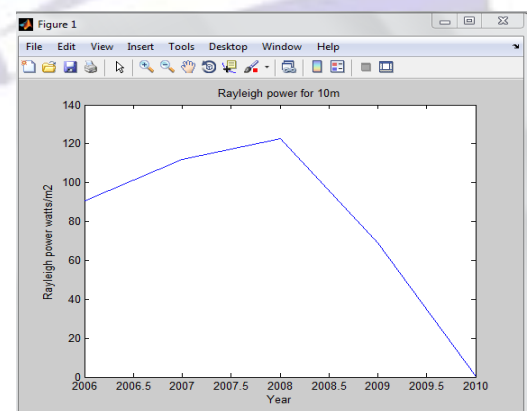


Figure 7f

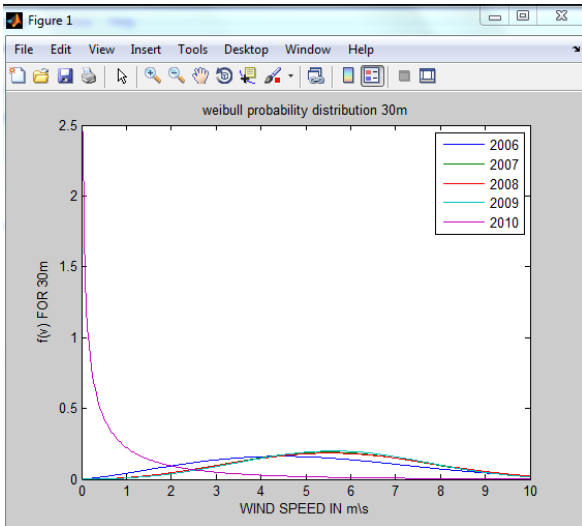


Figure 8a

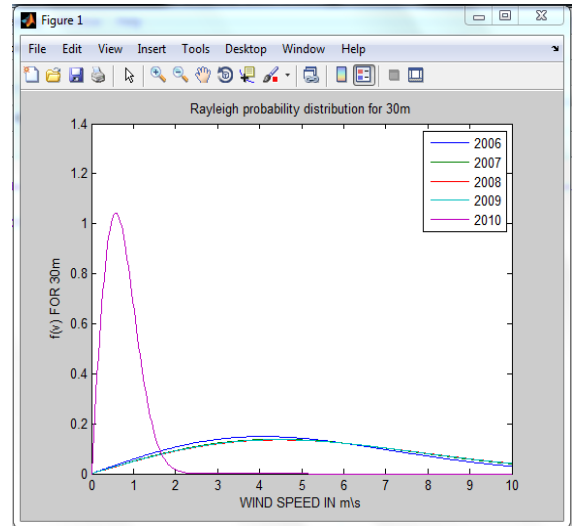


Figure 8d

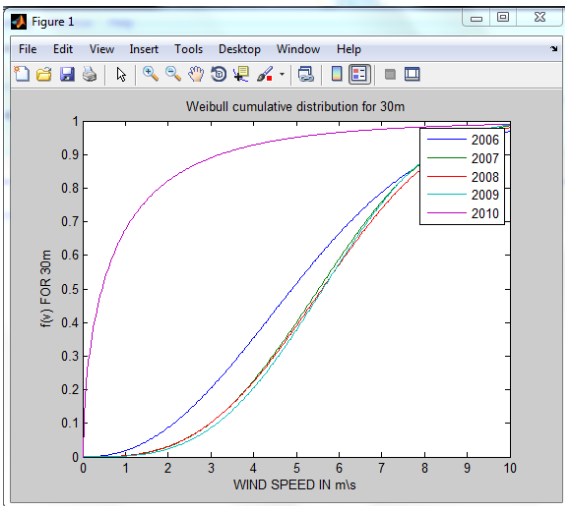


Figure 8b

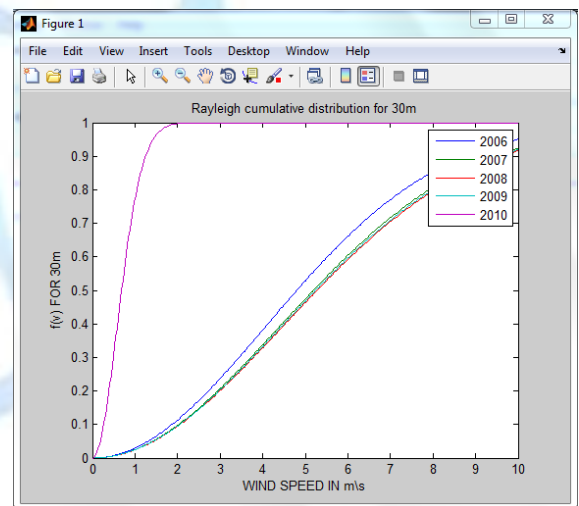


Figure 8e

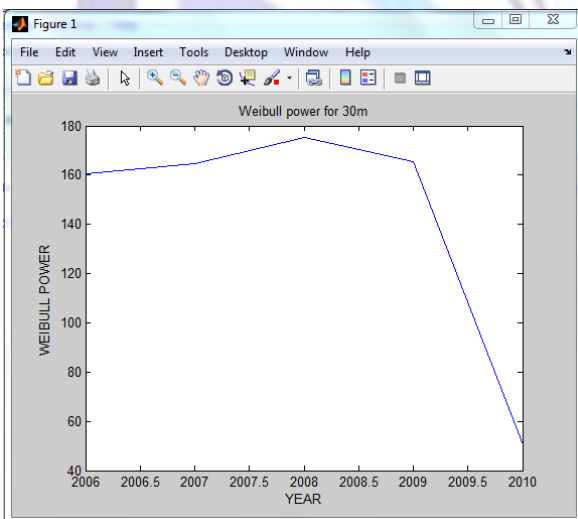


Figure 8c

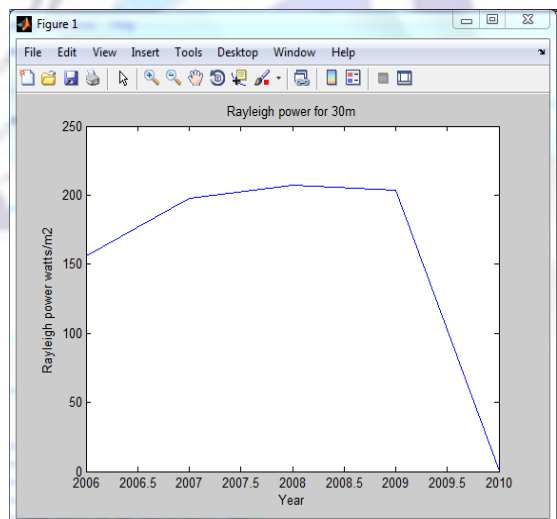


Figure 8f

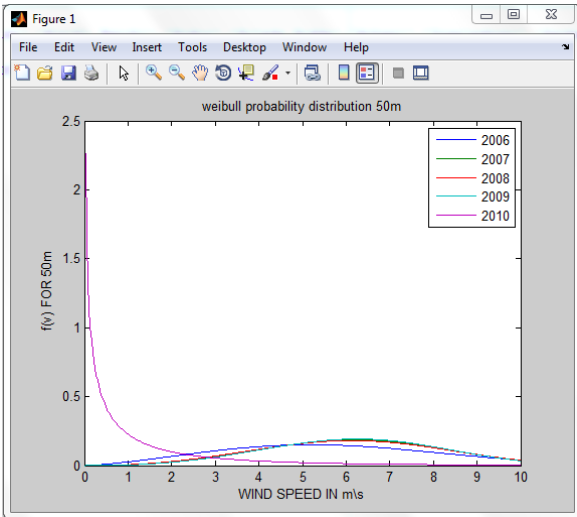


Figure 9a

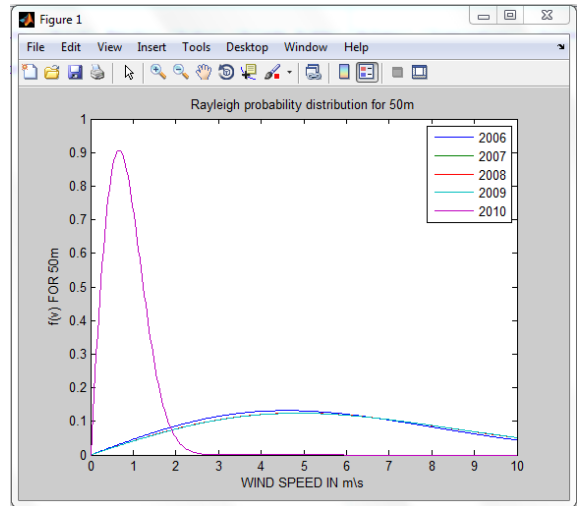


Figure 9d

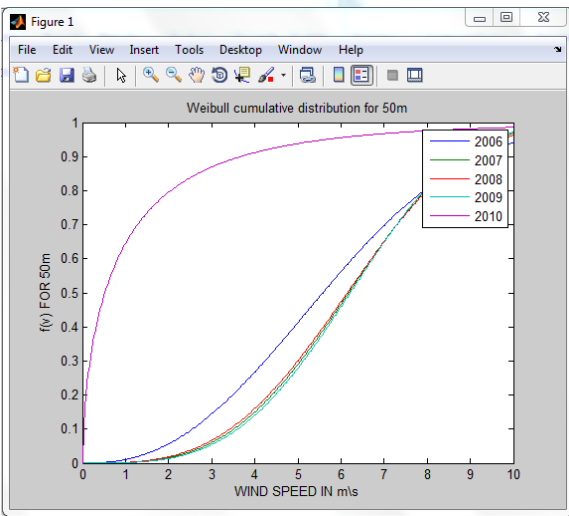


Figure 9b

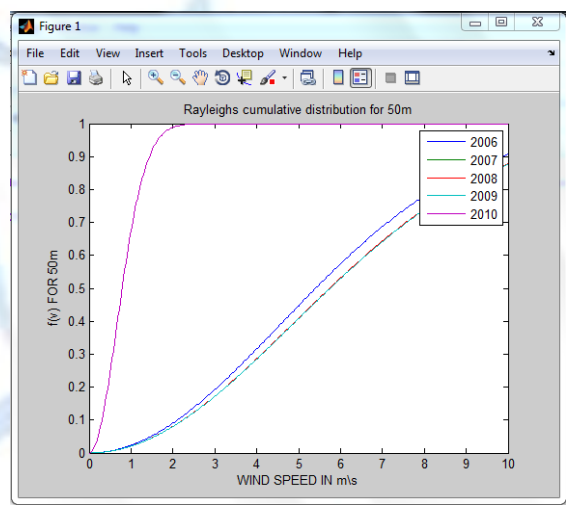


Figure 9e

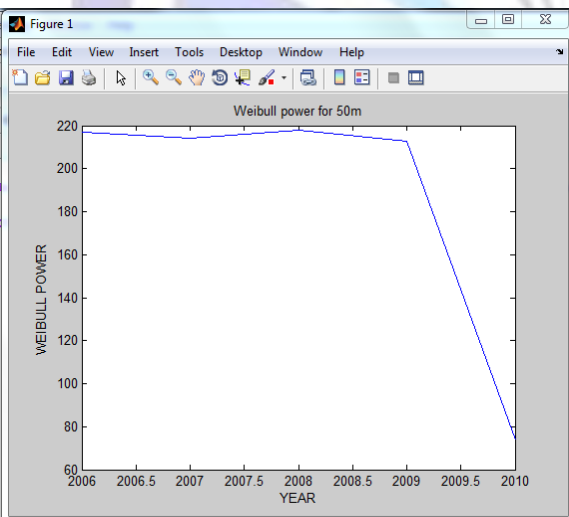


Figure 9c

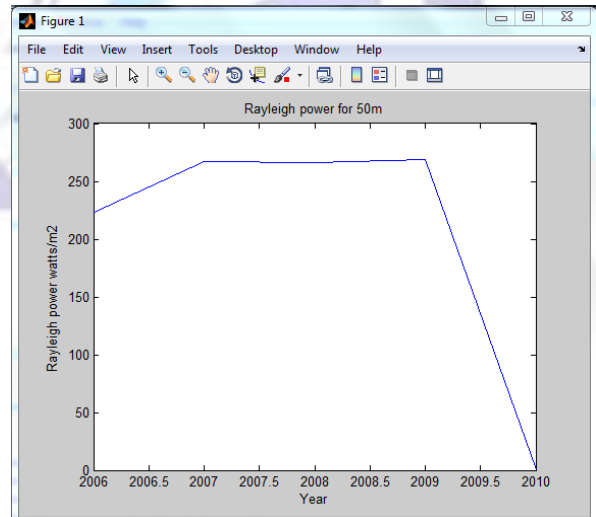


Figure 9f



Table 4: Annual Weibull shape parameter, scale parameter, Weibull Power & Rayleigh Power obtained for given site.

Years	K			c			Weibull Power Watts/m ²			Rayleigh Power Watts/m ²		
	10	30	50	10	30	50	10	30	50	10	30	50
2006	2.1146	2.2595	2.4093	4.8083	5.7674	6.4948	99.4687	160.7305	217.0505	90.5088	156.1874	223.0529
2007	2.7297	3.0551	3.3383	5.1632	6.2357	6.8950	99.6743	164.8519	214.2014	112.0651	197.4023	266.2717
2008	2.6457	2.9810	3.2055	5.3213	6.3400	6.8932	111.3556	175.4348	217.7393	122.6754	207.4783	266.6622
2009	0.8986	3.2396	3.4553	4.4948	6.3009	6.9119	647.6085	165.5295	212.9170	69.1049	203.6621	268.8349
2010	0.6113	0.6126	0.6156	0.6509	0.8226	0.9463	25.5073	50.5417	73.7960	0.2245	0.4531	0.6900

Step 4: ANN plots

Wind speed data was given as input in Matlab for the site; resulting graphs for Levenberg-Marquardt backpropagation are shown below.

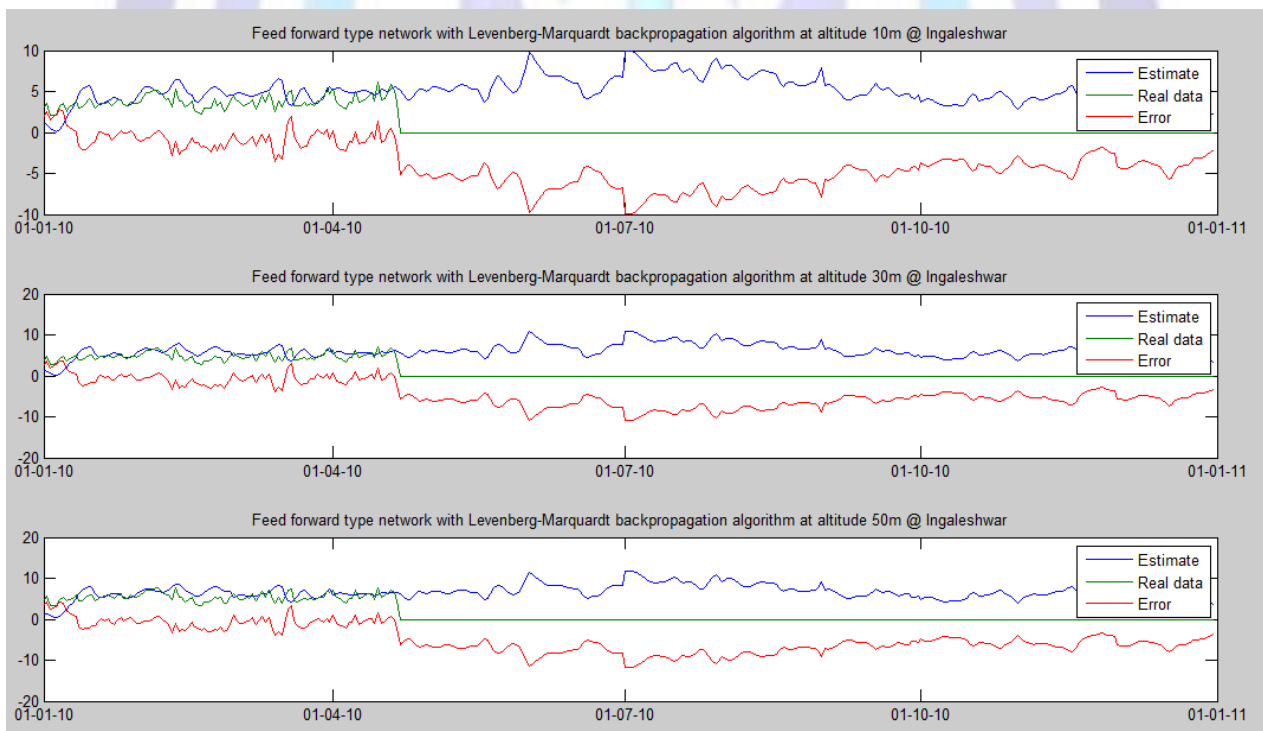


Figure 10: Predicted ANN plots for various heights

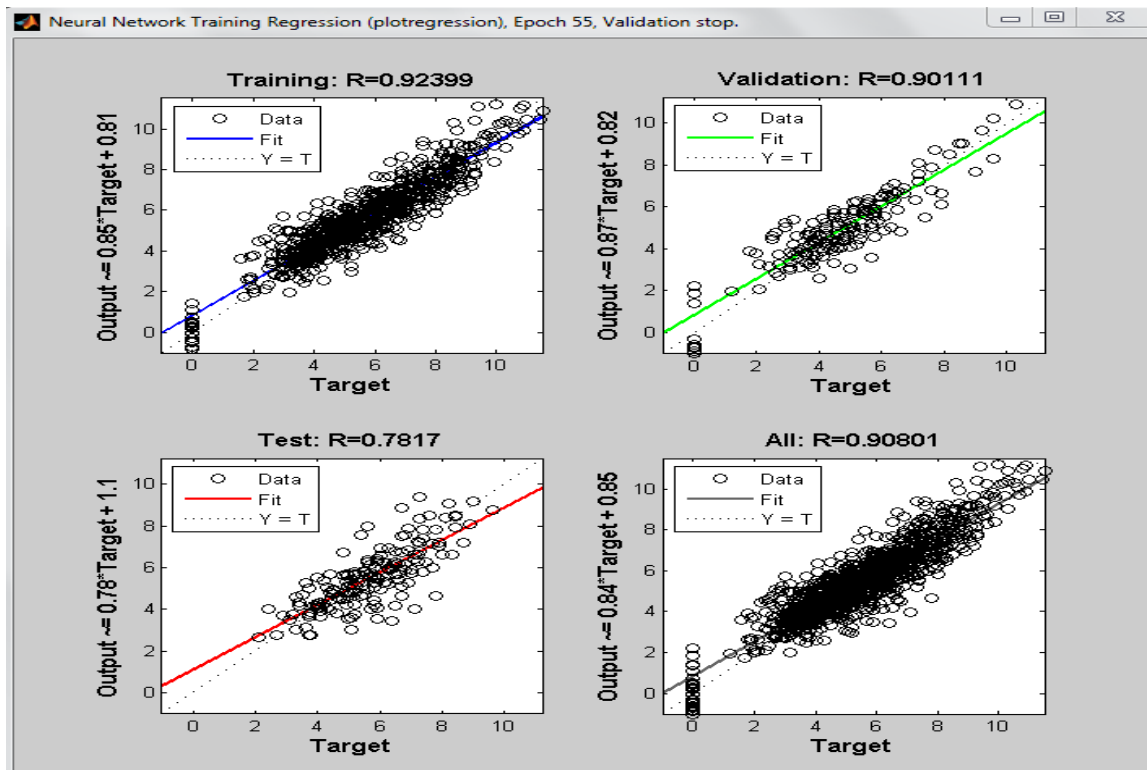


Figure 11: Regression plot for given Data

6. RESULTS AND DISCUSSION

In this paper proposes a data-mining-based methodology for long term wind forecasting (ANN), which is suitable to deal with large real databases. We consider the last five years of wind speed, and results for wind speed forecasting at 10 minutes intervals at various heights. In this study, wind power density was determined by using the Weibull and Rayleigh probability density functions. Moreover, wind speed was predicted using wind speed data of 5 years for site under considerations with Datamining methodology using intelligent technology as Artificial Neural Networks (ANN) and A PROLOG program was also designed and developed to calculate the Monthly/Annual mean wind speed data of the site, which is the first step of prediction of wind speed data.

From the figure 6(a-e) prolog plot we have able to find the monthly mean value of the given site and found to be good in the months of May to August in all the years for all the heights of the site under considerations and the average mean value of wind speed data for various heights and standard deviations is been tabulated in table 3. The uniqueness of the prolog program is so designed that it gives flexibility to calculate Annual and monthly mean wind speed data of the given site.

From the figure 7(a-f), figure 8(a-f), figure 9(a-f) and table 4 shows all the values tabulated for 10meters, 30meters and 50meters Weibull probability distribution, Weibull cumulative distribution and Weibull power plots and similarly for Rayleigh probability distribution, Rayleigh cumulative distribution and Rayleigh power for the given site under consideration. Typically, the shape factor ranges between 1 and 3. Within this range, for small k values, the Weibull distribution tends towards an exponential distribution, and for high k values the Weibull distribution tends towards a Gaussian distribution. A Weibull distribution with a shape factor of 2 becomes a Rayleigh distribution. The Rayleigh distribution is often referred to as a reference frequency distribution. With higher k values, the Weibull curve becomes taller and narrower, which in the wind speed range of the investigated wind energy projects 8 – 10 m/s, implies that both very low wind speeds and very high wind speeds are less frequent than at Weibull curves with lower k values. Since the influence of wind speed on the power contained in the wind flow is cubic, especially the shape of the tails of the Weibull curves has impacts on the power density in these environments, and this is as seen in the power cure in the plot and tabulated values.

From the figure 10 and figure 11, the ANN plots of the wind speed prediction at the given site under consideration is been plotted and the, Existing literatures reveal that hitertoe, computer based Wind prediction are implemented various Datamining techniques. This requires a lot of computational power and several researches are ongoing to enhance this approach. This work lays the foundation for a novel approach in implementing prolog to find Mean wind speed and wind prediction using ANN. Since the efficient training of an ANN within an optimum time frame is always a major requirement specification, we have been able to empirically obtain a backpropagation training algorithm that meets this requirement(i.e. the Levenberg-Marquardt(trainlm) algorithm has the best performance with regression nearly equal to 1.



7. CONCLUSION

This paper presents a statistical and datamining based methodology for wind speed prediction. The paper introduces a datamining technique that is ANN to predict the wind speed using LM Backpropagation algorithm resulting in a good prediction with least error and a best backpropagation algorithm. The paper also introduces a novel concept of using prolog to estimate the mean values of the wind speed data. The Statistical analysis of wind speed prediction shows that Weibull distribution is more suitable than Rayleigh distribution and by seeing the values of the k we can conclude that Higher values of k imply a sharper maximum in the frequency distribution curve and consequently a lower wind power density.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the Karnataka Renewable Energy Development Limited (KREDL), a nodal agency by the ministry of non conventional energy resources Govt. of India for providing the wind data resource assessment. The authors also acknowledge Jain University and New Horizon College of Engineering for supporting in publishing of this research paper.

REFERENCES

- [1] Yuan-Kang Wu and Jing-Shan Hong, "A literature review of wind forecasting technology in the world," Power Tech, 2007 IEEE Lausanne, pp.504-509, 1-5 July 2007.
- [2] Kulkarni, Makarand A., Patil, Sunil, G. V Rama, P. N Sen, "Wind speed prediction using statistical regression and neural network," Journal of Earth System Science, vol. 117, issue 4, pp. 457-463, 2008.
- [3] Rajesh G. Kavasseri, and Krithika Seetharaman, "Day-ahead wind speed forecasting using f-ARIMA models," Renewable Energy, vol. 34, Issue 5, pp. 1388-1393, May 2009.
- [4] Mohammad Monfared, Hasan Rastegar, and Hossein Madadi Kojabadi, "A new strategy for wind speed forecasting using artificial intelligent methods," Renewable Energy, vol. 34, Issue 3, pp. 845-848, March 2009.
- [5] George Sideratos and N.D. Hatzigiorgiou, "Application of Radial Basis Function Networks for Wind Power Forecasting". ICANN 2006, Part II, LNCS 4132, pp. 726 – 735, 2006.
- [6] R. E. Abdel-Aal, M. A. Elhadidy, and S. M. Shaahid, "Modeling and forecasting the mean hourly wind speed time series using GMDH-based abductive networks," Renewable Energy, vol. 34, Issue 7, pp. 1686- 1699, July 2009.
- [7] C. Potter and M. Negnevitsky, "Investigating Using Stochastic Methods to Generate Training Data for Windpower Prediction", Proceedings of the Australasian Universities Power Engineering Conference, vol. 1, 2005, pp. 347-352.
- [8] G. N. Bathurst, J. Weatherill, and G. Strbac, "Trading wind generation in short term energy markets", IEEE Transactions on Power Systems, vol. 17, 2002, pp. 782-789.
- [9] S. Li et al., "Using neural networks to estimate wind turbine power generation", IEEE Transactions on Energy Conversion, vol. 16, 2001, pp. 276-282.
- [10] J. Waewsak, C. Chancham, M. Landary and Y. Gagnon, "An analysis of wind speed distribution at Thasala, Thailand" *Journal of Sustainable Energy and Environment* 2 (2011) 51-55.
- [11] Dr. S. Gomathinayagam, K. Boopathi " Course Material on Wind Resource Assessment and Wind technology for State Nodal Agencies (SNAs) India", organized by C-WET-2013.
- [12] M. Balduccini, M. Gelfond and M. Nogueira. A-Prolog as a tool for declarative programming.
- [13] In Proceedings of the 12th International Conference on Software Engineering and Knowledge Engineering (SEKE'2000), 63-72, 2000.
- [14] Mert Y, Usta KI, Analysis of wind speed distributions: wind distribution function derived from minimum cross entropy principles as better alternative to Weibull function, *Energy Conversion and Management* 49 (2008) 962-973.
- [15] Patel MR, *Wind and Solar Power Systems* (1999) CRC Press.
- [16] Ramchandra, T, V. Rajeev, K. J. Vansee, and Sruthi B, 2005 *Energy Education science and Technology*, 14, 61.
- [17] Akpınar EK, Akpınar S, Determination of the wind energy potential for Maden, Turkey, *Energy Conversion and Management* 45 (2004) 2901–2914.
- [18] Papoulis A, Pillai SU, *Probability, random variables, and stochastic processes* (2001) 4th Edition.
- [19] Joseph P, Hennessey J, Some aspects of wind power statistics, *Journal of Applied Meteorology* 16 (1977) 119-128.
- [20] Walker, J. F and Jenkins. S (1997): *wind Energy Technology*, John Wiley and Sons, Chichester.



- [21] Gipe. P, (2004) : Wind Power: Renewable energy for home farm and businesses, Chelsea Green USA.
- [22] Indumathy. D, Sessaiah. C. V, Sukkiramathi. K, "*Estimation of Weibull Parameters for wind speed calculation at Kanyakumari in India*", IJRSET, Volume 3, Issue 1, 2014 ISSN. NO. 2319-8753.
- [23] Bratko. PROLOG Programming for Artificial Intelligence, 3rd ed. Addison-Wesley, Harlow, 2001.

