Wind energy potential estimation with prediction of wind speed distribution

Manish Kumar* and Cherian Samuel

Department of Mechanical Engineering, Indian Institute of Technology (BHU), Varanasi – 221005, India
Email: mkumar.rs.mec13@itbhu.ac.in
Email: csamuel.mec@itbhu.ac.in
*Corresponding author

Abstract: Integration of renewable distributed generation technology with radial power distribution system is rapidly growing. Among the number of available renewable energy sources, the role of wind energy in the power sector is very important. In this work, authors have estimated the wind energy potential at Banaras Hindu University (BHU), India. Estimation of wind energy potential depends on statistical analysis of wind speed distribution, which will give us the best-fitted probability density function of the desired location. For statistical analysis of wind speed distribution, authors have taken hourly data of BHU centre from the source of National Renewable Energy Laboratory (NREL), USA for the 12-month period. A statistical R-programming language tool has been used for the computational work. In this work, authors have considered Weibull, Gamma and Lognormal probability distributions for getting best-fitted distribution. After the analysis, result shows for the given wind speed data of BHU area, Weibull distribution is the best-fitted one.

Keywords: statistical analysis; wind speed; wind energy potential; renewable energy; renewable distributed generation; probability distribution; goodness-of-fit.


Biographical notes: Manish Kumar is currently pursuing his PhD with the Mechanical Engineering Department, Indian Institute of Technology (Banaras Hindu University), Varanasi, India. He received his BTech in Mechanical Engineering from Biju Patnaik University of Technology, Rourkela, India, and MTech in Industrial Engineering and Management from the Indian Institute of Technology (Indian School of Mines), Dhanbad, India. His current research interests include renewable energy planning, energy economics, operations management, decision science, optimisation and soft computing.

Cherian Samuel is an Assistant Professor in the Department of Mechanical Engineering at the Indian Institute of Technology (Banaras Hindu University), Varanasi, India. He received the Bachelor’s in Mechanical Engineering from Calicut University, India; the Master’s in Production Engineering from Calicut University, India; and the PhD in Industrial Engineering and Management...
1 Introduction

In recent past, rapidly growing global warming has been the major threat to the human being. Overcoming from global warming issues, globally all developed and developing countries meet each year in a summit of environmental issues. Where they do policy revision to improve the current environmental condition and maintain our ecology system. With this kind of global meets, the world came together to achieving a milestone in renewable energy sector to reduce the load of carbon emissions and improving world environment condition. For promoting renewable energy, developed countries are supporting developing countries for increasing utilisation of renewable energy sources with technological and financial support. Environmental friendly nature of renewable energy sources is increasing their importance globally despite their high initial setup coast which is going down with technological advancement. Initially, our attraction over renewable energy sources was due to abundant availability of this and limited conventional fossil fuels. Now, environmental issues are being more concerned area to motivate for the adoption of renewable energy sources in the power sector. With the use of wind energy-based renewable distributed generation technology concept, we can reduce green house gasses and adverse environmental effects with power quality improvement (Atwa and El-Saadany, 2011). With technological advancement, increased renewable energy-based generation capacities are also helping to reduce per unit cost.

Small-scale generating capacity of renewable energy-based technology gives us the freedom of setup decentralised generating units, which called renewable distributed generation technology. This new renewable distributed generation technology concept is helping us to have access to rural or remote areas, where to set up of transmission line would not be possible or feasible (Asrari et al., 2012; Rajanna and Saini, 2016). Integration of renewable energy sources to urban building in India has been discussed in the work of Kumar and Ravikumar (2015). Among the different alternatives of renewable energy sources, wind energy is one of the major sources and having maximum contribution in power generation. In a work of Sholapurkar and Mahajan (2015), potential of wind energy and their growth in India has been discussed. Kose et al. (2014) analysed potential of wind energy at Selcuk University, Turkey in their work. Integration of wind generators with distribution networks can improve voltage stability through reactive power management (Roy et al., 2013). For the reactive power management, a new distributed multi-agent scheme has been developed by Rahman et al. (2014) to integrate the renewable energy sources in distribution networks. Maintaining resilience of transmission and distribution networks with increasing penetration of renewable distributed generation is very important (Akter et al., 2014). Managerial challenges of wind energy are availability of wind speed and economical feasibility of the desired location.

Developing countries are also following footsteps of developed countries to have adoption of renewable energy technologies and reducing load over conventional energy sources. Attention over renewable distributed generation technology has also been
Wind energy potential estimation with prediction of wind speed distribution

growing at the institution and industrial level. Researchers are investigating to harness more energy from the renewable sources and reducing its per unit cost. Globally, wind energy is the major contributor in generation of electricity and helping to reduce carbon emissions. In wind energy case, now we can harness electricity at offshore area or low wind speed region. The contribution of wind energy in the energy sector is growing rapidly over the worldwide. Now, total world’s wind power capacity has reached at 432,419 MW on December 2015 as per GWEC Global Wind Statistics (2015) report, and with rapid expansion of wind energy it has been reached around 4.7% contribution of total electricity usage over worldwide (The World Wind Energy Association Half-year Report, 2016). Countries like China, the USA, Germany, India, Spain and Denmark are contributing majorly in wind energy sector with China (145,362 MW), the USA (74,471 MW) and Germany (44,947 MW) are ahead of India (25,088 MW) in fourth position as per Global Wind Report (2015). Less installation period of wind turbines and functioning make this more favourable in India and a form of renewable distributed generation technology wind energy also helps to enhance reliability and performance of power distribution system (Khare et al., 2013). In a report of the expert group on 175 GW Renewable Energy by 2022, the Ministry of New and Renewable Energy (MNRE) set the target of 60 GW for wind power generation capacity to be achieved by the year 2022. Here in Figures 1 and 2, we can see the annual capacity addition of renewable energy and wind energy in Indian energy sector, respectively. Where x-axis represents samples of last six years and y-axis represents every year total installed capacity of renewable energy and wind energy in megawatt unit. We can see in Figures 1 and 2, the potential growth rate of renewable energy and wind energy in India for the past six years. We can analyse in last few years installation capacity of wind energy has significantly increased. In 2011, capacity addition of wind energy in India was record highest 3 GW of new installation capacity (Global Wind Report, 2015). In this sector, Indian states like Tamilnadu, Gujrat, Rajasthan and Maharashtra are contributing majorly.

Figure 1 Annual installed capacity of renewable energy in India

![Figure 1](image1)

Figure 2 Annual installed capacity of wind energy in India

![Figure 2](image2)
Wind power has been developed in India on 1986 with the first wind farms installation in coastal areas of Gujrat (Okha), Maharashtra (Ratnagiri) and Tamil Nadu (Tuticorin) with 55 kW capacities of Vestas wind turbines (Wind Power in India). Potential of the wind energy in India has been first analysed by Dr. Jami Hossain (Hossain and Sharma, 2015) with the help of GIS system around of 3000 GW in the year 2011. In next year, analysis of Dr. Jami Hossain has been subsequently revalidated by Lawrence Berkley National Laboratory (LBNL), the USA in an independent study. With this study, the MNRE created a committee for the reassessment of the wind energy potential (Hossain and Sharma, 2015) and had announced a new estimation of the potential wind resource in India from 49,130 MW to 300,000 MW at 100 m hub height in a report of Estimation of Installable Wind Power Potential at 80 m level in India. The analysis result of a demonstration project, sponsored by MNRE, has shown the increasing wind resource with increasing Hub heights.

Optimal utilisation of wind energy demands a good knowledge base for the installation areas. For this, wind speed distribution plays a key role to have optimal utilisation of wind energy. In this work, authors have investigated the BHU region for the feasibility analysis of wind energy potential to install renewable distributed generation units. Hourly data of wind speed of BHU area has been collected from the NREL for the period of 12 months in between year 2002 and 2011 (India Solar Resource Data: Hourly Data and TMYs). Authors proposed different probability density functions for statistical analysis and computed with R-programming language. For the statistical analysis, authors applied Weibull, Gamma and Lognormal continuous distributions which vary between zero and infinity. With the help of goodness-of-fit test, authors identified the best-fitted probability density function for wind speed data of BHU area. Authors used Kolmogorov, Anderson–Darling and Cramer–von Mises tests for the goodness-of-fit test.

Practically it is very important to analyse the variation of wind speed for optimising design of the systems, resulting in less energy generating costs. From the last few years, multiple researchers have been conducted various studies to assess wind power over the worldwide. Multiple previous works are showing the wind variation for a typical site is represents the Weibull distribution (Abdraman et al., 2016; Bilir et al., 2015; Keyhani et al., 2010; Kwon, 2010; Mohammadi et al., 2016; Shu et al., 2015; Usta, 2016; Weisser, 2003). Some works have been done over comparative analysis between Weibull and Rayleigh distributions to predict wind speed distribution (Celik, 2004; Kose et al., 2014; Olaofe and Folly, 2012; Pishgar-Komleh et al., 2015; Togrul and Ertekin, 2011; Türk Toğrul and İmaş Kizi, 2008). Some authors developed new distribution function and advocate other than conventional Weibull distribution for the estimation of wind speed distribution (Akgül et al., 2016; Akpinar and Akpinar, 2007; Akpinar and Akpinar, 2009; Chang, 2011; Harris and Cook, 2014; Kantar and Usta, 2008; Kollu et al., 2012). In these previous studies, major consideration has been given to the two parameters (shape and scale) Weibull distribution, because the previous study has been shown that fitness of Weibull distribution with a wide collection of wind speed data. Despite of having major contribution of Weibull distribution for the estimation of wind speed distribution, multiple researchers identified more fitted probability distribution other than the Weibull distribution. On the basis of these studies, selection of conventional Weibull distribution without any comparative analysis may be a wrong choice.

In this work, our objective of this study is to statistically analyse the wind speed data of BHU area to predict the wind speed distribution and be able to estimate the wind
energy potential. We will have to identify the best fitted probability density function for the region of BHU. Presentation of the paper is: Section 2 is about data collection and organising them; Section 3 is theoretical representation of proposed distribution models; Section 4 is about three goodness-of-fit tests; Section 5 is statistical analysis part; Section 6 is results and discussion; and Section 7 concludes this work.

2 Data collection and organisation

Estimation of wind energy potential for a particular location requires analysis of available wind speed data for long or small periods. Lalas (1985) stated that the drawing out of meteorological data based on average wind speed for a period, duration curves, direction, power spectra of wind speed and its variation with heights. Generally, for the assessment of wind speed at desired location 10-year period of measurement will be required (Nfaoui et al., 1998). As per Frandsen and Christensen (1992), with comparison of long period wind speed data a short period, for example, one-year wind speed data may be suitable for wind energy potential estimation with 5–15% uncertainty. Use of the various types of instruments, types of equipment and site specifications for assessment of wind energy has been briefly explained in the work of Alawaji (1996). For getting wind speed data for a specific location, meteorological towers with cup anemometers and wind vanes are the major factors for getting the information of wind energy potential at measurement height from the ground level.

Figure 3 Indian wind power density map at 50 m height (see online version for colours)
Recently National Centres for Environmental Prediction (NCEP) or National Centre for Atmospheric Research (NCAR) and European Centre for Medium-Range Weather Forecasts (ECMWF) databases of wind speed have been developed globally for many regions (Landberg et al., 2003). With high-resolution Sonic Detection And Ranging (SODAR), Light Detection and Ranging (LIDAR) or satellite like ground-based remote sensing instruments have used as alternatives to meteorological towers for assessment of wind energy potential (Angelis-Dimakis et al., 2011). Choisnard et al. (2004) and Christiansen et al. (2006) presented a methodology for assessment of wind energy potential using a series of satellite synthetic aperture radar (SAR) images, this technique is particularly useful for regions where unavailability of year-long time series data, such as offshore regions. A method of ‘round robin site assessment’ has been investigated by Lackner et al. (2008) for the use of alternative monitoring strategy in wind energy potential assessment. After collecting wind speed data from any given methods will further utilised for feasibility analysis. Figures 3–5 show availability of wind energy potential at different region for mast height. Figure 3 shows regional distribution of wind energy potential at 50 m height, where Figures 4 and 5 show wind energy potential in India for different region at height 80 and 100 m. This shows wind energy potential at BHU campus lies in moderate region.

Figure 4  Indian wind power density map at 80 m height (see online version for colours)
All the data have been taken in Local Standard Time (LST). India is ahead of Greenwich Mean Time (GMT) with 5.5 or five and a half (5:30) hours. All the data is given in the form of cells with particular latitude and longitude information. LST is selected based on the cells location with referenced coordinate information. 12-month hourly data files in between the period 2002 and 2011 are available for download in a compressed archive (tar.gz) files, and each archive file contains a list of comma separate values (CSV) files (India Solar Resource Data: Hourly Data and TMYs). Here file names show our locational longitude and latitude of the wind speed cell centre which is determined to be the nearest coordinates entered into the Indian wind energy resource and that produce hourly computer generated imagery (CGI) data. Each and every cell covers an area on the surface of the Indian continent measuring 0.1° of latitude by 0.1° of longitude with the given cell coordinate which represents the cell’s centre point. In this study, the cell of BHU area represents the area of 25.16°N to 25.26°N and 82.89°E to 82.99°E.

We considered at least 1-year long measurement of data from an observation centre is required to determine the wind power potential and project feasibility. Wind speed data of the BHU region have been taken from the source centre by use of anemometers placed at the 2-m height of the measurement mast from the ground level. Elevation of the BHU centre is 265 ft from the sea level, and it is located in the centre of gangetic region. With 10-minute intervals, data logger recorded measurement of the parameters at the observation centre in every second, their average values, minimum and maximum values with their standard deviation. For computing purpose, CALLaLOG 98 software program was used. Hourly data have been recorded in archive file by the CALLaLOG 98 software and stored in daily and monthly folders. Statistical analysis of wind speed data given by CALLaLOG 98 software has been done from the statistical analysis tool R-programming language. Statistical R-programming language is widely using by modern time researcher
for the statistical analysis of data or pattern recognition. Missing data have been interpolated in this work, and this was only 0.06% in proportion, which was far less than of acceptable limit of 10% (AWS Scientific, Inc., 1997). 12-Month period of hourly data of meteorological parameters like air temperature, solar irradiation, barometric pressure and relative humidity apart from the wind speed data have been measured by the CALLaLOG 98 software and cup anemometer. The temperature variation at BHU area lies between maximum and minimum monthly average temperature 44°C in summer and 6°C in winter, respectively.

An important factor for the characterisation of the wind resource is the variation of wind speed with height above the ground. In this respect, the wind speeds were measured by one anemometer located at 2 m height on the mast. The sample of hourly distributed wind speeds measured between 1 January, 2002, and 31 December, 2011. For the desired height at 100 m average monthly variation of wind speed has been shown in Figure 6. For the conversion of wind speed data from the reference point to the desired level, we used power law. In Figure 6, we can see the monthly variation of average wind speed at BHU area, where maximum speed in month April and minimum in January month.

Figure 6  Month average wind speed values at BHU campus

The power law is used in wind power assessments for getting the wind speed values at wind turbine height from the surface level observation. On the basis of design of different kind of wind turbines it lies in between 50 m and 100 m heights, whereas observation of data may be over 2 m or 10 m height in normal cases (Manwell et al., 2010). For analysis purpose, we use a standard height to convert ground level data to standard level, which is 100 m in our case, because of normal height of maximum wind turbines and limitation of logarithmic profile of wind speed within surface layer of atmospheric boundary layer.

We use generated wind profiles in the form of numerous atmospheric pollution dispersion models. Wind profile gives the variation or changes of horizontal wind speed with vertical distribution. The wind profile of the atmospheric boundary layer is logarithmic in nature which is best represented by the log wind profile equation with consideration of surface roughness and atmospheric stability. In unavailability of surface roughness and atmospheric stability information, wind profile power law equation is often used in a place of log wind profile. Representation of the wind profile power law is (Justus and Mikhail, 1976):
Here \( u \) is the wind speed (m/s) of the desired level of height \( h \) (m) and \( u_r \) (m/s) represents the available wind speed information of the reference level height \( h_r \) (m). An empirically derived coefficient \( p \) (unit-less) varies with stability of the atmosphere. In neutral stability conditions, value of \( p \) is approximately 1/7, or 0.143 (Hsu et al., 1994). In simplified form, estimation of the wind speed at the desired level of height \( z \), the above power law relationship would be rearranged to:

\[
\frac{u}{u_r} = \left( \frac{h}{h_r} \right)^p
\]

The difference between the reference level and desired level is less (<100 m) for introducing substantial errors into the estimation of wind speed at desired level, so 1/7 value of \( p \) becomes constant in wind resource assessments. The constant value of \( p \) does not account for the roughness of the surface, the stability of the atmosphere, or in the presence of obstacles the displacement of calm winds from the surface. In places like mountain or forest regions, where trees or structures impede wind speed near the surface level, the use of constant value 1/7 of \( p \) may yield an error in estimates; the log wind profile would be preferable. Even in neutral stability conditions, for offshore wind farms over open water, 0.11 would be more appropriate rather than 0.143 (1/7). In our case, we have the wind speed data at the height of 2 m and requirement is at 100 m and our location lies in a plane area with the neutral condition. So, from the power law with \( p \) value 1/7, we have calculated the hourly wind speed data of BHU area at 100 m height.

Measured wind speed data are commonly available in time-series format, in which each data point represents either an instantaneous sample wind speed or an average wind speed over a given period. In some places, we can see wind speed data in frequency distribution format. This represents the frequency of wind speed distributed within various ranges (bins).

## 3 Wind speed distribution models

### 3.1 Statistical analysis

Wind speed distribution modelling requires analysis of hourly distributed wind speed data for the analysis period. For minimising the expenses and time to process long-term wind speed data, we can use statistical probability density functions for the prediction of randomly distributed wind speed data. The primary tools to describe wind speed characteristics are probability density functions. The parameters of probability density functions which describe wind speed frequency distribution are estimated using statistical data of analysis period. Many probability distribution functions have been proposed in recent past, but in the present study for the statistical analysis Weibull, Gamma and Lognormal distributions are used to identify the appropriate probability distribution. For this, authors used R-programming language to analyse the distribution of wind speed data and goodness-of-fits results. Parameters of each distribution have been calculated.
from the same computational tool. Representations of the mathematical formulation of Weibull, Gamma and Lognormal distributions are:

### 3.1.1 Weibull distribution
Mathematical representation of the probability density function of two parameters Weibull distribution with \( c \) being the scale parameter and \( k \) the shape parameter is

\[
f(x; c, k) = \left(\frac{k}{c}\right)^k \left(\frac{x}{c}\right)^{k-1} \exp\left[-\left(\frac{x}{c}\right)^k\right]
\]

Weibull CDF is written as

\[
F(x; c, k) = 1 - \exp\left[-\left(\frac{x}{c}\right)^k\right]
\]

where mean and variance of the distribution with \( \Gamma \) (Gamma function) are

\[
E[X] = c \Gamma\left(1 + \frac{1}{k}\right)
\]

\[
Var[X] = c^2 \left[\Gamma\left(1 + \frac{2}{k}\right) - \left(\Gamma\left(1 + \frac{1}{k}\right)\right)^2\right].
\]

### 3.1.2 Gamma distribution
Probability density function of Gamma distribution with \( \alpha \) (shape) and \( \beta \) (rate) parameters is

\[
f(x; \alpha, \beta) = \frac{\beta^{\alpha} x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}
\]

Gamma CDF is written as

\[
F(x; \alpha, \beta) = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}
\]

where mean and variance of the distribution is

\[
E[X] = \frac{\alpha}{\beta}
\]

\[
Var[X] = \frac{\alpha}{\beta^2}
\]

### 3.1.3 Lognormal distribution
The lognormal distribution is a probability distribution of a random variable whose logarithm is normally distributed. Lognormal probability distribution function with \( \mu \) as location and \( \sigma \) as scale parameters is given by
\[ \ln(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp \left[ -\frac{(\ln x - \mu)^2}{2\sigma^2} \right] \] (11)

Lognormal CDF is written as
\[ LN(x; \mu, \sigma) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{\ln x - \mu}{\sigma \sqrt{2}} \right) \right] \] (12)

where mean and variance of the distribution is
\[ E[X] = e^{\mu + \frac{1}{2} \sigma^2} \] (13)
\[ \text{Var}[X] = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2}. \] (14)

### 3.2 Goodness-of-fit tests

Goodness-of-fit tests allow us to identify the suitable distribution function among the different alternatives. In reality, there is no any perfectly matched distribution. Probability density function is the most used tools to estimate and modelling the wind speed distribution. Many authors proved Weibull distribution as a best-fitted probability density function for the wind speed distribution in their work. On the basis of the suitability of data, authors proposed Weibull, Gamma and Lognormal distributions for the study to identify best-fitted distribution in this case. It would be wrong to choose from the list of previous results of specific areas in a different environment. Because the variation of wind speed depends on their regional factors and it may vary region to region and may not fit with respective referenced regional distribution. For error minimisation, we should have to analyse the data and get best-fitted probability distribution. In this work, authors selected continuous distribution functions lies between zero and infinity and have been investigated with the aid of three goodness-of-fit tests namely, Kolmogorov–Smirnov, Anderson–Darling and Cramer–von Mises tests. Brief information of selected goodness-of-fit tests are given below.

#### 3.2.1 Kolmogorov–Smirnov test

The Kolmogorov–Smirnov test is useful when parameters of the distribution did not estimate. With modified Kolmogorov–Smirnov test we can use this where the parameters are estimated from the data. The Kolmogorov–Smirnov test is mainly based on empirical cumulative distribution function (CDF). For the given \( n \) ordered data points \( X_1, X_2, \ldots, X_n \), the empirical CDF is defined as (Daniel, 1990)
\[ E_x = n(i) / N \] (15)
where \( n(i) \) represents the number of points less than \( X_i \) and \( X_i \) are ordered between smallest and largest value. This is increases by \( 1/N \) for the each value of ordered data between smallest to largest.

The advantage of this test is that the distribution of Kolmogorov–Smirnov test statistic will not have need of cumulative function distribution test. Compared to
Cramer–von Mises test it is an exact test, it only applies to continuous distributions. The test is calculated as

\[ D = \max_{1 \leq i \leq N} F\left( X_i \right) - \frac{i}{N} \]  

where \( F \) is the theoretical cumulative distribution for what we are conducting this test which must be a continuous distribution.

### 3.2.2 Anderson–Darling test

Anderson–Darling tests the data whether this lies to the given probability distribution. The test is most often used to test of a given family of distribution, in which case parameters of that family needs to be estimated. Anderson–Darling is a modification of the Kolmogorov–Smirnov test and gives more weight to the tails where as Kolmogorov–Smirnov test gives more to the centre of the distribution. This has the advantage of allowing a more sensitive test. The Anderson–Darling test statistic is defined by (Anderson and Darling, 1952):

\[ A^2 = -n - S \]  

where

\[ S = \sum_{i=1}^{n} \frac{2i-1}{n} \left[ \ln F\left( X_i \right) + \ln \left( 1 - F\left( X_{n-i+i} \right) \right) \right] \]  

and \( F \) represents the theoretical CDF of the given family of distribution. Note that the \( X_i \) is the ordered data.

### 3.2.3 Cramer–von Mises test

The Cramer–von Mises test is used to test the fitness of distribution of a given family of distribution. In statistics the Cramér–von Mises test is used for testing the goodness-of-fit for a cumulative distribution function \( F \) compared to an empirical distribution function \( F_n \) for the given data, or comparing of two empirical distributions. It is also used in part of another algorithm as minimum distance estimation. This is represented as (Anderson, 1962):

\[ w^2 = \int_{-\infty}^{\infty} \left[ F_n(x) - F(x) \right]^2 dF(x) \]  

In this study, here \( F \) is the theoretical CDF for the wind speed data of BHU region, and \( F_n \) is the empirically observed distribution. Two distributions can be empirically estimated ones in a two-sample case. This Cramér–von Mises test is an alternative to the Kolmogorov–Smirnov test.

Let \( x_1, x_2, \ldots, x_n \) are the observed values, in ascending order. The numerical values of test statistic is found as

\[ T = nw^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[ \frac{2i-1}{2n} - F\left( x_i \right) \right]^2 \]  

In equation (19), $w^2$ is derived as difference between empirical distribution function and cumulative distribution for one sample application. Integration of equation (19) and multiply with $n$ will give $T$ as total differences for $n$ observed values.

4 Results and discussion

Figure 7 shows a histogram of wind speed data in 94 class intervals. Histogram has been draw from the use of R-programming language computation. Hines et al. (2008) state that is choosing the number of class intervals approximately equal to the square root of the sample size often work well in practice. Authors used this theory to find out the class intervals to draw the histogram for wind speed data. If you have very few intervals for the larger data, then this will give you the blocky or coarse figure, or if you have very large intervals compared to not much larger data, then this will show you ragged figure. For the selection of best fit, we should choose appropriate class intervals for the given sample size. In this case authors have total sample size is equal to 8762 which gives the value of 94 with approximation after the square root of sample size. Figure 8 is showing graphical representations of Weibull, Gamma and Lognormal distributions of the given data with the help of histogram and theoretical density plots. Authors have drawn the histogram and density curves in 94 class intervals for the given three distributions. From Figure 8, we can see Weibull is the best-fitted curve among the other proposed distributions. It is very easy to see that the other two, i.e., Gamma and Lognormal distributions are not showing a best-fitted curve for the given data. Figure shows, if we will have the wrong selection, there will be more risk. Wrong selections will lead more chances of error occur. If we rank them from Figure 8, then there will be Weibull leads to Gamma followed by Lognormal distribution.

Figure 7  Histogram of hourly wind speed data
Figure 8  Fitness of theoretical distributions with observed data (see online version for colours)

The result showing in Figure 9 is the graphical representation of empirical and theoretical CDFs. Variation of theoretical CDFs of these three proposed distributions with empirical distribution function is showing fitness of probability distribution of given data. From this we can identify Weibull distribution is more fitted with empirical distribution function. On the basis of this, we can say Weibull distribution is best fitted one in a case of BHU area.

Figure 9  Empirical and theoretical CDFs (see online version for colours)
Figure 10 shows P–P (probability–probability) plot of the given data. In a P–P plot, this assesses how closely two datasets agree for the two CDFs against each other. Here in Figure 10, Weibull distribution is showing a close relation with the given dataset for the CDF compared to other two Gamma and Lognormal distributions. With the help of P–P plots, we can evaluate the skewness of the distribution. Figure 11 is showing the Q–Q (quantile–quantile) plot. In a Q–Q plot, this also assess how closely two datasets agree for the two CDFs against each other with their quantile plots. Q–Q plot also helps to compare the shapes of distributions with the graphical presentation to view how properties like location, scale and skewness are similar or different in the two distributions. Q–Q plots compared between collected data and theoretical distributions. Q–Q provides an assessment of ‘goodness of fit’ graphically, rather than having a numerical summary. With the interpretation of Figure 11, we can find the best-fitted Weibull distribution in this study compared to other two Gamma and Lognormal distributions.

Figure 10  P–P plot (see online version for colours)

Table 1 is showing the results of parameters for the proposed distributions. Mean and standard deviation of 1-year period hourly wind speed data of BHU campus is 3.53 and 1.87, respectively. These parameters have been calculated by the maximum likelihood methods because of its more accurate values.

With goodness-of-fit test results in Table 2, authors have ranked these three distributions as per the individual test results of Kolmogorov–Smirnov, Anderson–Darling and Cramer–von Mises test. In Table 2, we can see all these three tests are showing Weibull distribution is best one with less error value and ranked them in order of Weibull, Gamma and Lognormal distribution. From this, we can say that the Weibull distribution is the better one among the proposed three distributions.
Figure 11  Q–Q plot (see online version for colours)

Table 1  Parameter values of Weibull, Gamma and Lognormal distributions

<table>
<thead>
<tr>
<th>Weibull</th>
<th>Gamma</th>
<th>Lognormal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shape</td>
<td>Scale</td>
<td>Shape</td>
</tr>
<tr>
<td>1.874745</td>
<td>3.979186</td>
<td>2.8788796</td>
</tr>
</tbody>
</table>

4.1 Goodness-of-fit test results

Table 2  Goodness-of-fitness tests results of Kolmogorov–Smirnov, Anderson–Darling and Cramer–von Mises

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull</td>
<td>0.02106212</td>
<td>3.87999752</td>
<td>0.65250478</td>
</tr>
<tr>
<td>Gamma</td>
<td>0.02778126</td>
<td>7.76586366</td>
<td>1.22543889</td>
</tr>
<tr>
<td>Log normal</td>
<td>0.06997673</td>
<td>78.86732540</td>
<td>12.59881876</td>
</tr>
</tbody>
</table>

5 Theoretical power density

Estimation of the wind power depends on wind turbine blade rotational cross-sectional area, wind speed and wind flow kinetic energy. It can also be represented by multiplication of kinetic energy with total swept volume of the wind per unit time (Ackermann, 2005; Ramirez and Carta, 2005)

\[ P(v) = \left(0.5 \rho v^3\right)vA = 0.5 \rho v^3 A \]  \[ (21) \]

where \( \rho \) is the air density (kg m\(^{-3}\)).
Variation of air density depends on upon pressure, temperature, and humidity. For the first attempt of resource calculation variation of air density is not greatly significant. With this assumption on a constant pressure surface, the error is probably less than 5% (Hennessey Jr., 1977). In standard conditions at sea level with 15°C temperature, a typical value of air density is $\bar{\rho} = 1.225 \text{ kg m}^{-3}$.

Statistical analysis of wind power potential with mean air density $\bar{\rho}$ is proportional to the cube of the random variable wind speed values. We can calculate the distribution of wind power with random wind speed values by multiplying the wind power density of each wind speed with the probability of each wind speed. So, this can be written in mathematical term

$$P(v) = 0.5\bar{\rho}v^3 f(v, \alpha, \beta) \quad [\text{Wm}^{-3}\text{s}]$$

(22)

With this, we can calculate the annual mean wind power potential from the wind resources available at a potential site. This can be determined through

$$\bar{P} = 0.5\bar{\rho}\int_0^\infty v^3 f(v, \alpha, \beta) \, dv = 0.5\bar{\rho}\mu' \quad [\text{Wm}^{-2}]$$

(23)

where $\mu'$ represents the mean of the cube of the wind speed ($\text{m}^3\text{s}^{-3}$)

$$\mu' = \beta^3\Gamma\left(1 + \frac{3}{\alpha}\right)$$

(24)

An annual power production from the wind turbines is a maximum wind power density for the aerodynamically efficient design at $v = v_{\text{me}}$

$$v_{\text{me}} = \beta\left(1 + \frac{2}{\alpha}\right)^{\frac{1}{3}} \quad [\text{ms}^{-1}]$$

(25)

### 6 Wind energy potential

Potential of wind energy can be estimated with randomly distributed wind speed by multiplication of power output from the hourly wind speed dataset and their probability distribution (Kwon, 2010),

$$\frac{E_{\text{annual}}}{A} = \frac{T}{A}\int_0^\infty P(v) f(v) \, dv$$

(26)

where $T$ denotes the period of hours and $A$ equals to the total area swept by wind turbine blades. In wind energy estimation, we neglect the value of air density because it does not affect the wind speed values. In unavailability of wind speed situation, we can exclude that period and can identify total number of hours for availability of wind speed in a year,

$$T = 365 \times 24 \times \{1 - f(v = 0)\}$$

(27)

Here $f(v = 0)$ means the probability of hours for unavailability of wind speed. We can calculate annual wind energy potential for the $T$ hours without having any probability distribution from this equation,
Here Figure 12 is showing mean monthly wind power density distribution for BHU area. We can see maximum potential of wind power for the April month and minimum in January month. This will help in optimal capacity installation of renewable distributed generation unit at BHU campus and have better demand–supply management.

**Figure 12** Monthly average wind power potential at BHU campus

7 Conclusion

A decision on making an investment in the renewable distributed generation technology planned to be built at BHU region will be made by the help of wind speed data measured from the region. Therefore, distribution of the wind speed at BHU region is statistically analysed for 12 months in between 2002 and 2011 on an hourly basis. Because of randomness of the wind speed values, it is expected to fit a probability density function which will give better fitness. For that, three probability density functions have been proposed in this analysis which is Weibull, Gamma and Lognormal distributions. Significance level alone cannot help to make a decision for selection of any distribution without going for any statistical goodness-of-fit test and graphical analysis. For making scientific decisions, the reliability of the decision is very important factor in estimation and investment projects to be made by the help of this decision. Since decisions made with the help of statistical analysis are in certain confidence level, in this case, there was 95% confidence level. In this study, authors used three goodness-of-fit tests named Kolmogorov–Smirnov, Anderson–Darling and Cramer–von Mises along with the graphical analysis to identify the fittest distribution. At the end of the comparison, Weibull has been determined to be the best-fitted distribution representing given wind speed data of the BHU region. Best-fitted distribution will reduce chances of error and risk in making an investment decision for setup of renewable distributed generation units.
References


Websites


Nomenclature

### Abbreviations

- **BHU**: Banaras Hindu University
- **CDF**: Cumulative density function
- **CGI**: Computer generated imagery
- **CSV**: Comma separate values
- **DG**: Distributed generation
- **ECDF**: Empirical cumulative density function
- **ECMWF**: European centre for medium-range weather forecasts
- **GIS**: Geographic information system
- **GMT**: Greenwich mean time
- **GoF**: Goodness-of-fit
- **LBNL**: Lawrence Berkley National Laboratory
- **LIDAR**: Light detection and ranging
- **LST**: Local Standard Time
- **MNRE**: Ministry of New and Renewable Energy
- **NCAR**: National Centre for Atmospheric Research
- **NCEP**: National Centres for Environmental Prediction
- **NREL**: National Renewable Energy Laboratory
- **pdf**: Probability density function
- **P–P**: Probability–Probability
- **Q–Q**: Quantile–Quantile
- **RDG**: Renewable distributed generation
- **RES**: Renewable energy sources
- **SAR**: Synthetic aperture radar
- **SODAR**: Sonic detection and ranging

### Symbols

- $u$: Wind speed (m/s) at height $h$
- $u_r$: Wind speed (m/s) at height $h_r$
- $h$: Desired level height (m)
- $h_r$: Reference level height (m)
- $p$: Empirically derived coefficient
- $c$: Scale parameter of Weibull distribution
- $k$: Shape parameter of Weibull distribution
### Wind energy potential estimation with prediction of wind speed distribution

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F$</td>
<td>Theoretical cumulative distribution function</td>
</tr>
<tr>
<td>$F_n$</td>
<td>Empirical distribution function</td>
</tr>
<tr>
<td>$T$</td>
<td>Period of hours</td>
</tr>
<tr>
<td>$P$</td>
<td>Wind power (W)</td>
</tr>
<tr>
<td>$v$</td>
<td>Wind velocity (m/s)</td>
</tr>
<tr>
<td>$A$</td>
<td>Swept area of wind turbine blades (m$^2$)</td>
</tr>
<tr>
<td>$E$</td>
<td>Annual wind energy potential (Joule)</td>
</tr>
</tbody>
</table>

**Greek letters**
- $\Gamma$: Gamma function
- $\alpha$: Shape parameter of gamma distribution
- $\beta$: Rate parameter of gamma distribution
- $\mu$: Location parameter of lognormal distribution
- $\sigma$: Scale parameter of lognormal distribution
- $\rho$: Air density (kg m$^{-3}$)
- $\mu'$: Mean of the cube of the wind speed (m$^3$s$^{-3}$)