Effect of reducing the accuracy of wind speeds on distribution functions: A case study

Ali Mostafaeipour
Industrial Engineering Department
Yazd University
Yazd, Iran
mostafaei@yazd.ac.ir

Omid Alavi
Department of Electrical Engineering,
K.N. Toosi University of Technology
Tehran, Iran
omidal1370@gmail.com

Abstract

Accuracy of wind speed data has important impact on determining wind power output from a wind turbine. There are many researches on four widely used wind speed distribution models described by gamma, lognormal, Rayleigh and Weibull for assessing wind potentials. However, there is lack of studies to evaluate sensitivity of these models with respect to accuracy of the measured wind data. In this paper, wind speed data are measured by national data buoy center (NDBC) over ten years, from 2004 to 2014, for four offshore stations in the east of the U.S. Two methods of maximum-likelihood estimator (MLE) and method of moments (MOM) are utilized for calculating parameters involved with these four distribution functions. For reducing the accuracy, a truncated set of wind data is generated by removing the decimal digits of the wind data; reducing the resolution to 1 m/s. Also, the best distribution functions in terms of performance are selected by examining nine goodness-of-fit statistics. From the outcomes, it is concluded that the Weibull function offers a better fit to the both actual and truncated data. Additionally, the Rayleigh distribution function exhibits suitable fit with the truncated wind speed data.

Keywords
Wind speed; Weibull distribution; Anemometer accuracy; Probability density functions; Massachusetts.

1. Introduction

Replacing renewable energy with fossil fuel has become an important issue over the recent decades. The fossil fuel consumption has resulted in, global warming, environmental pollution, and many other crucial problems in the world [1]. Renewable energy development is a main measure in climate change mitigation and greenhouse gas emission. The issues related to reduction of the anthropogenic impact on climate change, and increasing the implementation of renewable energy sources, have been the most debated topics in the world [2]. Most of the total energy consumption in the world is based on fossil fuels, which has numerous negative impacts on the environment. Recently, the production and consumption of these fuels are increasing as they are necessary for the maintenance of the global economy [3-5]. Also, the vast majority of the world’s energy consumption is related to the heating and cooling of the residential and commercial sector [6]. The renewable energy policy is a main purpose of sustainability, along-side energy efficiency and sufficiency [7]. Unknown amounts of conventional energy reserves are buried deep in the ground or under the ocean. It is extremely difficult to identify and exploit new sources due to very high cost and dangerous conditions while drilling under the ocean. Also, large amounts of natural gas would burn in order to refine when it comes to oil sands [8]. Therefore, renewable energy resources must be developed to prevent negative effects on the environment.
Wind and solar are the most important sources of renewable energies which are abundant in many countries. Wind is an important renewable energy source because of many advantages, such as low cost, clean, abundant, inexhaustible and environmental benefits. Wind turbine technology has increased over the last few decades in many countries. Many governments have decided to enhance knowledge wind turbine technologies for electricity generation [9]. Among the methods of electricity generation from renewable energies, wind turbines are believed to be the best option and are the most cost effective. Measuring an accurate value of solar radiation or wind speed can lead to a precise designing or planning of solar and wind power plants [10].

There are different studies on the assessment of distribution functions for modeling wind speeds. In the meanwhile, the Weibull function is the most widely used method in order to provide a suitable model for a wind speed dataset [11-15]. Shamshirband et al. [16] investigated the application of extreme learning machine for estimation of wind speed distribution. They conducted an analysis for computing two parameters of Weibull function using extreme learning machine (ELM). The results showed that the employing of ELM will result in more accuracy presentation of these parameters. Alavi et al. [17] evaluated the sensitivity analysis of four different distribution functions for five locations in Iran. They found that the Weibull distribution function exhibits greater flexibility in comparison with other ones. Only a few studies have been conducted on the effect of using various types of data on distribution functions, and almost all of them were related to the Weibull function and data with different sampling intervals [18, 19]. Thus, there is no comparative research on the effectiveness of data with different accuracies on the well-known distribution functions particularly in the field of wind speed modeling.

One mandatory device to be employed in meteorological measurements is anemometer. The available data from anemometers can be used in feasibility studies of wind energy. Each anemometer has prescribed accuracy and time intervals that can collect wind speeds. Generally, the range of measuring wind speed for the anemometers varies between seconds and hours. There is a lack of research on the importance of wind data accuracy in the performance and sensitivity of wind speed distributions models as the mandatory application tools on wind resource assessments. The main objective of the present article is to carry out an evaluation on sensitivity of four wind speed distributions functions to accuracy of anemometer measured wind data at four locations. A detailed comparison of the introduced distribution functions is undertaken to illustrate their aptness for describing wind speed characteristics. There have been numerous works regarding wind energy in the past[20-27].

The remainder of the paper is organized as follows: In Section 2, geographical location and wind characteristics of four offshore stations in the eastern part of the U.S. are explained. Section 3 presents four commonly used distribution functions to describe wind speed characteristics using the two parameter estimation methods of the maximum likelihood estimator and the method of moments. In Section 4, to investigate the sensitivity of the distribution functions using actual and truncated wind data, their statistical indicators are compared. Section 5 concludes the paper.

2. Case studies and wind speed data

The United States as a vast country leads the way in the energy industry. The USA despite the presence of countries like China and Germany in renewable energy development has achieved the first place in this field [28]. This country has plans to allocate more financial resources to expand the use of renewable energies. The U.S. state of Massachusetts is situated in the eastern part of the country, and has a remarkable onshore wind power potential as well as offshore. In this study, four different locations are considered to demonstrate the suitability of each distribution function for modeling wind speeds. The wind data were recorded by an anemometer installed on a buoy at the height of 5 m above the sea level. The geographical location of the examined locations can be shown in Fig. 1.
Additionally, Table 1 provides the wind characteristics and descriptive statistics, such as standard deviation, maximum, median, mean, skewness, and kurtosis for each location.

Table 1. Statistical characteristics of wind data for the selected locations

<table>
<thead>
<tr>
<th>No.</th>
<th>Station</th>
<th>Mean (m/s)</th>
<th>Max (m/s)</th>
<th>Standard deviation</th>
<th>Median (m/s)</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>#44005</td>
<td>5.53</td>
<td>22.8</td>
<td>3.19</td>
<td>5.1</td>
<td>0.65</td>
<td>3.08</td>
</tr>
<tr>
<td>2</td>
<td>#44007</td>
<td>6.31</td>
<td>23.7</td>
<td>3.38</td>
<td>5.9</td>
<td>0.63</td>
<td>3.17</td>
</tr>
<tr>
<td>3</td>
<td>#44013</td>
<td>6.30</td>
<td>22.7</td>
<td>3.35</td>
<td>5.9</td>
<td>0.66</td>
<td>3.29</td>
</tr>
<tr>
<td>4</td>
<td>#44018</td>
<td>6.64</td>
<td>23.0</td>
<td>3.28</td>
<td>6.3</td>
<td>0.55</td>
<td>3.19</td>
</tr>
</tbody>
</table>

Wind speed data at the height of 5 meters for the sites under consideration are obtained from national data buoy center (NDBC) over ten years, from 2004 to 2014 [29].

### 3. Wind speed distribution models

To evaluate the suitability of wind speed in a location for wind power generation, the wind speed is investigated by fitting on a distribution model. Generally, wind speeds over several years are assessed by a long-term analysis. In order to model wind speeds, various distribution functions are utilized that introduced in the literature. In this section, four commonly used distribution functions of gamma, lognormal, Rayleigh, and Weibull are proposed. Two methods of maximum likelihood estimator (MLE) and method of moments (MOM) are described along with each distribution function.

#### 3.1 Weibull distribution

The Weibull distribution is known as a strong distribution function which is widely used in different sectors, such as wind resource assessments to offer an appropriate wind speed probability distribution. The probability density function (PDF) and cumulative distribution function (CDF) of Weibull distribution can be expressed, respectively, by [30]:

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

\[
F(v, k, c) = 1 - \exp\left[-\left(\frac{v}{c}\right)^k\right] \tag{2}
\]

where \(k\), \(c\) and \(v\) are shape parameter, scale parameter and wind speed value, respectively. Weibull shape and scale parameters can be calculated using (3) and (4) based on the maximum likelihood method (MLE), respectively as [30]:

\[
\frac{1}{c} = \left(\frac{k}{\bar{v}}\right)^{1/k} \tag{3}
\]

\[
\frac{1}{c} = \frac{1}{\Gamma(1 + 1/k)} \int_0^\infty \frac{1}{x^k} \bar{v}^k \exp\left[-\left(\frac{\bar{v}}{c}\right)^k\right] dx \tag{4}
\]
\[ k = \left( \frac{\sum_{i=1}^{n} v_i \ln(v_i)}{\sum_{i=1}^{n} v_i} - \frac{\sum_{i=1}^{n} \ln(v_i)}{n} \right)^{-1} \]  
\[ c = \frac{\sum_{i=1}^{n} v_i^k}{n} \]

where \(v_i\) is the wind speed in \(i\)-th time step (m/s) and \(n\) is the number of nonzero wind speeds.

The estimation of Weibull parameters based on the method of moments (MOM) can be performed by solving these equations [31]:

\[ k = \left( \frac{0.9874 v^{-0.0983}}{\sqrt{S^2}} \right)^{1.0983} \]  
\[ c = \frac{v}{\Gamma(1+\frac{1}{k})} \]

where \(\bar{v}\), \(S^2\) and \(\Gamma\) are the mean wind speed, the variance of wind speed and the gamma function, respectively. The variance of wind speed can be found as follows:

\[ S^2(v) = \frac{\sum(v-v)^2}{n-1} \]  

where \(n\) is the number of samples.

In addition, the gamma function is provided as follows:

\[ \Gamma(t) = \int_{0}^{\infty} x^{t-1} e^{-x} dx \]

### 3.2 Gamma distribution

The PDF of gamma for wind speed, \(v\), with two parameters can be given by [32]:

\[ f(v, \alpha, \beta) = \frac{v^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} \exp\left(-\frac{v}{\beta}\right) \]

where \(\alpha\), \(\beta\) and \(\Gamma\) are the shape parameter, scale parameter and the gamma function, respectively. Also, the CDF of gamma distribution function can be expressed by [33]:

\[ F(v, \alpha, \beta) = \int_{0}^{\frac{v}{\beta}} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) dx \]

According to the MLE (maximum likelihood estimator) method, the shape and scale parameters can be obtained by solving simultaneously the following equations [33]:

\[ \alpha \beta = \bar{v} \]  
\[ n \ln(\beta) + n \psi(\alpha) = \sum_{i=1}^{n} \ln(v_i) \]

where \(n\) is number of samples, and \(\psi\) is the digamma function. Here, \(v_i\) is the wind speed in \(i\)-th time step (m/s), and \(\bar{v}\) is the mean wind speed. The digamma function can be determined as follows:

\[ \psi(\alpha) = \frac{d}{d\alpha} \ln(\Gamma(\alpha)) \]

The parameters of the gamma distribution when MOM is using can be given as follows [34]:

© IEOM Society International

1232
\[ \alpha = \left( \frac{\bar{v}}{S} \right)^2 \]  
\[ \beta = \frac{s^2}{\bar{v}} \]  

where \( \bar{v} \) and \( S \) are the mean wind speed and the standard deviation, respectively.

### 3.3 Lognormal distribution

The PDF of lognormal distribution can be presented by [35]:

\[ f(v, \mu, \sigma) = \frac{1}{v \sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left( \frac{\ln(v) - \mu}{\sigma} \right)^2 \right\} \]  

where \( \sigma \) and \( \mu \) are the shape and the scale parameters, respectively. The CDF of lognormal is given by [35]:

\[ F(v, \mu, \sigma) = \frac{1}{2} + \frac{1}{2} \text{erf} \left[ \frac{\ln(v) - \mu}{\sigma \sqrt{2}} \right] \]  

where the error function \( \text{erf} \) is defined by:

\[ \text{erf}(v) = \frac{2}{\sqrt{\pi}} \int_0^v \exp(-t^2) \, dt \]  

The parameters of lognormal distribution based on the MLE method are calculated by the following equations [35]:

\[ \mu = \frac{1}{N} \sum_{i=1}^{N} \ln(v_i) \]  
\[ \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [\ln(v_i) - \mu]^2} \]  

The MOM parameter estimation of the lognormal distribution can be performed by solving the following two equations [36]:

\[ \mu = -\frac{\ln(\sum_{i=1}^{n} v_i^2)}{2} + 2 \ln(\sum_{i=1}^{n} v_i) - \frac{3}{2} \ln(n) \]  
\[ \sigma = \sqrt{\ln(\sum_{i=1}^{n} v_i^2) - 2 \ln(\sum_{i=1}^{n} v_i) + \ln(n)} \]  

where \( n \) and \( v_i \) are the number of total data and the \( i \)-th value of wind speed data, respectively.

### 3.4 Rayleigh distribution

The Rayleigh distribution is the simplest distribution function and is extensively used to describe average wind speeds. The probability density function and cumulative distribution function of the Rayleigh can be written as [37]:

\[ f(v, a) = \frac{v}{a^2} \exp \left( -\frac{v^2}{2a^2} \right) \]  
\[ F(v, a) = 1 - \exp \left[ -\frac{1}{2} \left( \frac{v}{a} \right)^2 \right] \]  

where \( a \) is the single scale parameter.

By considering the MLE method, the scale parameter can be calculated by:

\[ a = \sqrt{\frac{1}{2n} \sum_{i=1}^{n} v_i^2} \]  

The value of the scale parameter based on the MOM is related directly to the mean wind speed, and is given by [37]:
\[ a = \sqrt[2\pi]{\hat{\sigma}} \] (26)

4. Results and Discussion

As mentioned before, the effectiveness of various factors on the performance of wind speed distributions is significant. However, there is a lack of study on the sensitivity of the commonly used functions for accurate modeling of data set. To address this, the wind-speed data for four offshore locations in the east of U.S. are examined as case studies for the present analyses.

In order to prioritize the introduced distribution functions, it seems a need to take into account some statistical indicators in the comparisons. In this paper, nine different forms of goodness of fit indices are used to choose the best and the most capable distribution function for each location. These indicators include root mean squared error (RMSE), the mean squared error (MSE), the normalized mean squared error (NMSE), the normalized root mean squared error (NRMSE), the mean absolute error (MAE), the mean absolute relative error (MARE), the coefficient of correlation (R2), the coefficient of determination (D), and the coefficient of efficiency (E) [38]. Whatever, the six first indicators are lower and the three last indicators are higher, is better and exhibits a proper fit on the empirical data.

In this study, we intend to compare four distribution functions in terms of the performance in the presence of inaccurate (or truncated) data. Additionally, it should be determined that by using inaccurate data, how can we found the same results as the precise data. For reducing the accuracy of the measured wind speeds, we can remove the decimal part of wind speeds or round these values. When the inaccurate data are utilized in the analysis, their outcomes should be compared by the accurate wind speed data.

Table 2 lists the obtained parameters from two analyses of Actual and Truncated wind speeds.

Table 2. Parameters of four distribution functions for the selected stations

<table>
<thead>
<tr>
<th>Station</th>
<th>Distribution function</th>
<th>#44005</th>
<th>#44007</th>
<th>#44013</th>
<th>#44018</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Actual</td>
<td>Truncated</td>
<td>Actual</td>
<td>Truncated</td>
<td>Actual</td>
</tr>
<tr>
<td>Gamma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOM</td>
<td>1.823</td>
<td>1.809</td>
<td>1.795</td>
<td>1.825</td>
<td>1.771</td>
</tr>
<tr>
<td>Rayleigh</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLE</td>
<td>4.427</td>
<td>4.219</td>
<td>5.047</td>
<td>4.783</td>
<td>5.036</td>
</tr>
<tr>
<td>MOM</td>
<td>4.522</td>
<td>4.331</td>
<td>5.067</td>
<td>4.841</td>
<td>5.050</td>
</tr>
<tr>
<td>Lognormal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLE</td>
<td>1.371</td>
<td>1.518</td>
<td>1.720</td>
<td>1.658</td>
<td>1.719</td>
</tr>
<tr>
<td>MOM</td>
<td>0.533</td>
<td>0.542</td>
<td>0.500</td>
<td>0.516</td>
<td>0.497</td>
</tr>
<tr>
<td>Weibull</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLE</td>
<td>1.831</td>
<td>1.791</td>
<td>1.982</td>
<td>1.908</td>
<td>1.994</td>
</tr>
<tr>
<td>MOM</td>
<td>6.244</td>
<td>5.944</td>
<td>7.130</td>
<td>6.757</td>
<td>7.121</td>
</tr>
</tbody>
</table>

According to Table 2, it is obviously clear that the shape parameter of the gamma distribution based on MOM method increases when the truncated data are used. This is despite the fact that the gamma scale parameter has fluctuations with actual and truncated data. Also, the gamma scale parameter using MLE method decreases with the truncated wind data. The Rayleigh parameter will always have a lower value by replacing the inaccurate data instead of the actual measured data. Unlike the shape parameter of the lognormal function, the scale parameter will go down with the inaccurate data. For the Weibull distribution function, we can see that two parameters of shape and scale follow an upward trend.

By performing an analysis, a rank was assigned to each distribution function based on the nine statistical indicators. The best performance for all locations with the actual data is related to the Weibull base on the MLE method. However, the most proper function for presenting the wind data when the inaccurate data are used, is related to the Weibull (MLE), Rayleigh (MLE), and Weibull (MLE). It is evident that the Weibull distribution function based on the MLE estimation method is the best option for both actual and truncated wind speeds. Also, it can be concluded that the MOM method is not suitable for all locations and the both types of data. For this reason, the presentation of this method has been refused. Fig. 2 (a-d) illustrates the fitted PDFs for the four nominated locations.
5. Conclusions

A precise knowledge of the wind speed distribution model based upon the wind regime at any wind farm is crucial to select ideal wind turbines and increase energy efficiency. In the meantime, determining the influence of different factors involved in choosing wind distributions is a notable exercise. In the present study, four widely used distribution functions of gamma, Rayleigh, lognormal, and Weibull, using two parameter estimation methods of MOM and MLE, are evaluated. The analysis was carried out with two types of wind speed data, the actual and truncated wind speed data. The actual wind data exhibit an accuracy of ±0.1 m/s; however, the truncated wind data have the accuracy of ±1 m/s. The results from analyzing wind data for the four studied stations showed that the Weibull distribution function with MLE method is definitely a better option to be used in wind speed analysis with actual data. Additionally, it is evident that the Rayleigh function can be another suitable function for modeling wind speeds. The suggested wind speed distribution functions can be equally applied for any station in the U.S. or other countries.
References


© IEOM Society International

1236


Biography

Ali Mostafaeipour is an assistant professor of Industrial Engineering at Yazd University, Iran. He has been teaching at Yazd University since 1989. He studied at Winona State University (University of Minnesota) in state of Minnesota, USA; University of Wisconsin at Platteville, Wisconsin, USA; Alabama A&M, Alabama, USA; and Iran University of Science and Technology, Tehran, Iran. He has served as a committee member, guest speaker, and co-chairman of 117 international conferences. He has been reviewer of 17 international journals mainly Elsevier. He has presented 73 mostly International conferences throughout the world. He has undertaken and managed 18 research projects, and holds 3 patents. He has been editorial board of several professional journals. Finally, he has published 51 journal articles mostly at Elsevier (ISI), and he authored 4 books. He holds an award for excellence from Yazd University as the year 2013 distinguished researcher, also distinguished author of “Wind Energy” book (INTech publisher, 2012, Croatia) with more than 5000 downloads in six months. His research interest lies in renewable energies, wind energy, value engineering, economic evaluation, and feasibility study of projects.

Omid Alavi was born in Zahedan, Iran, in 1991. He received the B.S. degree in electrical engineering from Ferdowsi University of Mashhad, Iran, in 2014. He is currently with the department of electrical engineering, K.N. Toosi
University of Technology, Tehran, Iran. His research interests include renewable energies, wind turbine design, power electronic devices, reliability, and resource assessments. He published several international papers in highly ranked journals. He is also the reviewer of some journals, such as “Electric Power Systems Research”, “Energy” and “Energy, Sustainability and Society”.