Wind turbine power curve modeling using maximum likelihood estimation method

Seokho Seo a, b, Si-Doek Oh a, Ho-Young Kwak a, c, *

a Blue Economy Strategy Institute Co. Ltd., Focus Buld. 23-10, Hyoryeong-ro, 60-gil, Seocho-gu, Seoul, 06721, South Korea
b Dept. of Climate Change Energy Eng., Yonsei University, Seoul 03722, South Korea
c Mechanical Engineering Department, Chung-Ang University, Seoul, 06974, South Korea

Article info
Article history:
Received 27 June 2017
Received in revised form 27 August 2018
Accepted 24 September 2018
Available online 27 September 2018

Keywords:
Logistic function
Wind turbine power curve
Weibull distribution
Maximum likelihood estimation method

A B S T R A C T
Modeling of wind turbine power curve which shows the relationship between wind speed and its power output can be used as an important tool in monitoring and forecasting wind energy. A data-driven approach to find most probable probability distribution function (PDF) for wind speed and turbine power is presented in this study. Equations for the scale and shape parameters in the Weibull wind speed distribution and equations for the four parameters in the logistic function were obtained explicitly by maximum likelihood estimation (MLE) method. With help of a selected data set from the wind speed and the corresponding power output data which was collected over a period of a year, the values of the parameters were obtained by solving the equations by iteration procedures. The predicted powers by the obtained logistic function closely follow the measured turbine powers averaged at 5-min or 10-min. Monitoring turbine power output by the logistic function was also tested for the measured powers in other time duration.

© 2018 Elsevier Ltd. All rights reserved.

1. Introduction
The power generated by wind turbines is characterized by its power curve which shows the relationship between wind speed and its power output. Accurate model of power curve for wind turbines can serve as an important tool in forecasting and predicting wind energy, which allows for better grid planning and integration of wind energy into existing power systems [1,2]. The power curve model can also serve as an effective monitoring tool, which enables to detect anomalous conditions during the operation of wind turbine [1].

The power curve of a wind turbine can be estimated using the power coefficient (Cp) from the turbine blade parameters such as tip speed ratio and pitch angle, rotor dimensions and reference air density [1,3]. Shortcoming of this model is that it depends on many technical factors of wind turbines [4]. Another way to obtain the power curve can be done by means of fitting techniques [5]. However, the resulting power curve equations from these techniques are quite complex [6], which cannot be expressed by a generic expression.

The wind farm power curve may be approximated by a logistic function with four parameters [7–9]. The nonlinear parametric power curve can be adjusted to the operational conditions of a specific wind turbine installed at a site having different wind characteristics by modifying its parameters. The parameters of the logistic function have been obtained using the least square method (LSM) [7,8]. Evolutionary strategy (ES) algorithm was applied to minimize the cost function in LSM and maximum likelihood function (MLF) [8], which is a challenging problem because the logistic function is nonlinear and it contains four variables.

This study is an outcome of a project on the estimation of annual electricity produced from wind turbines which will be installed at proposed sites in Jeju Island, Korea by using modified MERRA (NASA) wind field data. In this study, maximum likelihood estimation (MLE) method [10,11] was used estimate the parameters in the logistic function. Equations for the scale and shape parameters in the Weibull wind speed distribution and equations for the four parameters in the logistic function were obtained explicitly by MLE method. Using iteration procedures, the values of the parameters were obtained using a selected data set from the wind speed and the corresponding power output data which was collected over a period of a year from wind turbines installed at specific sites in Jeju.
Island, Korea. The predicted powers by the logistic function with the parameter values obtained by MLE method closely follow the measured powers averaged at 5-min or 10-min.

2. Parametric model of power curve

2.1. Wind speed distribution

In this study, a data-driven approach to find the parameters in the logistic function to estimate the wind turbine power is presented. A measured data of wind speed and the corresponding power from a 800-kW and a 2-MW wind turbine are used. Though the data was sampled at a high frequency of 0.5 Hz, the data was averaged and stored at 5-min (800-kW) or at 10-min (2-MW) intervals. Let the selected N pairs of data points for the wind speed and the corresponding power output be \([u_i, P_i]\) where \(i = 1, 2, \ldots, N\). The data was selected in order for time duration where outliers are not included.

Assume that the wind speed in the data set follows the Weibull distribution with two parameters, the probability density function (PDF) for the wind speed, \(u\), is expressed as [6].

\[
f(u/\beta, \eta) = \frac{\beta}{\eta} \left(\frac{u}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{u}{\eta}\right)^\beta\right]
\]

where \(\beta\) is shape parameter and \(\eta\) is scale parameter, which are both positive quantities. Furthermore, assume that the average wind speeds in the data set are statistically independent of one another, the PDF for the selected wind speed of data set, can be expressed as a multiplication of PDF’s individual wind speed [11].

\[
f(u_1, u_2, \ldots, u_N/\beta, \eta) = f(u_1/\beta, \eta) \cdot f(u_2/\beta, \eta) \cdots f(u_N/\beta, \eta)
\]

where \(u_i\) is the \(i\)th averaged wind speed in the data set.

The most probable PDF for the wind speed data set can be found by estimating the parameter value of \(\beta\) and \(\eta\) by the maximum likelihood function (MLF), \(L(u)\). For the selected wind speed data set, MLF, \(L(u)\) is defined based on Eq. (2) as

\[
L(u/\beta, \eta) = \prod_{i=1}^{N} f(u_i/\beta, \eta) = \prod_{i=1}^{N} \frac{\beta}{\eta} \left(\frac{u_i}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{u_i}{\eta}\right)^\beta\right]
\]

MLE estimates are obtained by maximizing MLF. Since \(L(u)\) and \(\ln[L(u)]\) are monotonically related to each other so the same MLE estimates are obtained by maximizing \(\ln[L(u)]\). The \(\ln[L(u)]\) is written as

\[
\ln L(u/\beta, \eta) = N \ln(\beta/\eta) + \sum_{i=1}^{N} \ln \left(\frac{u_i}{\eta}\right)^{\beta-1} - \sum_{i=1}^{N} \left(\frac{u_i}{\eta}\right)^\beta
\]

Differentiating Eq. (4) with respect to \(\beta\) and \(\eta\), we obtain MLE estimates for the most probable PDF for the wind speed such as

\[
\eta = \left(\frac{1}{N} \sum_{i=1}^{N} \left(\frac{u_i}{\eta}\right)^\beta\right)^{1/\beta}
\]

\[
1/\beta = \left[\frac{\sum_{i=1}^{N} \left(\frac{u_i}{\eta}\right)^\beta \ln(u_i)}{\sum_{i=1}^{N} \left(\frac{u_i}{\eta}\right)^\beta} - \frac{1}{N} \sum_{i=1}^{N} \ln(u_i)\right]
\]

An optimal parameter value of \(\beta\) can be estimated through Eq. (6) by an iteration procedure. With known value of \(\beta\), one can obtain an optimal parameter value of \(\eta\) using Eq. (5). In Eqs. (5) and (6), the probability with wind speed of \(u_i\) is \(1/N\). With estimated value of \(\beta\) and \(\eta\) for the most probable PDF, the average and the dispersion for the wind speed can be obtained by the following relations.

\[
\bar{u} = \eta \cdot (1 + 1/\beta)
\]

\[
S = \eta^2 \cdot \left[(1 + 2/\beta) - (1 + 1/\beta)^2\right]
\]

where \(\Gamma\) in Eqs. (7) and (8) is gamma function.

2.2. Wind power distribution

Assume that the logistic function given in Eq. (9) provides reasonably the power from wind turbines [7,8].

\[
P(u/\theta) = \frac{1 + m \cdot \exp(-u/\tau)}{1 + n \cdot \exp(-u/\tau)}
\]

where \(P\) is the power generated by the wind turbine, and \(\theta = (a, m, n, \tau)\) is vector parameter for the logistic function that determines its shape. The PDF of the turbine power, \(f_P\) can be derived from the most probable PDF for the wind speed with help of the defined logistic function using the statistical inference theorem [10].

\[
f_P = f_u \left(g^{-1}(P)\right) \cdot \frac{d}{dP} \left(g^{-1}(P)\right)
\]

where

\[
g^{-1}(P) = u = -\tau \cdot \ln\left(\frac{a - P}{nP - am}\right)
\]

Explicit expression for the PDF of the turbine power calculated based on Eq. (10) is given by

\[
f_P(P/\theta) = \frac{a - P}{nP - am} \cdot \exp\left\{-\frac{\beta \tau a (n - m)}{\eta (a - P)(nP - am)} \cdot \left[-\tau \cdot \ln\left(\frac{a - P}{nP - am}\right)\right]^{\beta-1}\right\}
\]

A vector parameter, \(\theta = (a, m, n, \tau)\) in the logistic function needs to be estimated from the N pairs of data points \([u_i, P_i]\), and \(i = 1, 2, \ldots, N\). To estimate the parameters, the following MLF from the power distribution data set is defined based on Eq. (12).

\[
L(P/\theta) = \prod_{i=1}^{N} \frac{a - P_i}{nP_i - am} \cdot \exp\left\{-\frac{\beta \tau a (n - m)}{\eta (a - P_i)(nP_i - am)} \cdot \left[-\tau \cdot \ln\left(\frac{a - P_i}{nP_i - am}\right)\right]^{\beta-1}\right\}
\]

where \(P_i\) is the \(i\)th measured power value among the N data set.

The logarithm of the expression, Eq. (13) becomes
\[
\ln L(P/\theta) = \sum_{i=1}^{N} \ln \left[ \frac{\beta r a (n - m)}{(\eta(a - P_i)/(nP_i - am))} \right] + (\beta - 1) \sum_{i=1}^{N} \ln \left[ -\frac{\tau}{\eta} \ln \left( \frac{a - P_i}{nP_i - am} \right) \right]
- \sum_{i=1}^{N} \left[ -\frac{\tau}{\eta} \ln \left( \frac{a - P_i}{nP_i - am} \right) \right]^\beta
\]

(14)

The optimum parameters for the logistic function can be obtained by differentiating Eq. (14) with respect to \(\tau\), \(a\), \(m\) and \(n\). The explicit equations for the estimates obtained are given as

\[
\left( \frac{\tau}{\eta} \right)^\beta = \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[ -\ln \left( \frac{a - P_i}{nP_i - am} \right) \right] \right\}^{-1}
\]

\[\eta = \left\{ \frac{1}{N} \sum_{i=1}^{N} \left[ -\ln \left( \frac{a - P_i}{nP_i - am} \right) \right] \right\}^{-1}
\]

(15)

\[
N \sum_{i=1}^{N} \frac{(a - P_i)/(nP_i - am)}{(a - P_i)/(nP_i - am) + (\beta - 1) \sum_{i=1}^{N} P_i/(n - m)/(nP_i - am) \ln \left( \frac{a - P_i}{nP_i - am} \right)} - \beta \sum_{i=1}^{N} \left( \frac{\tau}{\eta} \right) = 0
\]

(16)

\[
N \sum_{i=1}^{N} \frac{(n - m)/(nP_i - am)}{(n - m)/(nP_i - am) + (\beta - 1) \sum_{i=1}^{N} P_i/(nP_i - am) \ln \left( \frac{a - P_i}{nP_i - am} \right)} - \beta \sum_{i=1}^{N} \left( \frac{\tau}{\eta} \right) = 0
\]

(17)

\[
N \sum_{i=1}^{N} \frac{m/(nP_i - am)}{(n - m)/(nP_i - am) + (\beta - 1) \sum_{i=1}^{N} P_i/(nP_i - am) \ln \left( \frac{a - P_i}{nP_i - am} \right)} + \beta \sum_{i=1}^{N} \left( \frac{\tau}{\eta} \right)^\beta \frac{P_i/(nP_i - am) \ln \left( \frac{a - P_i}{nP_i - am} \right)}{(nP_i - am) \ln \left( \frac{a - P_i}{nP_i - am} \right)} = 0
\]

(18)

The performance of the obtained logistic function has been evaluated using the following performance metrics \(E_{abs}\) and \(E_{rms}\), which indicate the error between measurement and prediction in wind powers.

\[
E_{abs} = \frac{1}{N} \sum_{i=1}^{N} \left| P_{i,mes}(u_i) - P_{i,pred}(u_i) \right|
\]

(20)

\[
E_{rms} = \left( \frac{1}{N} \sum_{i=1}^{N} \left( P_{i,mes}(u_i) - P_{i,pred}(u_i) \right)^2 \right)^{1/2}
\]

(21)

The subscripts, \(mes\) and \(pred\) denote measured and predicted turbine power, respectively.

\(E_{abs}\) is the mean of the absolute value of the difference between the measured and predicted values of the power, called the mean absolute error and \(E_{rms}\) is the square root of the mean of the squared difference between the measured and predicted value of power, called the root mean squared error [9].

3. Calculation results and discussion

A comparison of measured and estimated wind speed probabilities for 800-kW wind turbine installed at a site in Jeju Island, Korea is shown in Fig. 1. The shape parameter, \(\beta\), and scale parameter, \(\eta\) obtained from Eqs. (5) and (6) using the wind speed data of 448 points averaged at 5-min are 2.43 and 11.185, respectively. The average value of the wind speed calculated by Eq. (7) is approximately 9.92 m/s and the dispersion is approximately 18.9 m/s. The wind speed distribution at this site is almost Gaussian whose most probable speed is 8.6 m/s. The Weibull wind speed probability with the parameters obtained by MLE method and the measured wind speed distribution represented by histograms are nearly similar and overlapping each other as shown in this figure.

The measured and estimated power curves for a 800-kW wind
turbine are shown in Fig. 2. The vector parameter \( \theta \) estimated using Eqs. (15)–(18) for the logistic function given in Eq. (9) is \( \theta = (1.71, 800.5, -5.5, 48.0) \). The cut-in, rated and cut-off speeds of the 800-kW wind turbine are 3.5, 12.0 and 25.0 m/s, respectively. The estimated power curve closely follows the measured power curve except the values around the rated value of wind speed. The measured wind power shown in Fig. 2 is displayed in time series in Fig. 3 along with the predicted power values by the logistic function. Remarkably, the predicted power denoted by grey line closely follows the measured power represented by dotted line. The values of errors, \( E_{\text{abs}} \) and \( E_{\text{rms}} \) are 44.2 and 57.8 kW, respectively for the power data of 448 points. It is noted that somewhat different parameter values in the logistic function, \( \theta = (1.6, 820.0, -7.3, 54.4) \) which satisfies Eq. (15)–(18) with a certain criterion, provided similar power values depending on wind speeds and the value of errors. Monitoring power output from 800-kW wind turbine was tested using the logistic function whose parameters were obtained from 448 data points as shown in Fig. 2. Fig. 4 shows the observed and predicted power by logistic function in time series at different time duration from the case shown in Fig. 3. The values of the errors, \( E_{\text{abs}} \) and \( E_{\text{rms}} \) are 38.3 and 50.0 kW, respectively in this case. Both measured and predicted powers are closely fit each other. Kusiak et al.’s study [7] has revealed that the k-nearest neighbor model, combined with the principal component analysis approach predicts the time dependent power values which closely follow the

![Fig. 2. Measured (full circles) and estimated (dotted line) power (kW) curves from a 800-kW wind turbine. The cut-in, rated and cut-off speeds are 3.5, 12.0 and 25.0 m/s, respectively. The vector parameter for the logistic function is \( \theta = (1.71, 800.5, -5.5, 48.0) \).](image)

![Fig. 3. Measured (dotted line) and predicted (grey line) power (kW) by logistic function with vector parameter of \( \theta = (1.71, 800.5, -5.5, 48.0) \) from a 800-kW wind turbine in time series. The time scale is 5-min.](image)

![Fig. 4. Monitoring turbine powers at different time interval of 37.5 h. Measured (dotted line) and predicted (grey line) power (kW) by logistic function with vector parameter of \( \theta = (1.71, 800.5, -5.5, 48.0) \) from a 800-kW wind turbine in time series. The time scale is 5-min.](image)

![Fig. 5. Measured (full circle) and estimated (solid line) power (kW) curves from a 2-MW wind turbine. The cut-in, rated and cut-off speeds are 3.5, 13.0 and 25.0 m/s, respectively. The vector parameter for logistic function \( \theta = (1.916, 2100.5, -8.5, 75.3) \).](image)
observed ones. They used measured wind speed and the corresponding power output data which was averaged and stored at 10-min interval for one month. However, they obtained parameter values for the logistic function by using an evolutionary computation algorithm to minimize the cost function for the LSM given in Eq. (19).

The measured and predicted power curves for a 2-MW wind turbine installed in other site in Jeju Island, Korea are shown in Fig. 5. The vector parameter estimated by MLE method for the logistic function is \( \theta=(1.916, 2100.5, -8.5, 75.3) \). The cut-in, rated and cut-off speeds of the 2-MW wind turbine are 3.5, 13.0 and 25.0 m/s, respectively. The power curve produced from 488 data points and the predicted power curve from logistic function are almost identical as shown in Fig. 5. The value of the shape parameter (\( \beta \)) and scale parameter (\( \eta \)) used in the evaluation of the vector parameter are 2.98 and 11.922, respectively. The average value of the wind speed calculated by Eq. (7) is approximately 10.7 m/s and the dispersion is approximately 15.3 m/s. Gaussian distribution for the wind speed probability was also obtained in this location. The measured power averaged at 10-min with dotted line along with the predicted power by the logistic function with grey line is shown in Fig. 6. The predicted power appears to closely follow the measured power. The values of the errors, \( E_{\text{abs}} \) and \( E_{\text{rms}} \), are 94.2 and 174.3 kW, respectively. Monitoring wind power from 2-MW wind turbine is also shown in Fig. 7, which shows the observed and predicted powers by the logistic function. The value of the errors, \( E_{\text{abs}} \) and \( E_{\text{rms}} \) in this case are 233.6 and 397.4 kW, respectively. Quite close agreement between the measured and predicted powers can be seen, which indicates that the parametric model based on MLE method can be used as a reference power curve of a turbine in a normal status.

4. Conclusions

A nonlinear parametric model of a power curve using logistic function was constructed by data-driven approach in this study. Equations for the scale and shape parameters in the Weibull wind speed distribution and equations for the four parameters in the logistic function were obtained explicitly by MLE method. The values of the parameters needed in obtaining the most probable PDF for wind speed and power were obtained by solving the equations using a selected wind speed and the corresponding power data set. The predicted power closely follows the measured power averaged at 5-min for 800-kW and the measured power averaged at 10-min for 2-MW wind turbine, which indicates that the parametric model constructed in this study can monitor the online performance of a wind farm. Better forecasting of wind power production is possible by a wind power modeling method presented in this study with available wind speed forecasting [13]. Day-ahead load forecasts which have already been developed for the conventional grid systems [14] are required to plan maintenance of wind farms and to schedule grid maintenance and energy storage operations [15].

Acknowledgement

Financial support by the Grant-in-Aid from the Korea Institute of Energy Technology Evaluation and Planning (Project 2015201103100) is greatly appreciated.

Appendix

Differentiating Eq. (4) with respect to \( \beta \) and \( \eta \), we have

\[
N = \sum_{i=1}^{N} \left( u_i / \eta \right)^\beta
\]  
(\text{A1})

\[
1 / \beta = \left[ \sum_{i=1}^{N} \left( u_i / \eta \right)^\beta \ln(u_i / \eta) \right] - \frac{1}{N} \sum_{i=1}^{N} \ln(u_i / \eta)
\]  
(\text{A2})

From Eq. (A1), one can get Eq. (5) easily. In Eq. (A2), the \( \eta \) in the last two terms cancel each other so that Eq. (A2) becomes Eq. (6).
However, it is better to use Eq. (A2) instead of Eq. (6) to obtain proper solution in short time. Equations (15)–(18) can be obtained from Eq. (14) after some manipulations.

References
