Energy Conversion and Management 173 (2018) 56–64

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

Integrated approach for the determination of an accurate wind-speed distribution model

Nurulkamal Masseran

School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

ARTICLE INFO

Keywords:
Goodness of fit
Multi-criteria evaluation
Multiplicative aggregation
Ranking score
Wind-speed modeling

ABSTRACT

The distribution model of wind-speed data is critical for the assessment of wind-energy potential because it reduces uncertainties in the estimation of wind power output. Thus, an accurate distribution model for describing wind-speed data should be determined before a detailed analysis of energy potential is conducted. In this study, information from several goodness-of-fit criteria, e.g., the R^2 coefficient, Kolmogorov–Smirnov statistic, Akaike's information criterion, and deviation in skewness/kurtosis were integrated for the conclusive selection of the best-fit distribution model of wind-speed data. The proposed approach integrates standardized scores and subjects each criterion to multiplicative aggregation. The approach was applied in a case study to fit eight statistical distributions to hourly wind-speed data collected at two stations in Malaysia. The results showed that the proposed approach provides a good basis for the selection of the optimal wind-speed distribution model. Furthermore, graphical representations agreed with the analytical results.

1. Introduction

The use of wind energy was rapidly expanded worldwide to address with the crisis of energy shortages, environmental pollution, and climate change [1–3]. In fact, the global annual production rate of wind power was constantly increased in tandem with technology maturation and decreases in energy generation costs [4]. Nowadays, wind energy is a major resource of renewable energy that could supply more than 40 times of the annual global electricity consumption [5,6]. The increasing use of wind energy also reflects the growing awareness of many countries to provide clean and safe energy. In addition, the application and development of wind power as a renewable energy source provides several advantages, such as cost-effectiveness; facile transportation; and opportunities for employment, research, economic activity, and independence in the electricity sector [7–9].

The appropriate and accurate modeling of wind-speed data is an important step in investigating the potential of wind energy because the production of wind power is strongly dependent on the characteristics and the capacity of wind magnitude [10]. The distribution model represented the variation and uncertainty of wind speed data in estimating the available energy potential [11]. A review of the literature shows that numerous studies have described the application of various distribution models in estimating and evaluating the potential of wind power in a particular region. In fact, the Weibull distribution is a widely used model in the wind industry sector. For examples, the information from the Weibull distribution has been used to estimate the wind power corresponds to the wind turbine capacity factor [12–15]. Apart from that, the Weibull distribution has also been used in many applications such as, the model estimation for evaluating the wind power performance system [16,17], the failure model for the wind turbine [18], the power curves estimation for wind turbine related to the power output [19], statistical mapping of wind power characteristics [20–26], the degradation model for wind turbine [27] and many more, see [28–50].

However, not all wind regimes can be modeled using the Weibull distribution. For examples, Aries et al. [51] used eight distribution models, namely, Gamma, Weibull, Lognormal, Gumbel, Generalized Logistic, Nakagami, and Inverse Gaussian distribution to model wind-speed data collected from four different sites in Algeria. They found that different distribution models could be used to best fit the data collected from each different site. Jung and Schindler [6] comprehensively assessed global wind-speed data by using 24 different single-distribution models and 21 mixed-distribution models. They also used different distribution models to show the spatial information of different regions. Ouarda et al. [52] evaluated 13 parametric wind speed distribution models to describe the data of wind speed in several sites within the United Arab Emirates (UAE). Based on the goodness-of-fit assessment, they found that different sites have different suitable models for wind speed data. Kantar et al. [53] used the Gamma, Weibull, Rayleigh, Lognormal, and extended generalized Lindley distribution to model wind-speed data collected from four stations in Turkey. They also found

E-mail address: kamalmsn@ukm.edu.my.

https://doi.org/10.1016/j.enconman.2018.07.066
Received 1 March 2018; Received in revised form 10 July 2018; Accepted 21 July 2018

that distribution models other than the Weibull could be a good alternative model for wind-speed data collected from different locations. Alavi et al. [54] compare 8 different distribution models to determine the best model for offshore wind speed calculations in east and southeast parts of Iran. The results of their study conclude that the Gamma, Inverse Gaussian, Nakagami and Generalized Extreme Value distribution can perform better for modeling wind speed compared with the Weibull. Carta et al. [55] provided a comprehensive review of the wind speed distribution models. They have determined that 11 of the wind speed distribution models have been used by most of the researchers all over the world for modelling and assessing wind energy potential. Apart from that, there are many research studies that have performed an analysis regarding wind energy by considering several distribution models simultaneously to provide more accurate results for wind energy calculation and estimation (see [56–67]). In summary, the literature reviews concluded that different distribution models will tend to represent different characteristics of different wind regimes. Thus, the selection of an accurate distribution model is an important issue that needs to be addressed before performing further analysis.

The selection of an accurate distribution model for wind-speed data involves the measurement of several criteria, such as the R^2 coefficient, Akaike information criterion (AIC), and Kolmogorov–Smirnov (K–S) statistic. Given that different measurement techniques and criteria can be used to evaluate the goodness-of-fit of the distribution model, problems will be encountered when some methods do not provide conclusive results (variation among the results for each method). Therefore, this study attempt to resolve this issue by proposing an integrated approach in determining the best model selection for wind speed data. The proposed approach involves the integration of several main criteria.

2. Wind-speed distribution models and parameter estimators

The estimated power of wind energy is generally calculated on the basis of the wind power equation, which is given as:

$$P_E(X) = \frac{1}{2}A\rho_k X^3 C_p(\lambda, \beta) \quad (1)$$

Eq. (1) describes that wind power is dependent on a proportion of the cube for wind speed X with a constant air density ρ_k and the area (A) of the airstream that has been measured at a plane perpendicular to the direction of wind speed. $C_p(\lambda, \beta)$ is the power coefficient that considers Betz's law for a particular type of wind turbine being used [7,68]. In Eq. (1), the term $P_w(x) = \frac{1}{2}A\rho_k C_p(\lambda, \beta)$ describes the power curve for a specific wind turbine. The power curve $P_w(x)$ can be determined accurately on the basis of the specification of a particular wind turbine. Thus, the only stochastic component that influences the variation and uncertainty of wind-energy production is contributed by the wind-speed variable X. This implies that a distribution model for wind speed $f(x)$ has to be utilized to cope with the variation and uncertainty in the estimation of wind-energy production.

In the engineering practice, the estimated mean of the wind energy which produced by a wind turbine is associated with the distribution model which given by the following equation:

$$\bar{P}_w = \int_0^\infty P_w(x)f(x)dx \quad (2)$$

Thus, a highly accurate distribution model for wind speed $f(x)$ will yield an accurate estimate of the potential of wind energy [62,66,68].

This study evaluates some of the most commonly used distribution models in the previous studies on wind-speed data in Malaysia. These models include the Lognormal (LN), Weibull (WE), Rayleigh (RY), Exponential (EX), Burr (BR), Gamma (GA), Inverse Gaussian (IGU) and Inverse Gamma (IGA) [7,60,69]. Table 1 shows a list of the statistical distribution models and their maximum likelihood estimators (MLEs).

3. Data and parameter estimation

The data used in this study were obtained from the Department of Environment Malaysia. Two stations in Peninsular Malaysia were selected for this study: Kuantan and Balok Baru. Hourly wind-speed data collected over the period of January 1, 2000 to November 30, 2009 were used. The missing values in the data are found to be missing at random points. In fact, the data have a small percentage of missing value. Thus, to impute any particular missing values in the data, the method of single imputation based on the average of the last known and next known observations to the missing values are applied. This method is easy to be implemented and it's able to provide a good result for a missing data with random behaviors [70]. Apart from that, the locations of both meteorological stations are shown in Fig. 1.

The MLEs of the WE, GA, IGU, IGA, and BR parameter distributions can be numerically determined on the basis of the observed wind-speed data through several methods, such as Newton–Rapson, scoring, EM algorithm, quasi-Newton, and Nelder–Mead. This study used the Nelder–Mead method as an optimization technique to determine the MLEs of the parameters [71]. The MLEs can also be easily determined for other distributions, such as LN, RY, and EX. Table 2 shows the results of the parameters for each distribution estimated through the MLE method.

4. Criteria for the selection of the best wind-speed distribution

Various techniques and criteria have been applied to determine the best statistical models for wind-speed data. The most commonly used model for wind-speed model selection is measurement based on observed and predicted data from a fitted model using criteria, such as R^2 coefficient, mean absolute percentage error (MAPE), root-mean-square error (RMSE), or mean absolute (MAE). A low 1-R^2 value is indicative of the good fit of the theoretical distribution to the experimental data. The 1-R^2 value usually provides information similar to that provided by MAPE, RMSE, or MAE. The R^2 coefficient is used to quantify the correlation between observed probabilities and predicted data from a distribution model. The R^2 coefficient is characterized by

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{F}_i - \bar{F})^2}{\sum_{i=1}^{n}(\hat{F}_i - \bar{F})^2 + \sum_{i=1}^{n}(F_i - \hat{F})^2} \quad (3)$$

where $\bar{F} = \frac{\sum_{i=1}^{n}\hat{F}_i}{n}$. The estimated cumulative $\hat{F}$ is derived from the proposed distribution model. A high R^2 value indicates the good fit of the model of the cumulative function $\hat{F}$ to the empirical cumulative probability F. The R^2 coefficient has been used by numerous researchers to determine a suitable wind-speed distribution model as in [28,51–54,57,58,63,72]. However, in this study, the 1-R^2 value given that low criterion values are preferred.

Another criterion for model selection is based on the empirical distribution function (EDF), such as the K–S statistic. The EDF approach also includes the Cramer-von Mises and Anderson–Darling statistics. However, the K–S statistic is the most popular goodness-of-fit technique among all EDF techniques. It is calculated by comparing the cumulative distribution of the observed data versus that of the fitted data. The EDF F_n for n observations is defined as

$$F_n(x_i) = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \leqslant x_i} \quad (4)$$

where $I_{X_i \leqslant x}$ is an indicator function. The indicator function will be equal to 1 if $X_i \leq x$ and 0 otherwise. The K-S statistic for a given theoretical cumulative distribution function $F(x)$ is given by

$$D_n = \sup_{x_i}|F_n(x_i) - F(x_i)| \quad (5)$$