Energy Conversion and Management 93 (2015) 414–434

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Probability distributions of wind speed in the UAE

T.B.M.J. Ouarda [a,b,*], C. Charron [a], J.-Y. Shin [a], P.R. Marpu [a], A.H. Al-Mandoos [c], M.H. Al-Tamimi [c], H. Ghedira [a], T.N. Al Hosary [c]

[a] Institute Center for Water and Environment (iWATER), Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, United Arab Emirates
[b] INRS-ETE, National Institute of Scientific Research, 490 de la Couronne, Quebec City, QC G1K9A9, Canada
[c] National Centre of Meteorology and Seismology, P.O. Box 4815, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Article history:
Received 13 October 2014
Accepted 13 January 2015
Available online 5 February 2015

Keywords:
Probability density function
Model selection criteria
Wind speed distribution
Kappa distribution
Coefficient of determination
Mixture distribution
Non-parametric model

ABSTRACT

For the evaluation of wind energy potential, probability density functions (pdfs) are usually used to describe wind speed distributions. The selection of the appropriate pdf reduces the wind power estimation error. The most widely used pdf for wind energy applications is the 2-parameter Weibull probability density function. In this study, a selection of pdfs are used to model hourly wind speed data recorded at 9 stations in the United Arab Emirates (UAE). Models used include parametric models, mixture models and one non-parametric model using the kernel density concept. A detailed comparison between these three approaches is carried out in the present work. The suitability of a distribution to fit the wind speed data is evaluated based on the log-likelihood, the coefficient of determination R^2, the Chi-square statistic and the Kolmogorov–Smirnov statistic. Results indicate that, among the one-component parametric distributions, the Kappa and Generalized Gamma distributions provide generally the best fit to the wind speed data at all heights and for all stations. The Weibull was identified as the best 2-parameter distribution and performs better than some 3-parameter distributions such as the Generalized Extreme Value and 3-parameter Lognormal. For stations presenting a bimodal wind speed regime, mixture models or non-parametric models were found to be necessary to model adequately wind speeds. The two-component mixture distributions give a very good fit and are generally superior to non-parametric distributions.

1. Introduction

The characterization of short term wind speeds is essential for the evaluation of wind energy potential. Probability density functions (pdfs) are generally used to characterize wind speed observations. The suitability of several pdfs has been investigated for a number of regions in the world. The choice of the pdf is crucial in wind energy analysis because wind power is formulated as an explicit function of wind speed distribution parameters. A pdf that fits more accurately the wind speed data will reduce the uncertainties in wind power output estimates.

The 2-parameter Weibull distribution (W2) and the Rayleigh distribution (RAY) are the pdfs that are the most commonly used in wind speed data analysis especially for studies related to wind energy estimation [35,27,42,51,45,7,16,23,36,3,2,1,25,40,10,31,43,46]. The W2 is by far the most widely used distribution to characterize wind speed. The W2 was reported to possess a number of advantages ([58] for instance): it is a flexible distribution; it gives generally a good fit to the observed wind speeds; the pdf and the cumulative distribution function (cdf) can be described in closed form; it only requires the estimation of 2 parameters; and the estimation of the parameters is simple. The RAY, a one parameter distribution, is a special case of the W2 when the shape parameter of this latter is set to 2. It is most often used alongside the W2 in studies related to wind speed analysis [27,16,3].

Despite the fact that the W2 is well accepted and provides a number of advantages, it cannot represent all wind regimes encountered in nature, such as those with high percentages of null wind speeds, and bimodal distributions [15]. Consequently, a number of other models have been proposed in the literature including standard distributions, non-parametric models, mixtures of distributions and hybrid distributions. A 3-parameter Weibull (W3) model with an additional location parameter has been used by Stewart and Essenwanger [55] and Tuller and Brett [58]. They concluded to a general better fit with the W3 instead of the ordinary W2. Auwera et al. [9] proposed the use of the Generalized Gamma distribution (GG), a generalization of the W2 with an additional

* Corresponding author at: Institute Center for Water and Environment (iWATER), Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, United Arab Emirates. Tel.: +971 2 810 9107.
E-mail address: touarda@masdar.ac.ae (T.B.M.J. Ouarda).

http://dx.doi.org/10.1016/j.enconman.2015.01.036
0196-8904/

Nomenclature

C_V	coefficient of variation	LN3	3-parameter Lognormal distribution
C_S	coefficient of skewness	MGG	mixture of two Gamma pdfs
C_K	coefficient of kurtosis	ML	maximum likelihood
cdf	cumulative distribution function	MM	method of moments
χ^2	Chi-square test statistic	MWW	mixture of two 2-parameter Weibull pdfs
D/M	distribution/method	n	number of wind speed observations in a series of wind speed observations
EV1	Gumbel or extreme value type I distribution	N	number of bins in a histogram of wind speed data
$f_{\hat{\theta}}()$	probability density function with estimated parameters $\hat{\theta}$	P3	Pearson type III distribution
$\hat{f}()$	estimated probability density function	pdf	probability density function
F_i	empirical probability for the ith wind speed observation	R^2	coefficient of determination
$\hat{F}_i$	estimated cumulative probability for the ith observation obtained with the theoretical cdf	R^2_{PP}	coefficient of determination giving the degree of fit between the theoretical cdf and the empirical cumulative probabilities of wind speed data.
$F()$	cumulative distribution function	R^2_{QQ}	coefficient of determination giving the degree of fit between the theoretical wind speed quantiles and the wind speed data.
$F^{-1}()$	inverse of a given cumulative distribution function	RAY	Rayleigh distribution
G	Gamma distribution	rmse	root mean square error
GEV	generalized extreme value distribution	v_i	the ith observation of the wind speed series
GG	generalized Gamma distribution	$\hat{v}_i$	predicted wind speed for the ith observation
GMM	generalized method of moment	W2	2-parameter Weibull distribution
K()	kernel function	W3	3-parameter Weibull distribution
KAP	Kappa distribution	WMM	weighted method of moments
KE	Kernel density distribution		
KS	Kolmogorov–Smirnov test statistic		
LN2	2-parameter Lognormal distribution		

shape parameter, for the estimation of mean wind power densities. They found that it gives a better fit to wind speed data than several other distributions. Recently, a variety of other standard pdfs have been used to characterize wind speed distributions [15,64,38,41,39,54]. These include the Gamma (G), Inverse Gamma (IG), Inverse Gaussian (IGA), 2 and 3-parameter Lognormal (LN2, LN3), Gumbel (EV1), 3-parameter Beta (B), Pearson type III (P3), Log-Pearson type III (LP3), Burr (BR), Erlang (ER), Kappa (KAP) and Wakeby (WA) distributions. Some studies considered non-stationary distributions in which the parameters evolve as a function of a number of covariates such as time or climate oscillation indices [30]. This approach allows integrating in the distributional modeling of wind speed information concerning climate variability and change.

To account for bimodal wind speed distributions, mixture distributions have been proposed by a number of authors [13,5,15,18,49]. The common models used are a mixture of two W2 and a mixture of a normal distribution singly truncated from below with a W2 distribution. In Carta et al. [15], the mixture models were found to provide a good fit for bimodal wind regimes. They were also reported to provide the best fits for unimodal wind regimes compared to standard distributions.

Non-parametric models were also proposed by a number of authors. The most popular are distributions generated by the maximum entropy principle [37,50,4,18,62]. These distributions are very flexible and have the advantage of taking into account null wind speeds. Another non-parametric model using the kernel density concept was proposed by Qin et al. [48]. This approach was applied by Zhang et al. [63] in a multivariate framework.

Because a minimal threshold wind speed is required to be recorded by an anemometer, null wind speeds are often present. However, for many distributions, including the W2, null wind speeds or calm spells are not properly accounted for because the cdf of these distributions gives a null probability of observing null wind speeds (i.e. $F_X(0) = 0$, where $F_X(x)$ is the cdf of a given variable X). Takle and Brown [57] introduced what they called the "hybrid density probability" to consider null wind speeds. The zero values are first removed from the time series and a distribution is fitted to the non-zero series. The zeros are then reintroduced to give the proper mean and variance and renormalize the distribution. Carta et al. [15] applied hybrid functions with several distributions and concluded that there is no indication that hybrid distributions offer advantages over the standard ones.

In order to compare the goodness-of-fit of various pdfs to sample wind speed data, several statistics have been used in studies related to wind speed analysis. The most frequently used ones are the coefficient of determination (R^2) [24,17,3,37,50,15,41,54,63], the Chi-square test results (χ^2) [9,20,22,3,18], the Kolmogorov–Smirnov test results (KS) [35,34,58,47,22,18,48,60] and the root mean square error (rmse) [35,34,9,52,3,18]. In most studies, a visual assessment of fitted pdfs superimposed on the histograms of wind speed data is also performed [42,6,59,7,36,32,18,48,19]. R^2 and rmse are either applied on theoretical cumulative probabilities against empirical cumulative probabilities (P–P plot) [50,15,41,54] or on theoretical wind speed quantiles against observed wind speed quantiles (Q–Q plot) [24,17,3,37,63]. These statistics are also sometimes computed with wind speed data in the form of frequency histograms [14,15,64,48,60].

In addition to the analysis performed on wind speed distributions, some authors have also evaluated the suitability of pdfs to fit the power distributions obtained by sample wind speeds or to predict the energy output [9,52,17,37,25,64,18,41,19]. In this case, pdfs are first fitted to the wind speed data. Then, theoretical power density distributions are derived from the pdfs fitted to wind speed. Finally, measures of goodness-of-fit are computed using the theoretical wind power density distributions and the estimated power distribution from sample wind speeds. Alternatively, analyses are also performed on the cube of wind speed which is proportional to the wind power [27,15].

A relatively limited number of studies have been conducted on the assessment of pdfs to model wind speed distributions in the Arabian Peninsula or neighboring regions: Algifri [6] in Yemen, Mirhosseini [40] in Iran, Sulaiman et al. [56] in Oman, and Şahin