Mixture probability distribution functions to model wind speed distributions

Ravindra Kollu[1*], Srinivasa Rao Rayapudi[1], SVL Narasimham[2] and Krishna Mohan Pakkurthi[1]

Abstract

Accurate wind speed modeling is critical in estimating wind energy potential for harnessing wind power effectively. The quality of wind speed assessment depends on the capability of chosen probability density function (PDF) to describe the measured wind speed frequency distribution. The objective of this study is to describe (model) wind speed characteristics using three mixture probability density functions Weibull-extreme value distribution (GEV), Weibull-lognormal, and GEV-lognormal which were not tried before. Statistical parameters such as maximum error in the Kolmogorov-Smirnov test, root mean square error, Chi-square error, coefficient of determination, and power density error are considered as judgment criteria to assess the fitness of the probability density functions. Results indicate that Weibull-GEV PDF is able to describe unimodal as well as bimodal wind distributions accurately whereas GEV-lognormal PDF is able to describe familiar bell-shaped unimodal distribution well. Results show that mixture probability functions are better alternatives to conventional Weibull, two-component mixture Weibull, gamma, and lognormal PDFs to describe wind speed characteristics.

Keywords: Mixture distributions, Probability density functions, Wind speed distribution

Background

Growing global population along with fast depleting reserves of fossil fuels is influencing researchers to search for clean and pollution-free sources of energy such as solar, wind, and bioenergies. Wind energy is a never ending natural resource which has shown its great potential in combating climate change while ensuring clean and efficient energy. Further, rapid advances in wind turbine technology led to significant growth of wind power generation across the world. However, wind energy is more sensitive to variations with topography and wind patterns compared to solar energy. Wind energy can be harvested economically if the turbines are installed in a windy area and suitable turbine is properly selected. Wind speed forecasting is a critical factor in assessing wind energy potential and performance of wind energy conversion systems. Several probability density functions (PDFs) have been used in literature to describe wind speed characteristics which include Weibull, Rayleigh, bimodal Weibull, lognormal, gamma, etc.

Islam et al. [1] used two-parameter Weibull distribution function for wind speed forecasting and assessed wind energy potentiality at Kudat and Labuan, Malaysia. Celik [2] used Weibull-representative wind data instead of the measured data in time-series format for estimating the wind energy and had shown that estimated wind energy is highly accurately. Celik [3] made statistical analysis of wind data at southern region of Turkey and summarized that Weibull model was better than Rayleigh model in fitting the measured data distributions. Akdag et al. [4] discussed the suitability of two-parameter Weibull wind speed distribution and the two-component mixture Weibull distribution (WW-PDF) to estimate wind speed characteristics. Carta et al. [5] used WW-PDF because it is able to represent heterogeneous wind regimes in which there was evidence of bimodality or bitangentiality or, simply, unimodality. Maximum likelihood and least-square methods were used to estimate WW-PDF parameters. In [6], wind speed distributions were shown to be satisfactorily described with a log-normal function, and in [7], Weibull and lognormal distribution functions were used to fit wind speed distributions. Kiss and Imre [8] used Rayleigh, Weibull, and gamma distributions to model wind speeds both over land and sea. They found that

* Correspondence: ravikollu@gmail.com
[1]Department of Electrical and Electronics Engineering, J.N.T. University Kakinada, Kakinada, Andhra Pradesh 533003, INDIA
Full list of author information is available at the end of the article

generalized gamma distribution, which has independent shape parameters for both tails, provides an adequate and unified description almost everywhere. Generalized extreme value (GEV) distribution that combines the Gumbel, Frechet, and Weibull extreme value distributions were used to model extreme wind speeds [9-12]. In recent past, mixture distributions were used to estimate wind energy potential that are quite accurate in describing wind speed characteristics. Jaramillo and Borja [13] used mixture Weibull distribution (WW) to model bimodal wind speed frequency distribution. Akpinar et al. [14] used mixture normal and Weibull distribution (NW), which is a mixture of truncated normal distribution, and conventional Weibull distribution to model wind speeds. Tian Pau [15] employed mixture gamma and Weibull distribution (GW) which is a combination of gamma and Weibull distributions, and also mixture normal distribution (NN) which is a mixture function of two-component truncated normal distribution for wind speed modeling.

The objective of this study is to propose three new mixture distributions, *viz.*, Weibull-lognormal (WL), GEV-lognormal (GEVL), and Weibull-GEV (WGEV) for wind speed forecasting. Comparison of the proposed mixture distributions with existing distribution functions is done to demonstrate their suitability in describing wind speed characteristics.

The rest of this paper is organized as follows: wind distribution models and goodness of fit tests used in this paper are presented in the section 'Methods'. Results derived from this study are discussed in 'Results and discussions' section. Details about the data used for the analysis are given in this section. Conclusions are presented in the 'Conclusions' section.

Methods
Significance
The most suitable wind turbine model which needs to be installed in a wind farm is selected by careful wind energy resource evaluation. Accurate evaluation could be done using best fit distribution model. Using inappropriate distribution models results in inaccurate estimation of wind turbine capacity factor and annual energy production which in turn leads to improper estimation of levelized production cost [13]. Hence, it is important to choose an accurate distribution model which closely mimics the wind speed distribution at a particular site.

Wind distribution models
Wind distribution modeling requires analysis of wind data over a number of years. To reduce the expenses and time required to process long-term wind speed data, it is desirable to use statistical distribution functions for describing the wind speed variations. The primary tools to describe wind speed characteristics are probability density functions. The parameters of probability distribution functions which describe wind-speed frequency distribution are estimated using statistical data of a few years. Many PDFs have been proposed in recent past, but in present study Weibull, Lognormal, gamma, GEV, WW-PDF, mixture gamma and Weibull distribution, mixture normal distribution, mixture normal and Weibull distribution, and three new mixture distributions, *viz.*, Weibull-lognormal, GEV-lognormal, and Weibull-GEV are used to describe wind speed characteristics. Parameters defining each distribution function are calculated using maximum likelihood method.

Weibull distribution
The Weibull function is commonly used for fitting measured wind speed probability distribution. Weibull distribution with two parameters is given by [1]:

Weibull PDF

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

Weibull cumulative distribution function (CDF):

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

Weibull shape and scale parameters are calculated using the maximum likelihood method [16] which is given by:

$$k = \left[\frac{\sum_{i=1}^{n} v_i^k \ln(vi)}{\sum_{i=1}^{n} v_i^k} - \frac{\sum_{i=1}^{n} \ln(vi)}{n}\right]^{-1} \quad (3)$$

where v_i is the wind speed in time step i and n is the number of data points. To evaluate (3) an iterative technique is used. Scale parameter is obtained by

$$c = \left[\frac{1}{n}\sum_{i=1}^{n} v_i^k\right]^{1/k} \quad (4)$$

Generalized extreme value distribution
GEV distribution is a flexible model that combines the Gumbel, Frechet and Weibull maximum extreme value distributions [11,17]. GEV PDF is given by

$$e(v, \zeta, \delta, 1) = \left(\frac{1}{\delta}\right)\left(1 + \frac{\zeta(v-1)}{\delta}\right)^{-\frac{1}{\zeta}-1}$$

$$exp\left[-\left(1 + \frac{\zeta(v-l)}{\delta}\right)\right]^{\frac{1}{\zeta}} \; if \; \zeta \neq 0 \quad (5)$$

GEV CDF [17] is given by

$$E(v, \zeta, \delta, l) = exp\left[-\left(1 + \frac{\zeta(v-1)}{\delta}\right)\right]^{\frac{-1}{\zeta}} \quad \text{if} \quad \zeta \neq 0 \tag{6}$$

GEV parameters are calculated using the maximum likelihood method which maximizes the Logarithm of likelihood function given by

$$LL = \ln\prod_{i=1}^{n}\{e(v_i; \zeta, \delta, l)\} = \Sigma_{i=1\ln}^{n}\{e(v_i; \zeta, 1)\} \tag{7}$$

Lognormal distribution

Lognormal distribution is probability distribution of a random variable whose logarithm is normally distributed.

Lognormal PDF is given by [18,19]

$$ln(v, \emptyset, \lambda) = \frac{1}{v\emptyset\sqrt{2\pi}} \quad exp\left[\frac{-(\ln(v) - \lambda^2)}{2\emptyset^2}\right] \tag{8}$$

Lognormal CDF is written as [18]

$$LN(v, \emptyset, \lambda) = \frac{1}{2} + \frac{1}{2} \quad erf\left[\frac{\ln(v) - \lambda}{\emptyset\sqrt{2}}\right]$$

where

$$erf(v) = \frac{2}{\sqrt{\pi}}\int_0^v \quad exp(-t^2)dt \tag{9}$$

Lognormal parameters λ and Φ estimated using maximum likelihood method which do not need an iterative procedure are given by [20]

$$\lambda = \frac{1}{N}\sum_{i=1}^{N}\ln(vi); \quad \emptyset^2 = \frac{1}{N}\sum_{i=1}^{N}[\ln(vi) - \lambda]^2 \tag{10}$$

Gamma distribution

The probability density function of gamma distribution is expressed using the below function [15]

$$g(v, \alpha, \beta) = \frac{v^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} \quad exp\left[-\frac{v}{\beta}\right] \tag{11}$$

The cumulative Gamma distribution function is given by [20]

$$G(v, \alpha, \beta) = \int \frac{v^{\alpha-1}}{\beta^{\alpha}\Gamma(\alpha)} \quad exp\left[-\frac{v}{\beta}\right]dv \tag{12}$$

Gamma distribution parameters are estimated using maximum likelihood method that maximizes the logarithm of likelihood function which is given by:

$$LL = \ln\prod_{i-1}^{n}\{h(v_i; \alpha, \beta)\} = \sum_{i-1}^{n}\ln\{h(v_i; \alpha, \beta)\} \tag{13}$$

Two-component mixture Weibull distribution

The probability density function, which depends on five parameters $(v; k_1, c_1, k_2, c_2, w)$ is given by [5]

$$\begin{aligned} ff(v; k_1, c_1, k_2, c_2, w) &= wf(v; k_1, c_1) \\ &+ (1-w)f(v; k_2, c_2) \end{aligned} \tag{14}$$

The cumulative distribution function is given by [5]

$$\begin{aligned} FF(v; k_1, c_1, k_2, c_2, w) &= wF(v; k_1, c_1) \\ &+ (1-w)F(v; k_2, c_2) \end{aligned} \tag{15}$$

Relevant likelihood function is

$$LL = \sum_{i=1}^{n}\ln\{wf(v; k_1, c_1) + (1-w)f(v; k_2, c_2)\} \tag{16}$$

Mixture gamma and Weibull distribution

The probability density function and cumulative distribution function of the mixture gamma and Weibull distribution are given by [15]

$$\begin{aligned} h(v; \alpha, \beta, k, c, w) &= wg(v; \alpha, \beta) \\ &+ (1-w)f(v; k, c) \end{aligned} \tag{17}$$

$$\begin{aligned} H(v; \alpha, \beta, k, c, w) &= wG(v; \alpha, \beta) \\ &+ (1-w)F(v; k, c) \end{aligned} \tag{18}$$

Relevant likelihood function is

$$LL = \sum_{i-1}^{n}\ln\{wg(v; \alpha, \beta) + (1-w)f(v; k, c)\} \tag{19}$$

Mixture normal distribution

The probability density function of singly truncated normal distribution is given by [15]

$$q(v; \mu, \sigma) = \frac{1}{I(\mu, \sigma)\sigma\sqrt{2\pi}}exp\left[-\frac{v-\mu^2}{2\sigma^2}\right] \quad for \quad v \geq 0, \tag{20}$$

where $I(\mu, \sigma)$ is the normalization factor that leads the integration of the truncated distribution to one is expressed as

$$I(\mu, \sigma)\frac{1}{\sigma\sqrt{2\pi}}\int_0^{\infty} \quad exp\left[-\frac{(v-\mu)^2}{2\sigma^2}\right]dv. \tag{21}$$

The cumulative truncated normal distribution is given by

$$Q(v; \mu, \sigma) = \int_0^v \frac{1}{I(\mu, \sigma)\sigma\sqrt{2\pi}} \quad exp\left[-\frac{v-\mu^2}{2\sigma^2}\right]dv \tag{22}$$

The mixture function of two component truncated normal distribution from the above can be written as [15]

$$r(v; \mu_1\sigma_1, \mu_2, \sigma_2 w) = wq(v; \mu_1\sigma_1) \\ + (1 - w)q(v; \mu_2\sigma_2) \quad (23)$$

The cumulative distribution function is given by

$$R(v; \mu_1\sigma_1, \mu_2, \sigma_2 w) = wQ(v; \mu_1\sigma_1) \\ + (1 - w)Q(v; \mu_2\sigma_2) \quad (24)$$

Relevant likelihood function to estimate the five parameters is

$$LL = \sum_{i=1}^{n} \ln\{wq(v; \mu_1\sigma_1) + (1 - w)q(v; \mu_2\sigma_2)\} \quad (25)$$

Mixture normal and Weibull distribution

The probability density function of the mixture distribution comprising of truncated normal and conventional Weibull is written as [15]

$$s(v; \mu, \sigma, k, c) = wq(v; \mu, \sigma) + (1 - w)f(v; k, c) \quad (26)$$

Its cumulative distribution function is given as

$$S(v; \mu, \sigma, k, c) = wQ(v; \mu, \sigma) + (1 - w)F(v; k, c) \quad (27)$$

Relevant likelihood function to estimate the five parameters is

$$LL = \sum_{i=1}^{n} \ln\{wq(v; \mu, \sigma) + (1 - w)f(v; k, c)\} \quad (28)$$

Mixture Weibull and GEV distribution

The probability density function of the mixture distribution comprising Weibull and GEV functions which is applied for the first time to model wind speed distribution is written as

$$t(v; k, c, \zeta, \delta, l) = wf(v; k, c) \\ + (1 - w)e(v; \zeta, \delta, l) \quad (29)$$

Its cumulative distribution function is given as

$$T(v; k, c, \zeta, \delta, l) = wF(v; k, c) \\ + (1 - w)E(v; k, c, \zeta, \delta, l) \quad (30)$$

Relevant likelihood function to estimate the six parameters is

$$LL = \sum_{i=1}^{n} \ln\{wf(v; k, c) + (1 - w)e(v; \zeta, \delta, l)\} \quad (31)$$

Mixture Weibull and lognormal distribution

The probability density function of the mixture distribution comprising Weilbull and lognormal of functions which is applied for the first time to model wind speed distribution is written as

$$u(v; k, c, \lambda, \phi) = wf(v; k, c) + (1 - w)l(v; \lambda, \phi) \quad (32)$$

Its cumulative distribution function is given as

$$U(v; k, c, \lambda, \phi) = wF(v; k, c) \\ + (1 - w)L(v; \lambda, \phi) \quad (33)$$

Relevant likelihood function to estimate the five parameters is

$$LL = \sum_{i=1}^{n} \ln\{wf(v; k, c) + (1 - w)L(v; \lambda, \phi)\} \quad (34)$$

Mixture GEV and lognormal distribution

The probability density function of the mixture distribution comprising GEV and lognormal functions which is applied for the first time to model wind speed distribution is written as

$$v(v; \zeta, \delta, l, \lambda, \phi) = we(v; \zeta, \delta, l) \\ + (1 - w)l(v; \lambda, \phi) \quad (35)$$

Its cumulative distribution function is given as

$$V(v; \zeta, \delta, l, \lambda, \phi) = wE(v; \zeta, \delta, l) \\ + (1 - w)L(v; \lambda, \phi) \quad (36)$$

Relevant likelihood function to estimate the five parameters is

$$LL = \sum_{i=1}^{n} \ln\{we(v; \zeta, \delta, l) + (1 - w)l(v; \lambda, \phi)\} \quad (37)$$

Goodness-of-fit tests

Goodness-of-fit tests are used to measure the deviation between the predicted data using theoretical probability function and the observed data. In this paper five statistical errors are considered as judgment criteria to evaluate the fitness of PDFs.

Kolmogorov-Smirnov test

The first one is the Kolmogorov-Smirnov test (K-S), which is defined as the maximum error in cumulative distribution functions [21].

$$K - S = max|C(v) - 0(v)| \quad (38)$$

where $C(v)$ and $O(v)$ are the cumulative distribution functions for wind speed not exceeding v calculated by distribution function and observed wind speed data respectively. Lesser K-S value indicates better fitness of the PDF.

Table 1 Computed parameter values of different probability density functions

PDF		Station 42056	Station 46012	Station 46014	Station 46054
Weibull	k	2.8941	1.8549	1.7047	1.9541
	c	7.7429	6.2994	6.9481	9.1251
GEV	ζ	−0.1033	−0.0439	−0.0226	−0.3184
	δ	2.4400	2.5942	3.0363	4.2437
	l	5.7867	4.1868	4.4836	6.6940
Lognormal	λ	1.8467	1.5285	1.5929	1.8877
	Φ	0.4643	0.6836	0.7607	0.7418
Gamma	α	5.8742	2.7481	2.3127	2.5656
	β	1.1778	2.0348	2.6804	3.1673
WW	w	0.0610	0.0001	0.5197	0.4556
	k_1	3.1485	1.8549	3.1775	4.6263
	c_1	7.8370	6.2994	10.0174	12.1321
	k_2	1.4328	1.3170	1.8689	1.7374
	c_2	5.9625	1.4795	4.1004	5.1573
GW	w	0.6058	0.0721	0.6233	0.4399
	α	12.6779	4.0155	2.3876	2.3507
	β	0.6129	0.4728	1.8325	1.9608
	k	2.3200	2.0125	3.3603	4.4201
	c	6.3285	6.6370	10.2677	11.9500
NN	w	1.0000	0.5122	0.8013	0.5256
	μ_1	6.8866	6.8843	6.5127	11.3795
	σ_1	2.6108	3.4583	4.0899	2.5038
	μ_2	0.0487	3.7791	1.6113	4.0575
	σ_2	1.0933	2.3035	2.4630	2.8832
NW	w	0.0501	0.0749	0.3818	0.4756
	μ	2.6721	8.3953	9.6324	11.4904
	σ	6.0388	3.2149	2.9104	2.4729
	k	3.1093	1.8491	1.8421	1.7367
	c	7.7920	6.0395	4.5859	5.7013
WGEV	w	0.9797	0.2878	0.5800	0.5037
	k	3.0702	2.1500	1.8891	1.7168
	c	7.8205	2.7925	4.2977	5.5608
	ζ	0.5375	−0.0704	−0.2131	−0.3100
	δ	1.2598	2.3404	2.7420	2.6007
	l	1.4598	5.6512	8.3962	10.4755
WL	w	0.5978	0.9026	0.8301	0.6502
	k	2.5095	2.0329	1.9671	3.7812
	c	7.0528	6.6999	7.7545	11.4111
	λ	2.0333	0.6824	0.8508	1.1627
	Φ	0.2564	0.6347	0.6293	0.7333
GEVL	w	0.6006	0.7307	0.5929	0.6781
	ζ	−0.1743	−0.0896	−0.2394	−0.3702
	δ	2.3409	2.4933	3.3233	3.4051

Table 1 Computed parameter values of different probability density functions *(Continued)*

l	5.0583	5.2851	6.5856	9.0274
λ	2.0746	0.8904	1.1108	1.1501
Φ	0.2461	0.6823	0.6766	0.7403

R^2 test

R^2 test is used widely for goodness-of-fit comparisons and hypothesis testing because it quantifies the correlation between the observed cumulative probabilities and the predicted cumulative probabilities of a wind speed distribution. A larger value of R^2 indicates a better fit of the model cumulative probabilities $\hat{F}$ to the observed cumulative probabilities F. R^2 is defined as [22]:

$$R^2 = \frac{\sum_{i=1}^{n} \left(\hat{F}_i - \bar{F} \right)^2}{\sum_{i=1}^{n} \left(\hat{F}_i - \bar{F} \right)^2 + \sum_{i=1}^{n} \left(F_i - \hat{F}_i \right)^2} \quad (39)$$

where $\bar{F} = \frac{1}{n} \sum_{i=1}^{n} \hat{F}_i$ The estimated cumulative probabilities $\hat{F}$ are obtained from cumulative distribution functions (CDFs).

Chi-square error

Chi-square error is used to assess whether the observed probability differs from the predicted probability. Chi-square error is given by

$$x^2 = \sum_{i=1}^{n} \frac{\left(F_i - \hat{F}_i \right)^2}{\hat{F}_i} \quad (40)$$

Root mean squared error

Root mean squared error (RMSE) provides a term-by-term comparison of the actual deviation between

Table 2 Statistical errors for different distribution functions of Station 42056

| PDF	K-S error	R^2	x^2	RMSE	PDE
Weibull	0.0275	0.9980	0.0010	0.0057	0.0035
GEV	0.0460	0.9925	0.0060	0.0102	8.8948
Lognormal	0.0994	0.9545	0.0400	0.0209	32.4888
Gamma	0.0675	0.9827	0.0127	0.0138	10.0000
WW	0.0192	0.9993	0.0002	0.0037	−0.1558
GW	0.0111	0.9998	0.0005	0.0018	−0.1072
NN	0.0189	0.9994	0.0001	0.0030	−0.4768
NW	0.0180	0.9993	0.0003	0.0037	−0.0559
WGEV	0.0201	0.9992	0.0002	0.0039	−0.6804
WL	0.0118	0.9998	0.0005	0.0019	−0.0331
GEVL	0.0098	0.9998	0.0003	0.0017	−0.2688
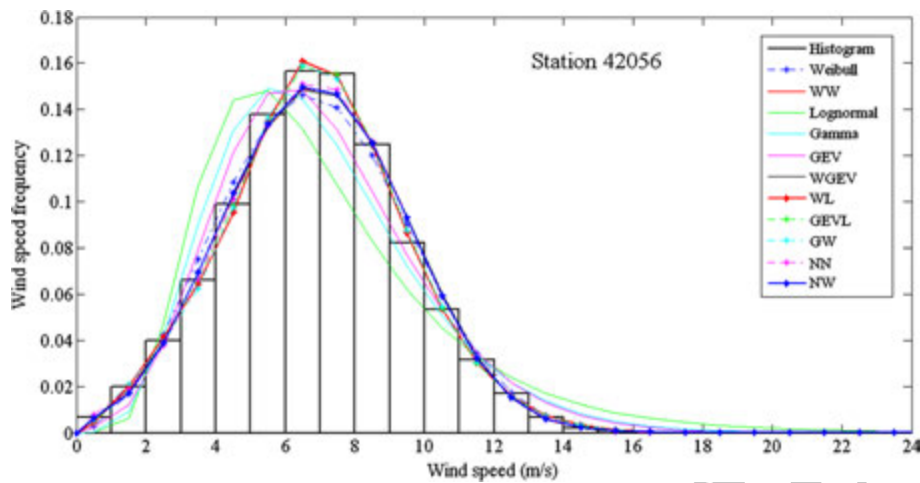
Figure 1 Predicted and observed wind frequencies of Station 42056.

observed probabilities anead predicted probabilities. A lower value of *RMSE* indicates a better distribution function model.

$$\text{RMSE} = \left[\frac{1}{n}\sum_{i=1}^{n}\left(F_i - \hat{F}_i\right)^2\right]^{1/2} \tag{41}$$

Power density error (PDE)
The relative error between the wind power density calculated from actual time-series data and that from theoretical probability function is expressed as [23]

$$\text{PDE} = \left(\frac{PD_{tp} - PD_{ts}}{PD_{ts}}\right)*100 \tag{42}$$

where PD_{ts} is the wind power density calculated from actual time-series data which is given by

$$PD_{ts} = \frac{1}{2}\rho\bar{v^3} \tag{43}$$

where PD_{tp} is the wind power density based on a theoretical probability density function $fn(v)$ which is given by

$$PD_{tp} = \frac{1}{2}\rho\int v^3 fn(v)dv \tag{44}$$

Results and discussion
Wind speed data from four wind stations were used in evaluating different PDFs to assess their suitability. Wind speed data provided by National Data Buoy Center [24] at five stations 42056 (Yucatan Basin), 46012 (Half Moon Bay, 24NM South Southwest of San Francisco, CA), 46014 (PT Arena, 19NM North of Point Arena, CA), and 46054 (Santa Barbara W 38 NM West of Santa

Barbara, CA) are used for wind speed analysis. Ten-minute mean wind speed data recorded at 5 m above the sea level are used for present studies.

- Wind data of over the period 2008 to 2010 is used for wind station 42056.
- For station 46012, wind data over a period of 10 years (2001 to 2010) is used for statistical analysis.
- Wind data of station 46014 over a period of three years (2008 to 2010) is analyzed for wind distribution modeling.
- For wind station 46054 data over the period (1999 to 2000) is used for statistical analysis.

In the present study, suitability of the PDFs is assessed using goodness-of-fit tests. All computational procedures are carried out in MATLAB software package.

Table 3 Statistical errors for different distribution functions of Station 46012

| PDF	K-S error	R^2	x^2	RMSE	PDE
Weibull	0.0169	0.9994	0.0009	0.0035	0.4728
GEV	0.0328	0.9969	0.0009	0.0079	1.6274
Lognormal	0.0838	0.9742	0.0310	0.0165	42.9656
Gamma	0.0458	0.9941	0.0082	0.0084	12.2546
WW	0.0169	0.9994	0.0009	0.0035	0.4726
GW	0.0107	0.9998	0.0003	0.0023	−0.3121
NN	0.0208	0.9992	0.0006	0.0057	−0.2180
NW	0.0156	0.9995	0.0009	0.0034	0.0310
WGEV	0.0082	0.9999	0.0002	0.0013	−0.1954
WL	0.0108	0.9998	0.0004	0.0024	−0.2721
GEVL	0.0116	0.9998	0.0001	0.0030	0.4372
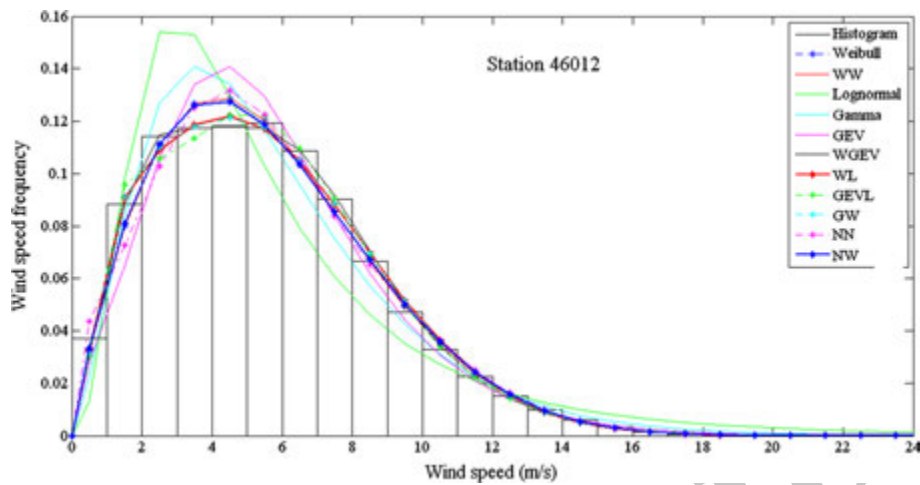
Figure 2 Predicted and observed wind frequencies of Station 46012.

Computed parameter values of different PDFs used for all the four stations are presented in Table 1.

The mean and standard deviation of observed wind speed for Station 42056 are 6.91888 m/s and 2.56768 m/s, respectively. Wind frequency histogram resembles familiar bell-shaped curve; hence, Weibull PDF fits the observed distribution well. The statistical parameters for fitness evaluation of PDFs currently analyzed are presented in Table 2. All the PDFs except lognormal, GEV, and gamma are able to describe the wind speed characteristics well which is evident from their small power density errors shown in Table 2. Considering K-S error, χ^2 error, RMSE and PDE, the distribution functions lognormal, GEV and Gamma have large errors indicating their inadequacy in modeling wind speeds. Results presented in Table 2 show clearly that proposed mixture GEVL PDF provided the best fit of observed wind speed distribution. From

Figure 1, it is evident that GEVL distribution provides a close fit throughout the entire wind speed spectrum when compared to other distributions. The higher value of R^2 and the lower values of K-S error, RMSE and chi-square error indicate that proposed GEVL distribution is more accurate than other PDFs in modeling wind speeds of Station 42056.

Station 46012 has a mean and standard deviation of 5.59185 m/s and 3.13391 m/s, respectively, for the observed wind speed. Statistical errors, K-S, R^2, χ^2, and RMSE given in Table 3 indicate that proposed mixture WGEV distribution provides best fit for the observed wind frequency distribution, which is closely followed by GW, WL, GEVL, and WW mixture distributions. Convention PDFs such as lognormal and gamma, over predicted wind speeds which are in the range of 2 to 5 m/s and 13 to 24 m/s, respectively. These PDFs have under predicted
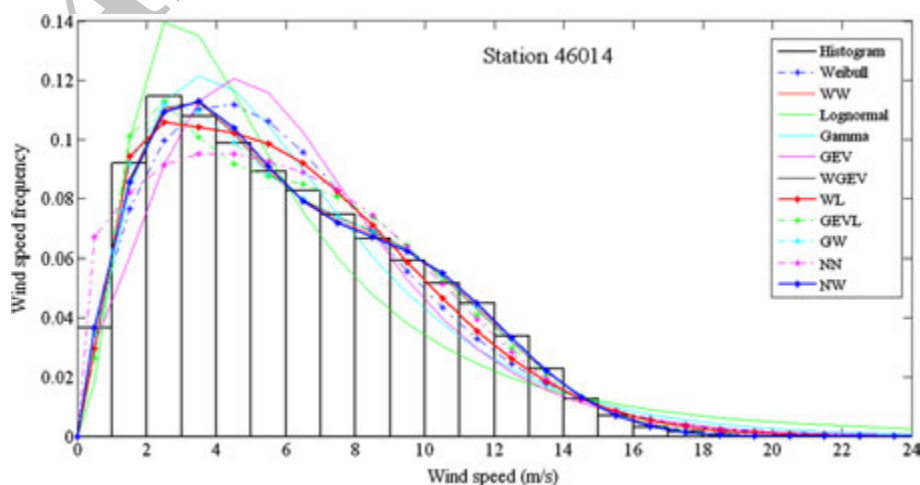

Figure 3 Predicted and observed wind frequencies of Station 46014.

Table 4 Statistical errors for different distribution functions of Station 46014

PDF	K-S error	R2	x^2	RMSE	PDE
Weibull	0.0330	0.9962	0.0004	0.0074	3.0473
GEV	0.0473	0.9913	0.0006	0.0121	2.4164
Lognormal	0.0729	0.9778	0.0246	0.0138	29.4110
Gamma	0.0447	0.9938	0.0024	0.0081	12.7017
WW	0.0111	0.9998	0.0002	0.0021	−0.0177
GW	0.0087	0.9999	0.0001	0.0018	0.5776
NN	0.0396	0.9973	0.0002	0.0096	−0.1938
NW	0.0122	0.9998	0.0003	0.0024	−0.2683
WGEV	0.0108	0.9998	0.0001	0.0021	0.1171
WL	0.0238	0.9986	0.0001	0.0049	2.1816
GEVL	0.0153	0.9994	0.0003	0.0045	0.7699

Table 5 Statistical errors for different distribution functions of Station 46054

PDF	K-S error	R^2	$x^2 x^2$	RMSE	PDE
Weibull	0.0804	0.9789	0.0172	0.0158	4.0047
GEV	0.0562	0.9886	0.0008	0.0137	−3.2096
Lognormal	0.1293	0.9317	0.0601	0.0242	10.3370
Gamma	0.1026	0.9633	0.0304	0.0188	11.4088
WW	0.0089	0.9999	0.0000	0.0019	−0.0171
GW	0.0112	0.9997	0.0000	0.0032	0.3890
NN	0.0149	0.9997	0.0000	0.0037	0.0305
NW	0.0108	0.9997	0.0000	0.0031	0.1555
WGEV	0.0093	0.9999	0.0000	0.0022	0.0499
WL	0.0208	0.9986	0.0005	0.0070	1.6018
GEVL	0.0191	0.9991	0.0002	0.0059	0.8829

speeds between 5 to 11 m/s. Apart from WGEV and WL, other mixture PDFs and conventional PDFs over predicted wind speeds in the range of 3 to 5 m/s which are reflected by the overestimated predicted probabilities as depicted in Figure 2. Results indicate that mixture PDFs perform better compared to conventional single PDFs.

Mean and standard deviation of wind speed for Station 46014 are 6.19907 m/s and 3.71888 m/s, respectively. From Figure 3, it is seen that WGEV, WG, WW, and WN distributions are able to model wind speed characteristics better than other PDFs. All other distributions have either over- or under-predicted wind speeds apart from these three. Results presented in Table 4 clearly show that, considering K-S error, χ^2, and RMSE, GW PDF has the smallest error followed by WGEV, WW, and NW. If R^2 error is considered, GW has a value very close to 1.0, confirming its superiority in performance followed by WGEV, WW, and NW distributions.

Wind regime of Station 46054 has bimodal distribution with mean and standard deviation of 8.1261 m/s and 4.2390 m/s, respectively. Compared to conventional single PDFs, mixture PDFs have performed well in modeling the wind speeds. Lognormal, gamma, Weibull, and GEV fared poorly in describing the wind characteristics compared to other mixture PDFs. As seen from Figure 4 and statistical parameters from Table 5, two component mixture Weibull distribution (WW) provided the best fit for the observed wind data, closely followed by proposed mixture function WGEV.

Figures 1 to 4 show that mixture PDFs fit much better than the conventional Weibull, lognormal, and gamma distributions. Proposed mixture distributions GEVL for Station 42056 and WGEV for Station 46012 have outperformed other existing mixture and conventional single distributions. For Station 46054, WW distribution
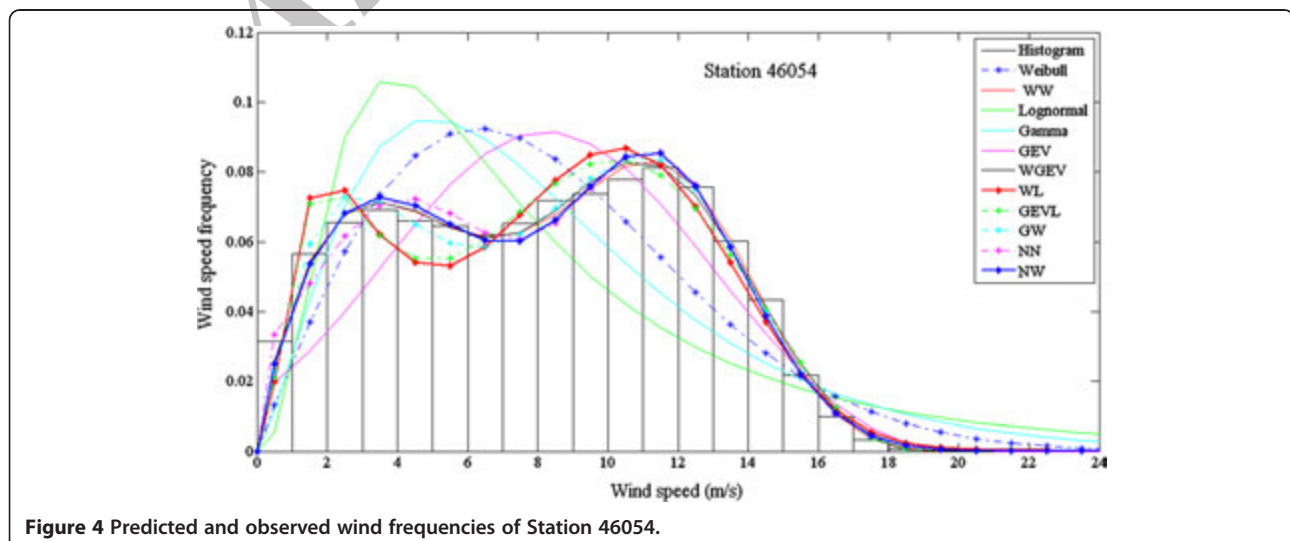

Figure 4 Predicted and observed wind frequencies of Station 46054.

provided a better fit than others, while WGEV being the close second best fit.

Conclusions

In the present article, a comparison of distribution models has been undertaken for describing wind regimes of four wind stations. Common conventional PDFs and mixture PDFs along with three proposed new mixture PDFs, *viz.*, WGEV, GEVL, and WL are used for wind speed modeling. It is shown that conventional PDFs, such as Weibull, lognormal, and gamma, are inadequate; hence, mixture functions are used to model the observed wind speed distributions better. Though the superiority of proposed mixture functions in Station 46014 is not very significant, the proposed mixture distributions GEVL for Station 42056 and WGEV for Station 46012 have provided better fit of the empirical data than other existing mixture distributions. For Station 46054, both WW and WGEV are found more suitable for describing wind speed distributions than other distributions. The performance difference between WW and WGEV distributions is not significant for this station. Results show clearly that proposed mixture PDFs, WGEV and GEVL, provide viable alternative to other mixture PDFs in describing wind regimes. Mixture PDFs which include GEV are able to provide close fit, particularly for high speed ranges of the wind spectrums. This is critical for wind speed applications as wind power is proportional to the cube of wind speed. Hence, mixture combinations of GEV with other conventional distributions need to be tried out for further analysis.

Abbreviations

α: shape parameter of gamma distribution (dimensionless); β: scale parameter of gamma distribution (m/s); χ^2: chi-square error; δ: scaleparameter of GEV distribution (m/s); Γ: gamma function; λ: mean of natural logarithm of wind speed (m/s); μ: mean of wind speed (m/s); Φ: standard deviation of natural logarithm of wind speed (m/s); ρ: air density (kg/m^3); σ: standard deviation of wind speed (m/s); ζ: shapeparameter of GEV distribution (dimensionless); *c*: scale parameter of Weibull (m/s); CDF: cumulative distribution function; *e*: GEV PDF; *E*: GEV CDF; *erf*: error function; $\hat{F}$: predicted cumulative probability at time stage i; $\bar{F}$: observed cumulative probability at time stage *i*; *f*: Weibull PDF; *F*: Weibull CDF; *ff*: two-component mixture Weibull PDF; *FF*: two-component mixture Weibull CDF; *fn*: theoretical PDF; *g*: gamma PDF; *G*: gamma CDF; GEV: generalized extreme value distribution; GEVL: mixture GEV and lognormal distribution; GW: mixture gamma and Weibull distribution; *h*: gamma-Weibull PDF; *H*: gamma-Weibull CDF; *k*: shape parameter of Weibull (dimensionless); K-S: Kolmogorov-Smirnov; *l*: location parameter; *ln*: lognormal PDF; *LN*: lognormal CDF; *LL*: logarithm of likelihood function; MLM: maximum likelihood method; *n*: number of data points; NN: mixture normal distribution; NW: mixture normal and Weibull distribution; PDF: probability density function; *q*: singly truncated normal PDF; *Q*: singly truncated normal CDF; *r*: normal-normal PDF; *R*: normal-normal CDF; R^2: coefficient of determination; RMSE: root mean square error; *s*: normal-Weibull PDF; *S*: normal-Weibull CDF; *t*: Weibull-GEV PDF; *T*: Weibull-GEV CDF; *u*: Weibull-lognormal PDF; *U*: Weibull-lognormal CDF; v_i: wind speed in time stage i; $\overline{v^3}$: mean of wind speed cubes; *v*: GEV-lognormal PDF; *V*: GEV-lognormal CDF; *w*: weight of mix parameter (dimensionless); W: Weibull distribution; WW: two-component Weibull distribution;

WL: mixture Weibull and lognormal distribution; WGEV: mixture Weibull and GEV distribution.

Authors' contributions
RK conceived the concept, performed the modeling, carried out the software simulation, and drafted the manuscript. KMP provided assistance for the software simulation and modeling. SRR checked the modeling analysis, verified the results, and helped in drafting the manuscript. SVLN reviewed the entire manuscript and offered technical help. All the authors read and approved the final manuscript.

Authors' information
RK is an assistant professor in the Electrical and Electronics Engineering Department, Jawaharlal Nehru Technological University, Kakinada, India. His areas of interest include distribution system planning and distributed generation. SRR is an associate professor of the same university. His areas of interest include electric power distribution systems and power systems operation and control. SVLN is a professor of Computer Science and Engineering in the School of IT, Jawaharlal Nehru Technological University, Hyderabad, India. His areas of interests include real time power system operation and control, IT applications in power utility companies, web technologies, and e-governance. KMP is a postgraduate student of the Department of Electrical and Electronics Engineering at JNTU Kakinada, India. His areas of interest include probabilistic DG planning and energy systems.

Author details
[1]Department of Electrical and Electronics Engineering, J.N.T. University Kakinada, Kakinada, Andhra Pradesh 533003, INDIA. [2]Computer Science and Engineering Department, School of Information Technology, J.N.T. University Hyderabad, Hyderabad, Andhra Pradesh 500085, INDIA.

References
1. Islam, MR, Saidur, R, Rahim, NA: Assessment of wind energy potentiality at kudat and Labuan, Malaysia using weibull distribution function. Energy **36**(2), 985–992 (2011)
2. Celik, AN: Energy output estimation for small-scale wind power generators using Weibull-representative wind data. J Wind Eng Ind Aerodyn **91**(5), 693–707 (2003)
3. Celik, AN: A statistical analysis of wind power density based on the weibull and Rayleigh models at the southern region of turkey. Renew Energy **29**, 593–604 (2003)
4. Akdağ, SA, Bagiorgas, HS, Mihalakakou, G: Use of two-component Weibull mixtures in the analysis of wind speed in the Eastern Mediterranean. Applied Energy **87**(8), 2566–2573 (2010)
5. Carta, JA, Ramı́rez, P: Analysis of two-component mixture weibull statistics for estimation of wind speed distributions. Renew Energy **32**(3), 518–531 (2007)
6. Luna, RE, Church, HW: Estimation of long term concentrations using a 'universal' wind speed distribution. J Appl Meteorol **13**(8), 910–916 (1974)
7. Garcia, A, Torres, JL, Prieto, E, De Francisco, A: Fitting wind speed distributions: a case study. Solar Energy **62**(2), 139–144 (1998)
8. Kiss, P, Jánosi, IM: Comprehensive empirical analysis of ERA-40 surface wind speed distribution over Europe. Energy Conversion and Management **49**, 2142–2151 (2008)
9. Embrechts, P, Klüppelberg, C, Mikosch, T.: Modelling Extremal Events for Insurance and Finance. Springer, New York (1997)

10. Kotz, S, Nadarajah, S: Extreme Value Distributions: Theory and Applications. World Scientific Publishing Company, Singapore (2001)
11. Ying, A, Pandey, MD: The *r* largest order statistics model for extreme wind speed estimation. J Wind Eng Ind Aerodyn **95**(3), 165–182 (2007)
12. Hyun Woo, P, Hoon, S: Parameter estimation of the generalized extreme value distribution for structural health monitoring. Probabilistic Engineering Mechanics **21**(4), 366–376 (2006)
13. Jaramillo, OA, Borja, MA: Wind speed analysis in La Ventosa, Mexico: a bimodal probability distribution case. Renew Energy **29**(10), 613–630 (2004)
14. Akpinar, S, Akpinar, EK: Estimation of wind energy potential using finite mixture distribution models. Energy Conversion Management **50**(4), 877–884 (2009)
15. Tian Pau, C: Estimation of wind energy potential using different probability density functions. Applied Energy **88**(5), 1848–1856 (2011)
16. Kececioglu, D: Reliability engineering handbook, vol.1 and 2. Destech Pubblications, Pennsylvania (2002)
17. Cheng, E, Yeung, C: Generalized extreme gust wind speeds distributions. J Wind Eng Ind Aerodyn **90**, 1657–1669 (2002)
18. Liu, F-J, Chen, P-H, Kuo, S-S, De-Chuan, S, Chang, T-P, Yu-Hua, Y, Lin, T-C: Wind characterization analysis incorporating genetic algorithm: a case study in Taiwan Strait. Energy **36**, 2611–2619 (2011)
19. Celik, AN, Makkawi, A, Muneer, T: Critical evaluation of wind speed frequency distribution functions. J. Renewable Sustainable Energy () (2010). doi:10.1063/1.3294127
20. Carta, JA, Ramírez, P, Velázquez, S: A review of wind speed probability distributions used in wind energy analysis case studies in the canary islands. Renew Sustain Energy Rev **13**(5), 933–955 (2009)
21. Sulaiman, MY, Akaak, AM, Wahab, MA, Zakaria, A, Sulaiman, ZA, Suradi, J: Wind characteristics of Oman. Energy **27**, 35–46 (2002)
22. Morgan, C, Matthew, L, Vogel, RM, Baise, LG, Eugene: Probability distributions for offshore wind speeds. Energy Conversion and Management **52**(1), 15–26 (2011)
23. Seyit, A, Akdag, Ali, D: A new method to estimate Weibull parameters for wind energy applications. Energy conversion and management **50**(7), 1761–1766 (2009)
24. National Data Buoy Center: http://www.ndbc.noaa.gov. Accessed on 10 October, 2011