2. Literature Review

Wind is a natural phenomenon that varies continuously with time and space. Wind resource assessment is the estimation of wind conditions based on several factors namely, available wind data, topographical, and meteorological conditions etc. at a given site that enables in effective power generation by the wind power plant during its service life. Scientific community treats the available wind data as a continuous random variable and therefore defines the wind statistics as a science that describes the patterns of the wind regime over considerable time period **[3]**. The prime motto of wind statistics is that it enables the correct assessment of wind flow pattern, which is essential for safe designing of mega-structures, and also enables in the extraction of maximum power from the kinetic energy of the wind. In the language of statistics, the wind flow pattern is unstable in the short term. However, it has a regular and stable pattern in the long-term **[3]**. The wind data behaves as a continuous random variable; therefore, the continuous probability distribution also known as the probability density function (pdf) is used to model the predictable pattern of the wind. The pdf of a continuous random variable describes the relative likelihood for this random variable to take on a given value **[5]**. Since the 1940s' till present, several kind of literatures have been published that use varieties of pdf to describe the wind speed frequency distributions. The continuous distributions which include Weibull, Rayleigh, Normal, Log-Normal, and Gamma distributions are most commonly employed to model unimodal wind speed data. However, recent advancement and availability of sophisticated computing system have motivated the researchers to fit bimodal (extra hump) and bi-tangential (extra bump) data using mixture distributions such as Weibull-Weibull, Truncated Normal-Weibull, and Gamma-Weibull distributions for wind speed analysis. These mixture distributions are mixed in different weight proportions to form mixture models.

Similarly, for wind direction analysis, von Mises, and wrapped Cauchy distributions are commonly used **[9]**. In the field of wind direction also, researchers **[10-14]** tried mixture distributions; the mixture of von Mises distribution has recently been used for wind directional analysis. Along with the pdfs to fit wind speed data, researchers **[15-24]** used Maximum Entropy Principle (*MEP*) for both wind speed and directional analysis. However, due to its certain limitations, *MEP* does not gain as much popularity as pdfs have in defining the wind data. The wind turbine characteristics, namely, *CF* and C_p are the two essential aspects need to be considered before installation of the wind turbine at a given site. Several researchers **[25-30]** proposed analytical models to estimate the *CF* of the wind turbine. In this present scope of work, a review of various continuous distributions that have been used to model heterogeneous as well as non-heterogeneous wind speed data, methods employed to estimate the parameters of these distributions, analytical models available to describe the wind turbine characteristics, extreme wind climate distributions, and various continuous distributions to model wind directional data have been studied.

2.1 Review of Continuous Distributions

2.1.1 Weibull distribution (W.pdf)

The 2-parameter Weibull distribution (W(k,s)) is a versatile distribution for wind speed data analysis. This distribution is helpful in describing the unimodal frequency distribution of wind speed data at many sites **[31].** The *W.pdf* is originally a 3-parameter distribution comprises of shape, scale, and location parameters. The location parameter is the lowest value of wind speed which can be equated to zero due to the presence of calm hours, and the 3-parameter Weibull distribution gets reduced to 2-parameter *W.pdf*. This 2-parameter *W.pdf* has been extensively employed to model wind speed data which enables to calculate wind power potential at a given site **[25, 32-53]**.

Weibull distribution is the flexible distribution and can be converted into other distributions such as Exponential, Rayleigh, Gaussian and Inverse Weibull distribution by varying its shape parameter. The $W(k,s)$ is an adjustable distribution and can be converted into another distribution by varying its parameters, such as $W(1,s)$ distribution is an exponential distribution with mean value '*s*'. The $W(2,s)$ distribution is a Rayleigh distribution with mean zero and variance '$s^2/2$'. In addition, the $W(3.6,s)$ is very close to the Gaussian distribution. **Stewart, and Essenwanger [54]** proposed 3-parameter *W.pdf* and found to be superior to 2-parameter *W.pdf* for estimating certain probability over thresholds. However, **Chadee, and Sharma [55]** pointed out that the added location parameter introduces difficulties in the estimation of its parameters and a positive value for it gives rise to an unrealistic condition of zero probability of wind speeds less than the parameter value.

The *W.pdf* has a wide variety of application in the estimation of wind power potential or to be more precise, in the statistical estimation of wind characteristics **[56].** The *W.pdf* enables in (a) wind speed data modelling. (b) estimation of variation of vertical wind characteristics **[57, 58],** (c) estimation of the *CF*, C_p, and output power of the wind turbines **[25-27, 59-68]**, (d) bivariate analysis of both wind direction and speed **[7, 16, 69-73]**, etc.

2.1.2 Rayleigh Distribution

Rayleigh distribution is a single parameter distribution and can be reckoned as a special case of *W.pdf*. The 2-parameters *W.pdf* can be converted into Rayleigh distribution by keeping the shape parameters of *W.pdf* to 2.

2.1.3 Gamma Distribution (G.pdf)

Gamma distribution is a 2-parameter distribution. The characteristic of this distribution is that its curve drops off much gradually than the *W.pdf* for shape parameters $k > 1$ and more quickly for $k < 1$. Therefore, the Gamma distribution also has the fair chance to fit the wind speed data and can act as a viable distribution to model wind speed data. This could be explained by the graphical representation as shown below:

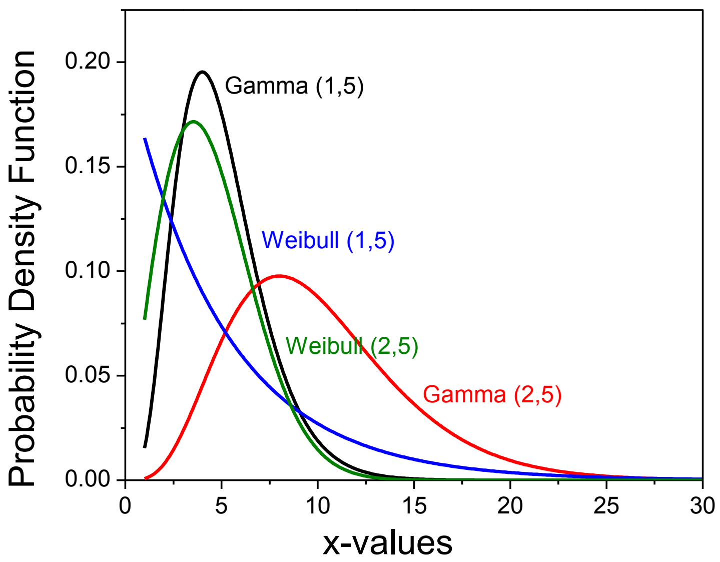

Figure 2.1: Comparison of Gamma and Weibull densities.

2.1.4 Normal Distribution/Gaussian and its Evolved Distributions (N.pdf)

The Normal distribution is a continuous distribution which is perfectly bell-shaped. All the measuring quantities included wind speeds available in nature are non-negative. The Normal distribution will not be going to work to model these empirical distributions, as the probability of negative value is greater than zero in the Normal distribution.

This axiom led the researchers **[5, 74]** to derive other distributions which evolved from the Normal distribution such as Lognormal distribution, Square root normal distribution etc. Further, some researchers **[16, 75]** truncate the normal distribution from its left limb to overcome such problems.

2.1.4.1 Log-Normal Distribution

If a random variable V follows the Lognormal distribution, then $Y = ln(V)$ follows the normal distribution, where $ln(V)$ is the natural logarithm to the base e. Likewise, if Y has a normal distribution, then the exponential function of Y, $V = exp(Y)$, follows the log-normal distribution.

2.1.4.2 Square Root Normal Distribution

If a random variable V follow square root distribution, then $Y = \sqrt{V}$ follows the normal distribution, Likewise, if Y follows the normal distribution, then the square of Y, $V = (Y)^2$, follows the square root normal distribution.

2.1.4.3 Truncated Normal Distribution

When the normal distribution is truncated either from the left or from the right limb that results in the formation of truncated normal distribution. For wind speed data modelling, the normal distribution is truncated from the left limb.

2.1.5 Mixture Distributions

Mixture distribution is a useful model to represent the heterogeneous wind data. The *W.pdf* is not sufficient to describe all the wind regimes encounter in nature. Therefore, several researchers **[7, 10, 13, 16, 69-73, 75-84]** continued to propose new mixture distributions of two-component mixed at different weight proportion for wind data analysis. These mixture distributions have an advantage that the number of parameters that constitute these distributions are five and can enable to fit any types of heterogeneous data.

2.1.5.1 Weibull-Weibull Distribution (WW.pdf)

Weibull-Weibull distribution (*WW.pdf*) constitutes of two *W.pdf*s that are mixed in different weight proportions. It can be expressed in closed form and considerably requires shorter duration for its calculation. **Jaramillo, and Borja [69]** used this distribution for the first time for wind speed data analysis of La Ventosa, Mexico. Sites selected have an extra hump in its frequency distribution. The monthly average data for the year 2000-2001 were taken for analyses. **Carta, and Ramirez [70]** discussed three widely known numerical methods, namely, Least Square Method (*LSM*), Method of Moments (*MOM*), and Maximum likelihood Method (*MLM*) to estimate the parameters of *WW.pdf*. They found that the *LSM* gives the highest degree of fit for all the four stations under consideration. However, there was no significant difference with the other two methods, and all methods were equivalently good at describing the wind power density distributions. **Akdag, *et al.* [71]** analysed the wind speed data of nine buoys, located in Ionian and Aegean Sea (Eastern Mediterranean) with *WW.pdf* and compared the same with conventional *W.pdf*.

Qin, *et al.* [72] proposed two new mixture *W.pdf*s that were based on the existing concept of mixing two distributions at different weight proportion. One new distribution was the mixture of 2, 3-parameter Weibull distribution [*mW(2, 3)*] and the other new distribution was the mixture of two component of 3, 3-parameter Weibull distribution [*mW(3, 3)*].

2.1.5.2 Truncated Normal-Weibull Distribution (TNW.pdf)

Truncated Normal-Weibull distribution constitute of Truncated Normal and Weibull distribution. The advantage of this distribution is that it can take into account the null wind speed. **Carta, and Ramirez [75]** analyzed the wind speed data of 16 locations of Canarian Archipelago that comprises of both unimodal and bimodal frequency distributions with *TNW.pdf* and compared the same with *WW.pdf* and *W.pdf*. The similar comparison was made by **[16, 73]** at six weather stations of Canary island and four locations of Elazig, Turkey respectively.

2.1.5.3 Gamma-Weibull Distribution (GW.pdf)

Gamma-Weibull constitutes of Gamma and Weibull distribution. The two distributions were earlier used independently to define the wind speed data. **Chang [7]** proposed this distribution along with another new distribution namely, mixture Truncated Normal distribution and were compared with *WW.*, *TNW.*, *MEP*. (constraint moment = 4) and *W.pdfs*. The hourly wind speed data recorded at three locations around Taiwan over the period of 2 years. They found that in estimating the wind speed distribution the proposed *GW.pdf* performed best followed by *TNW.pdf* and *WW.pdf*, while the *TNN.pdf* performs worst, while estimating the wind power density. The performance of *MEP* and *GW.pdf* was best, followed by the *WW.*, and *TNW.pdfs*.

2.1.6 Maximum Entropy Principle (MEP)

The principle of maximum entropy as stated by **[85, 86]** was based on the concept of entropy. **Shannon [87]** defined the *MEP* as when one was having some partial information about a random variate, one should choose that probability distribution for it, which was consistent with the given information, but has otherwise maximum uncertainty (entropy) associated with it. **Kapur [88]** demonstrated its use in the field of science and engineering. This principle lead to the generation of distributions that have further been utilized in the estimation of wind energy potential. **Li, and Li [18]** used the *MEP.pdf* (constraint moment = 3) to assess the wind power potential of Waterloo region. Wind characteristics such as monthly, annual, and diurnal behaviour were also examined. **Akpinar, and Akpinar [17], Li, and Li [24]** discussed the new family of *MEP.pdf* called as *MEP*-type distribution. The conventional *MEP.pdf* was modified by adding the wind speed to r^{th} power (where *r* varies from 0-5) as the pre-exponential term. The *MEP*-type distribution was compared with conventional *MEP* and *W.pdfs*. They found that *MEP*-type distribution was a suitable alternative for wind energy assessment. However, there was no fixed value of '*r'* that was applicable to all the locations under studies. Moreover, there were certain values of '*r*' that perform worst as compared to *W.pdf* for the assessment of wind power potential.

2.2. Comparative Study of Various Distributions for Wind Speed Assessment

Several unimodal continuous distributions are available to model the wind speed data. Among them, Weibull, Rayleigh, Gamma, and log-Normal are the most commonly used distributions. Among these, *W.pdf* is considered as the most favourable distribution to fit the wind speed data. The **International Standard IEC 61400-12 [6] and other International Agency [89]** recommend the 2-parameter *W.pdf* to model wind speed data for wind power generation.

Tar [57] compared Weibull, Rayleigh, Normal, Log-Normal, Square-Root Normal, and Gamma distributions and found that Rayleigh distribution to be the worst distribution to define the wind speed data across seven location of Hungary. **Zhou, *et al.* [90]** compared eight conventional distributions and eight *MEP*-derived distributions (constraint moment = 2-9) namely, Weibull, Inverse Gamma, Inverse Gaussian, Gumbel-maximum, Rayleigh, Lognormal, Erlang, Gamma distributions. For the analyses, long-term data of five locations of North Dakota were taken. They found that different goodness of fit statistics show different preferential order among the distributions under study for all the locations. However, performance of the Rayleigh distribution was worst on the criteria of goodness of fit. **Lo Brano, *et al.* [41]** compared seven distributions namely, Weibull, Rayleigh, Lognormal, Gamma, Inverse Gaussian, Pearson type V, and Burr distributions. They found that Burr distribution to be the most suitable for monthly data analysis of the urban area of Palermo, in the south of Italy. **Zamani, and Badri [53]** compared Weibull, Normal, Lognormal and Rayleigh distributions and found that Weibull to be the best distribution to describe the wind speed data. **Yim [91], Oner, *et al.* [92]** compared Weibull, Normal, and Rayleigh distribution and found that Weibull and Normal distribution performs equally good, whereas, Rayleigh distribution shows poor fit as compared to other two distributions for a given site; whereas, in 2^{nd} article Weibull distribution has been found to be superior than other two in accurately simulating the actual data. **Sohoni, *et al.* [93]** compared Weibull, Rayleigh, Gamma, Lognormal and inverse Gaussian and found that Weibull, Rayleigh, and Gamma, provide a better fit to the data as compared to the other two distributions. **Philippopoulos, *et al.* [45]** also compared the five distributions mentioned above and proposed that Gamma distribution was a suitable alternative to *W.pdf*.

Leite, and das Virgens [40] compared Weibull, Rayleigh, Beta, Gamma, and Normal distribution and found that *W.pdf* is the best distribution for fitting wind speed data at a given site. **Safari [48]** compared Weibull, Rayleigh, Gamma, Lognormal, and Normal distributions. They found that the suitable distribution is site dependent, Gamma and Weibull were two distributions that are most suitable for the selected sites. **Kiss, and Janosi [94]** compared Weibull, Rayleigh, Lognormal and generalized Gamma distributions. They found that generalized Gamma distribution proves to be an adequate and unified model for the description of all the sites under study. **Amaya-Martinez, *et al.* [34], Alavi, *et al.* [95]** compared Gamma, Lognormal, Rayleigh and Weibull for assessing wind potential. Several researchers **[25, 39, 42-44, 46, 47, 50, 52, 58, 60, 61, 96-112]** made comparison between the Weibull and Rayleigh distribution and found that *W.pdf* performed best to fit the measured wind speed data for the estimation of the wind energy potential. Their comparison was based on the different sites of the world for the different duration of time, whereas, very few **[111, 113]** found Rayleigh to be more suitable than the *W.pdf*. Table 2.1 shows the unimodal continuous distributions that have been frequently used to model wind speed data.

The expressions for probability density function and cumulative distribution function of various continuous distributions are given as:

Table 2.1: The common continuous distributions that are used to model wind speed data.

Sl. No.	Types of Distributions	
	Probability Density Function	**Cumulative Distribution Function**
1.	Weibull **[34, 40, 41, 45, 48, 53, 57, 90-94]** $f(v;k,s)=\left(\frac{k}{s}\right)\left(\frac{v}{s}\right)^{k-1}\exp\left[-\left(\frac{v}{s}\right)^{k}\right];\ v>0;k,\ s>0,$	$F(v;k,s)=1-\exp\left[-\left(\frac{v}{s}\right)^{k}\right]$
2.	Inverse Weibull **[114, 115]** $f(v;k,s)=\left(\frac{k}{s}\right)\left(\frac{v'}{s}\right)^{-(k+1)}\exp\left[-\left(\frac{v'}{s}\right)^{-k}\right];v'>0;k,\ s>0$	$F(v;k,s)=\exp\left[-\left(\frac{v'}{s}\right)^{-k}\right]$
3.	Normal **[40, 48, 53, 57, 91, 92]** $q(v;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\left(\frac{(v-\mu)^2}{2\sigma^2}\right)\right]$	$Q(v;\mu,\sigma)=\frac{1}{2}+\frac{1}{2}erf\left[\frac{v-\mu}{\sqrt{2}\sigma}\right]$
4.	Log Normal **[34, 41, 45, 48, 53, 57, 90, 93, 94]** $q(v;\mu,\sigma)=\frac{1}{v}\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\left(\frac{(\ln v-\mu)^2}{2\sigma^2}\right)\right];\ v>0$	$Q(v;\mu,\sigma)=\frac{1}{2}+\frac{1}{2}erf\left[\frac{\ln v-\mu}{\sqrt{2}\sigma}\right]$
5.	Square Root Normal **[57]** $q(v;\mu,\sigma)=\frac{1}{2\sqrt{v}}\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\left(\frac{(\sqrt{v}-\mu)^2}{2\sigma^2}\right)\right];\ v>0$	$Q(v;\mu,\sigma)=\frac{1}{2}+\frac{1}{2}erf\left[\frac{\sqrt{v}-\mu}{\sqrt{2}\sigma}\right]$

Table 2.1: Continued

6.	Truncated Normal **[75]** $$q(v;\mu,\sigma)=\frac{1}{I_0(\mu,\sigma)}\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\left(\frac{(v-\mu)^2}{2\sigma^2}\right)\right]$$	$$Q(v;\mu,\sigma)=\frac{1}{I_0(\mu,\sigma)}\left\{\frac{1}{2}+\frac{1}{2}erf\left[\frac{v-\mu}{\sqrt{2}\sigma}\right]\right\}$$
7.	Gamma **[34, 40, 41, 45, 48, 57, 90, 93, 94]** $$\begin{cases} g(v;\zeta,\beta)=\dfrac{v^{\zeta-1}}{\beta^{\zeta}\Gamma(\zeta)}\exp\left[-\dfrac{v}{\beta}\right]; & v>0;\ \varsigma,\beta>0 \\ 0; & v\le 0 \end{cases}$$	$$G(v;\zeta,\beta)=\int_0^v\frac{v^{\zeta-1}}{\beta^{\zeta}\Gamma(\zeta)}\exp\left[-\frac{v}{\beta}\right]dv$$
8.	Rayleigh **[34, 40, 41, 45, 48, 53, 57, 90-94]** $$f(v;s)=\frac{2v}{s^2}\exp\left[-\left(\frac{v}{s}\right)^2\right];\quad v>0$$	$$F(v;s)=1-\exp\left[\frac{-v^2}{2s^2}\right]$$

2.3 Comparison of the *MEP* Distribution with the Weibull Distribution

Ramirez, and Carta [19] applied the *MEP* (constraint moment = 4) to the wind speed data of three meteorology stations of Canarian Archipelago (Spain) and found that *MEP* could be used as an alternative to the *W.pdf*. **Chellali, *et al.* [23]** compared the *MEP.pdf* (constraint moment = 2-4; order of pre-exponential term = 4) with *W.pdf*. For comparison, hourly data of 2 years of six different sites of Algeria were selected. They found that *MEP.pdf* provides a better alternative than the *W.pdf*. However, it has certain drawbacks associated with it such as difficult to handle, optimality of configuration is a major challenge, and this distribution have been optimized for certain range of wind speed, beyond this range, it shows unpredictable behaviour.

Zhang, *et al.* [15] also checked the performance of *MEP.pdf* (constraint moment = 3, 4) and *W.pdf* in the intertidal zones. They found that *MEP.pdf* was more suitable than the *W.pdf* with the increase in height. The *MEP.pdf* with four constraint moment give least percentage error in estimating the wind power density. **Chang [20]** also compared *W.*, *WW.* and *MEP.pdfs* (constraint moment = 4). The sites selected were the three locations of Taiwan. The author found that *WW.* and *MEP.pdfs* were the useful alternatives to the W distribution in wind power assessment.

Table 2.2 shows the summary of the research carried out till-date comparison of mixture distribution, and distribution derived from *MEP*. The span of years at which researches have been carried out has also been displayed. In this table '*' mark shows the order of performances of the distribution, the more the '*' beside the abbreviation of distribution, the better is the distribution in describing the wind data of the given site.

Table 2.2: Summary of research carried out till-date in mixture modelling

Authors	Duration	Types of Probability Density Function					
Akpinar, and Akpinar [16]	8 years	*W.*	*WW.**	*TNW.***			*MEP.* (cm = 3)
Vicente [73]		*W.*	*WW.*	*TNW.*			
Chang [7]	2 years	*W.*	*WW.**	*TNW.***	*TNN.*	*GW.****	*MEP.**** (cm = 4)
Akdag, *et al.* [71]	1 year (3 hourly avg.)	*W.*	*WW.***				
Carta, and Ramirez [75]	16 years	*W.*	*WW.***	*TNW.***			
Jaramillo, and Borja [69]	1 year	*W.*	*WW.***				
Rajapaksha, and Perera [82]	1 year	*W.*	*WW.***	*(lnN-W.)**		*GW.*	

‘*’ order of performance

2.4. Methods for Estimating the Parameters of the Distributions

Numerous methods are available that can be used to determine the parameters of the distributions that enables in describing the characteristics of the wind. Practically, in the field of sustainable energy, researchers tried to estimate the single value of the parameters that can be used to define the entire sample data **[116]**. Several researchers **[117-133]** discussed various methods to estimate the parameters of the distribution. However, all these methods are solely based on three primary methods that are widely accepted and are used till-date. These three primary methods are discussed below:

2.4.1 Graphical / Least Square Method (LSM)

Graphical/*LSM* method selects those parameters of pdf as optimum, which enables in minimization of the sum of the squares of the difference between the observed cumulative relative frequency and theoretical cumulative distribution function (*CDF*). The Least Square Method is the only method that applied to the cumulative distribution function ($F(v,\theta)$) of the distribution to estimate its unknown parameters. The expression for minimization of non-linear an objective function $SSE(V_{max};\theta)$ under linear inequality constraints for wind speed data is given as:

$$MinSSE(V_{\max};\theta) = Min\left\{\sum_{i=1}^{N}[F_i - F(V_i;\theta)]^2\right\} \tag{2.1}$$

where vectors F_i and θ contain the empirical cumulative relative frequencies and the unknown parameters of the distribution function respectively. The $F(V_i;\theta)$ is the *CDF* of the theoretical distribution. The vector V_{max} is the maximum wind speed at each interval. For a unimodal distribution such as Weibull, Rayleigh, Lognormal, Gamma etc., It is very easy to linearize the *CDF* of the theoretical distribution and the unknown parameters can be estimated by minimizing the sum of the squares of the deviations between the linearized cumulative relative frequency and those obtained from the linearized *CDF* of the distribution. Perhaps, this is the oldest method to estimate the parameters of the pdfs that could be expressed in its closed form, namely, Weibull, Rayleigh etc. **[70, 77, 118-123, 128, 134-138]**. However, the problem arises with the mixture distributions which are used to model bimodal or bi-tangential data. In that case, the linearization of the *CDF* of distributions is difficult to obtain, as the logarithmic transformation of additive distribution is difficult to linearize. Therefore, the researchers **[70, 75]** used Lavenberg–Marquardt algorithm (*LMA*) to solve nonlinear equations.

2.4.2 Method of Moments (MOM)

Method of Moments is a very common statistical estimation technique to estimate parameters of any distribution. The idea consists in comparing the theoretical moments of a distribution to their empirical counterparts, the latter being only based on data $v_1, v_2, v_3, ..., v_n$. The number of moment equations should be equal to the numbers of parameters to be estimated. For example: for two parameter distributions, two statistical moments are required. Similarly, for five parameter mixture distributions, five statistical moments are required. The Quasi-Newton algorithm is commonly used to solve the equation of statistical moments **[12]** numerically. The advantage of this method lies in its simplicity to use. However, its limitation is that it does not consider the entire sample data **[70, 118-120, 122, 126-130, 132, 136, 139-150]**.

2.4.3 Maximum Likelihood Method (MLM)

The principle of Maximum Likelihood Method states that it calculates the values of the parameters that maximize the total probability of obtaining the particular sample. The likelihood of the sample is the product of all the individual probabilities **[23, 41, 78, 83, 117-119, 121, 122, 124, 131, 132, 136, 140, 141, 145, 151-155].**

The likelihood function *L* is:

$$L = \prod_{i=1}^{N} f_i(v;\theta) \tag{2.2}$$

Taking logarithmic will convert the product of all the individual probability into the summation of these probabilities.

$$\ln L = \sum_{i=1}^{N} \ln(f_i(v;\theta)) \tag{2.3}$$

To maximize the log-likelihood function, the differential of ln*L* with respect to θ is taken and equated it to zero, which has been used to estimate the values of the parameters that maximize the likelihood function. The mixture distributions contain the sum of logarithms of pdf. Therefore, elaborate techniques namely, the Expectation-Maximization (*EM*) algorithm, the Newton–Raphson method, and method of scoring are commonly used for its solution.

Based on above discussions three basic methods of the estimation of parameters, other researchers **[133, 139, 156, 157]** proposed few other methods (secondary methods) to estimate the Weibull parameters. The mathematical expressions for all these methods are discussed in details in chapter 4, section 4.2. Comparison between secondary as well as primary methods to estimate the Weibull parameters have been discussed in subsequent section. However, these secondary methods will work for *W.pdf* only. For other distributions, only the primary methods are applicable.

2.4.4 Comparison of Different Methods to Estimate Parameters of the Weibull Distribution

Several authors **[57, 133, 142, 156, 158-160]** discussed various methods to estimate the Weibull parameters. These methods were the Least Squares Method (*LSM*) also known as Graphical method, the Empirical Method (*EM*), the Method of Moments (*MOM*), the Power Density Method (*PDM*), the Maximum Likelihood Method (*MLM*) and the Modified Maximum Likelihood Method (*MMLM*). **Seguro, and Lambert [133]** compared *MLM*, *MMLM*, and *LSM* on a sample size of 72 and concluded that *MLM* was the best estimation method for the Weibull parameters based on the criterion of total energy output. They also found the *MMLM*, and *MLM* to be more accurate than the *LSM* regardless of the class width. However, **Cook [161]** discarded the claim made by **Seguro, and Lambert [133]** and concluded that *LSM* was equally good as *MLM* for estimating the Weibull parameters. **Akdag, and Dinler [139]** proposed a new method, the *PDM*, and compared it with *LSM*, *MLM*, and *MOM*. The sample data size taken for the analysis were less than 100,000. They stated that the *PDM* was an adequate method to estimate the Weibull parameters and that it might have better suitability than other methods. **Jowder [66]** compared the Empirical and Graphical Methods by taking monthly and annual wind speed data and determined the wind power density at 10, 30, and 60 *m* heights in the Kingdom of Bahrain. He found that the *EM* provides the more accurate prediction of average wind speed and power density than the Graphical Method. **Genc, *et al.* [136]** compared *LSM*, *MOM* and *MLM* based on 10,000 simulated samples generated in 10, 30, 50, 100, and 120 units. For estimation of $F(v)$ approximations plotting position, namely, $i/(n+1)$, and $(i\text{-}0.3)/(n+0.4)$ were used and compared. They found that *MLM* with $F(v)$ $(i\text{-}0.3)/(n+0.4)$ was more sensitive technique for estimating the Weibull parameters.

Chang [118] checked the accuracy of all the methods mentioned above on the simulated as well as observed data with sample size 10,000. He found that *MLM* performs best on both simulated as well as observed data and *LSM* performs worst based on the criteria of root mean squared error and the Kolmogorov-Smirnov test. However, the simulated data taken for the comparison had variation in seed values, and an average value of the parameters were taken for the analysis. **Akdag, and Guler [126]** proposed a novel energy pattern factor for estimating the 2-parameter *W.pdf*. However, this method is only a simplification of the *PDM* or Energy pattern factor method (E_{pf}) with the pre-determined coefficient. Wrapping up, from all the methods discussed in the literature to estimate the parameters of the 2-parameter *W.pdf*, not a single method stands out as being the universally accepted best one.

A standard method for Weibull parameter estimation used in industries is the Wind Atlas Analysis and Application Program (*WAsP*) developed by the Department of Wind Energy at the Denmark Technical University earlier known as Riso National Laboratory of Denmark. Various researchers compared this tool with statistical methods, for instance, **Bagiorgas, *et al.* [162]** compared the *WAsP* algorithm used in the WindoGrapher software with four statistical methods to estimate Weibull parameters, namely *LSM*, *MOM*, *MLM* and an alternative Maximum Likelihood Method. They found that the *MOM* and *MLM* produce the same results while only shape parameter is concerned, however, for the scale parameter all the methods produce almost identical results. **Bénédicte [163]** compared *MLM* and *MOM* (based on the first and third Moments) to the *WAsP* method based on the error in the production of real wind farms. They found that the *MOM* strongly agrees with the *WAsP* method in estimating the Weibull parameters.

Furthermore, the *WAsP* was employed to simulate the wind climate by reverse modelling using wind data **[164, 165]**. However, the methods employed by them to estimate the Weibull parameters for the wind energy potential are respectively the *MOM* and *LSM*. *WAsP* programming has also been commonly used for wind energy resources assessment **[166-168]**

Table 2.3: Summary of the research up till now on methods employed in parameters estimation of *W.pdf*.

Sl. No.	Methods Employed to Estimate Weibull Parameters
1.	Least Square Method **[66, 117-119, 122, 124, 133, 136, 139, 142, 153, 157, 161, 162]**
2.	Method of Moments **[117-119, 122, 124, 127, 136, 139, 142, 150, 162]**
3.	Empirical Method **[66, 117-119, 122, 150, 157]**
4.	Power Density Method **[117-119, 122, 124, 139, 150]**
5.	Maximum Likelihood Method **[117-119, 122, 124, 127, 133, 136, 139, 150, 153, 161, 162]**
6.	Modified Maximum Likelihood Method **[117-119, 122, 124, 133, 161]**
7.	Alternative Maximum Likelihood Method **[157, 162]**
8.	Wind Atlas Analysis and Application Program **[153, 161, 162]**

From above discussions, it has been clear that a good estimating method should satisfy the following criteria:

(i) The method should be accurate for estimating the Weibull parameters.

(ii) The formulation of the method should be simple and explicit.

(iii) The solution should be efficient in terms of calculations and preferably does not involve any iterative procedure.

(iv) The method should be free from the binning problem, as too high bin size results in information loss, and too low bin size results in sampling errors.

In this study, the performance of all the six statistical methods has been compared based on these four criteria. Table 2.4 shows the comparative assessment of various methods to estimate Weibull parameters with respect to each criterion discussed above in a better way.

Table 2.4: Comparative assessment of various methods to estimate Weibull parameters on four different criteria.

Sl. No.	Methods	Whether accurate?	Whether explicit formulation?	Whether efficient for calculations?	Whether free from binning?
1	LSM	Yes	Yes	Yes	No
2	MLM	Yes	No	No (Involves many iterations)	No
3	MOM	Yes	No	No	Yes
4	MMLM	Yes	No	No	No
5	EM	No (Approximate Method)	Yes	Yes	Yes
6	PDM	Yes	Yes	Yes	Yes

The criterion (i) may be considered as an essential pre-requisite, though not sufficient, for selecting the method. It has been observed from Table 2.4 that only empirical method does not satisfy criterion (i). However, from the table, it has been concluded that the other methods namely, least square method, maximum likelihood method, a method of moments, and modified maximum likelihood method do not satisfy criteria (ii), (iii) and (iv) simultaneously, though they satisfy criterion (i). Only the Power Density Method (PDM) satisfies all the criteria. The PDM was proposed by Akdag and Dinler [139] who did not claim that the method is more accurate than other conventional methods. However, they claimed that it satisfies the criteria (ii), (iii) and (iv) along with its comparable accuracy with other methods. Hence, from Table 2.4 the Power Density Method (PDM) can be considered as the best competitor of our proposed method followed by the Least Square Method (LSM) though the accuracy of the latter is highly influenced by binning, and there is no explicit method available in the literature to determine the size of the bin.

2.5 Wind Turbine Characteristics

Chang, *et al.* [28] shows the importance of wind turbine characteristics for the selection of wind turbine at a given site. They showed that *CF* and the wind turbine efficiency, also known as Power Coefficient of a specific wind turbine, are equally important to define the wind turbine characteristics completely, whereas, Availability Factor is an important characteristic to define the power plant efficacy.

The characteristics of the wind turbine primarily depend upon the turbine power curve. This turbine power curve is designed by the manufacturer and solely depends on three parameters, i.e., cut-in (V_c), rated (V_r) and cut-out (V_f) wind speeds, and therefore, specifically depend upon the type of turbine used at a particular site which is difficult to be generalized. However, the entire power curve follows two mechanisms; namely, pitch controlled and stall controlled mechanism. In the pitch controlled wind turbine, wind speeds between V_r and V_f are fairly constant whereas, in stall controlled wind turbines wind speeds between V_r and V_f are of variable types. However, wind speeds between V_c and V_r are of continuously increasing character type.

Several researchers **[25-30, 59-63, 69, 169-171]** attempted to represent the ascending segment of the turbine power curve using the general model. By the general model, the authors mean that the segment of turbine power output, as the percentage of rated power output that is described using V_c and V_r only, without the prior knowledge of turbine output throughout the ascending segment. The general model available in the literature along with other models are given in Table 2.4:

Table 2.5: The different analytical models to predict normalized power curve.

Linear [30]	Quadratic [26]	Cubic [63]	General [25, 59-62]
$P_n(v)=\frac{v-V_c}{V_r-V_c}$	$P_n(v)=\frac{v^2-V_c^2}{V_r^2-V_c^2}$	$P_n(v)=\frac{v^3}{V_r^3}$	$P_n(v)=\frac{v^k-V_c^k}{V_r^k-V_c^k}$

However, several other researchers **[28, 29]** proposed the polynomial function of order *n* to regress the ascending segment of the manufacturer provided power curve data points. The order and values of the polynomial constant have been solely dependent upon the type of wind turbine selected for a particular site. Table 2.5 shows the commercially available wind turbines and the order of best-fitted polynomial function with its constant values.

Table 2.6: Types of the commercially available turbine and the analytical models to express the normalized power.

Type of Turbines	Constant Value $P_n(v) = a_0+a_1v+a_2v^2+a_3v^3\ldots a_6v^6$
Vestas V47–660 kW **[28]** $P_n(v)=\sum_{i=0}^{4}a_iv^i$	a_0 = 0.39209; a_1 = -0.24764; a_2 = 0.04446; a_3 = -0.00169
Enercon53-800kW **[29]** $P_n(v)=\sum_{i=0}^{6}a_iv^i$	a_0 = 1.5771e-01; a_1 = -2.8988e-01; a_2 = 1.7668e-01; a_3 = -4.8053e-02; a_4 = 6.6426e-03; a_5 = -4.2704e-04; a_6 = 1.0185e-05.
Enercon 101-3000kW **[29]** $P_n(v)=\sum_{i=0}^{6}a_iv^i$	a_0 = -3.3450e-02; a_1 = 7.0904e-02; a_2 = -5.3329e-02; a_3 = 1.8260e-02; a_4 = -2.8726e-03; a_5 = 2.3246e-04; a_6 = -7.3596e-06
AAER-A-1000kW **[29]** $P_n(v)=\sum_{i=0}^{6}a_iv^i$	a_0 = -4.0530e-02; a_1 = 8.3688e-02; a_2 = -5.8759e-02; a_3 = 1.7661e-02; a_4 = -2.3803e-03; a_5 = 1.6461e-04; a_6 = -4.5615e-06.

Table 2.5: Continued

GE 15XLE **[29]** $$P_n(v)=\sum_{i=0}^{6}a_iv^i$$	*a_0 = 7.7273e-03; a_1 = -2.0437e-03; a_2 = -9.3940e-03; a_3 = 4.8722e-03; a_4 = -6.3600e-04; a_5 = 5.3199e-05; a_6 = -2.0425e-06.*
GE36SL **[29]** $$P_n(v)=\sum_{i=0}^{6}a_iv^i$$	*a_0 = 7.1937e-02; a_1 = -1.2362e-01; a_2 = 6.8510e-02; a_3 = -1.7299e-02; a_4 = 2.3291e-03; a_5 = -1.4201e-04; a_6 = 3.1326e-06*
Vestas90-3000kW **[29]** $$P_n(v)=\sum_{i=0}^{6}a_iv^i$$	*a_0 = 4.0821e-02; a_1 = -5.6299e-02; a_2 = 2.0951e-02; a_3 = -3.1933e-03; a_4 = 4.0487e-04; a_5 = -2.2914e-05; a_6 = 4.1853e-07.*
Generic model **[169, 170]** $$P_n(v)=\sum_{i=0}^{n}a_iv^i$$	
Pitch controlled turbine (V_i = 4, V_r = 15; V_f = 25) **[69]** $$P_n(v)=a_2+\frac{a_2-a_1}{1+\exp\left(v-a_3/a_4\right)}$$	*a_1 = -0.035; a_2 = 1.0; a_3 = 9.00; a_4 = 1.50.*
Pitch controlled (V_i = 4.5, V_r = 12; V_f = 25) **[171]** $$P_n(v)=a_1+\exp\left(\frac{-(v-a_2)}{a_3}\right)$$	a_1 = 1.00; a_2 = 8.70; a_3 = 1.15.
For Stall controlled turbine (V_i = 4, V_r = 14; V_f = 25) **[171]** $$P_n(v)=a_1+a_2-a_1+10^{(a_6(v-a_4))}+a_3-a_1+10^{(a_7(a_6-v))}$$	*a_1 = −112.77; a_2 = −0.05; a_3 = 0.82; a_4 = 14.00; a_5 = 14.00; a_6 = 0.19; a_7 = 0.19.*

- $P_n(v)$ is the normalized power output throughout the ascending segment.

Since the two types of above-mentioned models have certain advantages as well as certain limitations too; the generic models, on the one hand, can approximately fit the manufacturer provided power curve data. The polynomial function, on the other hand, gives accurate fitting. However, the order of constant and values of constant are solely dependent upon the types of turbine used. The normalized power output has been used to estimate the *CF* and C_p of the wind turbine. The expressions for *CF* and C_p have been discussed in chapter 1, section 1.4.2 and 1.4.3 respectively.

2.6 Extreme Wind Climate Modelling

Extreme wind climate modelling is used to specify the design wind speed that enables in the calculation of design wind load for different structures. Extreme value theory was firstly proposed by **Gumbel [172]**. Extreme value limit distributions can be classified into three types.

Type-I Gumbel distribution
Type-II Fréchet distribution
Type III- Weibull distribution
The mathematical expressions for the extreme value distribution have been discussed in chapter 7 of the current work. The distributions of types II and III each form a family of curves with specific characters. Compared to the type I distribution, they have a special feature, i.e., a certain curvature which can best be seen in a plot on Gumbel probability paper. A probability paper is a graph paper with vertical axis specially ruled to transform the distribution function to a straight line when it is plotted against the variate as abscissa. While the curves for the type II distribution bend in a concave shape in respect to the axis of the reduced variate in Gumbel probability paper, the curves corresponding to type III show a distinct convex character (see Figure 2.2). These two types are separated from the type I distribution which appears in the plot as a straight line **[173]**.

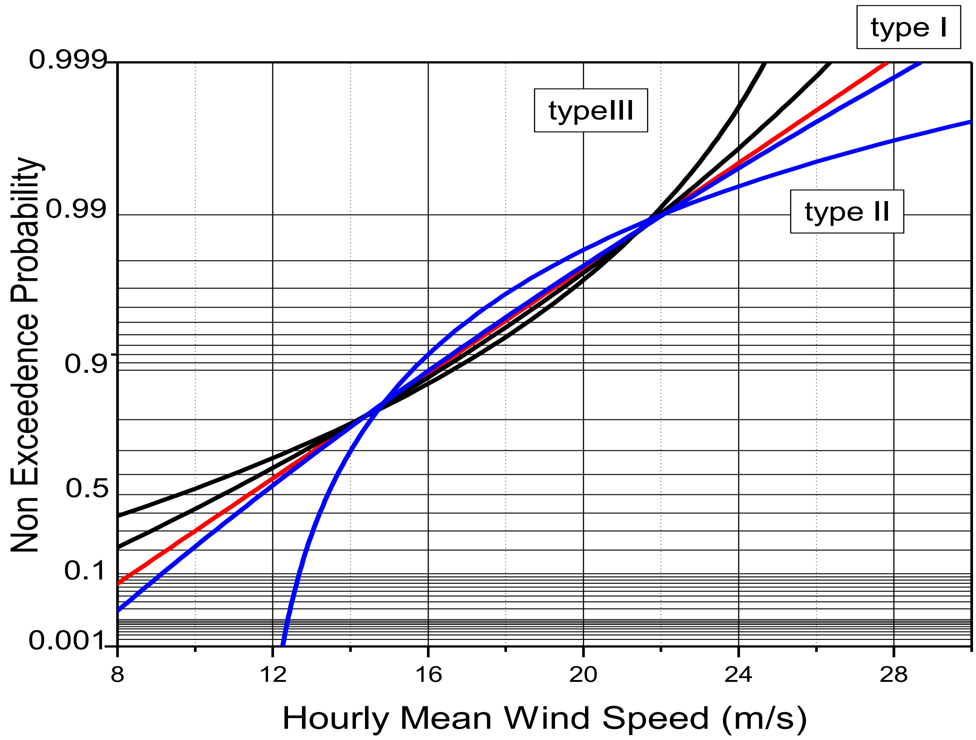

Figure 2.2: Traces of the extreme value limit distributions in a Gumbel probability paper **[173]**.

2.6.1 Selection Criteria for Appropriate Types of Extreme Value Distribution

Fisher [174] showed that a large sample of maxima or minima of the random variable is approximately distributed as a generalized extreme value distribution. However, the appropriate choice of types of *EVD* for extreme wind speed modelling is a challenging task. The wind engineers did not yet reach a common viewpoint on *EVD* models, i.e., most suitable distribution to predict V_d. The discrepancy between the researchers increases progressively as the sample taken for *EVA* is small in size or the or the duration of the return period becomes high **[175, 176]**. Several researchers **[172, 177, 178]** state that the extremes were best described by the type I, or Gumbel distribution, which has an infinite upper tail. The U.S. wind map also uses the *GEV*-type I distribution. Indian standard **[179]** also fits the gust wind speed data by a type I distribution to determine the design wind speed for a target return period. **Cook [180]** state that the *GEV*-type III (Reverse Weibull) distribution, which has a finite upper tail is inappropriate and that its best fit is due to the unavailability of sufficiently large data sets. However, earlier **Peterka [181]** found *GEV*-type III to be appropriate for describing extreme data. **Simiu, and Heckert [182]** also suggested that reverse Weibull distribution was more appropriate than Gumbel distribution to fit the extreme wind speed data. **Sarkar, *et al.* [183]** also employed *POT* approach and *EVD* to analyze gust wind speed data for three locations of India, namely Calcutta (Kolkata), Lucknow, and New Delhi. They found the Reverse Weibull and Fréchet distribution to be more suitable for these locations. **Lakshmanan, *et al.* [184]** fit maximum daily gust wind speed data of various locations in six wind zones of India by Fisher-Tippett type I distribution which was similar to Gumbel distribution. The Indian standard **[179]** also fits the type I distribution to the daily maximum gust wind speed data that may lead to the chances of underestimation or overestimation of design wind speed. Thus, it is important to verify the type of distribution suitable for Indian extremes.

2.7 Wind Directional Distribution

Wind directional assessment is also an essential aspect of planning, designing, and development of the wind farm. In fact, prior information about the direction of the wind enables the wind turbine installer to install the wind turbine in such a direction so that the total amount of energy captured will be maximized, that enables in the maximization of the performance of the wind farm. However, very less attention is paid towards this aspect of wind regime due to the complexity involved in circular data analyses. Some researchers **[9, 185]** proposed distributions to describe the circular data, namely, circular uniform, von Mises, wrapped-normal, and wrapped-Cauchy distributions. Among them, von Mises distribution proved to be the most prominent in describing the circular data. The brief description and distribution based on von Mises distribution have been discussed as:

2.7.1 von Mises Distribution (vM)

von Mises [185] introduced a pdf whose total probability is focused on the circumferential periphery of a unit circle. **Gumbel [186]** emphasized its similarities with the Normal distribution. The mathematical expressions for von Mises distribution have been discussed in chapter 8.

2.7.2 Mixture von Mises Distribution (mvM.pdf)

Mixture von Mises distribution comprises of two or more von Mises distributions. Several researchers **[10, 13]** mixed von Mises distribution in different weight proportion, i.e., they treated each von Mises distribution as one component and mixed two or more component to form mixture von Mises distribution. This model is particularly useful to fit wind direction data. The *mvM.pdf* comprises of a weighted sum of N von Mises probability densities $vM\ (\theta;\kappa_j,\mu_j)$.

Satari, *et al.* [187] used the von Mises distribution to fit the the circular data of Malaysia. They found that wind direction was mostly affected by the monsoon season, namely, southwest and northeast monsoon season. The wind direction has no influence on the wind speed in Peninsular Malaysia regardless from where it may blow. **Kamisan, *et al.* [188]** compared four types of circular distributions namely, circular uniform distribution, von Mises distribution, wrapped-normal distribution and wrapped-Cauchy distribution. The site selected to model the directional wind data were the regions of Malaysia. They found that the southwesterly monsoon wind direction of Malaysia was best described using von Mises distribution. **Williams, *et al.* [189]** also found the von Mises circular distribution best described the embedding dispersion of pollution source.

Carta, *et al.* [10] compared the Smith distribution with mixture von Mises distribution for two locations of the Canarian Archipelago; they found that finite mixtures composed of two von Mises distribution provide the better fit than the Smith distribution. However, with the increase in the number of modes (prevailing wind direction), the number of component of mixture von Mises has to be increased, to achieve the better fit. But, the maximum number of component of mixture von Mises distribution should not be greater than six. **Masseran, *et al.* [13]** compared the single von Mises distribution with the mixture von Mises distribution with different weight component (*N* = *1, 2, 3…8*) for the nine locations of Malaysia. They had also concluded the equivalent results as found by **Carta, *et al.* [10]** that with the increase in the number of modes the acceptability of the mixture distribution increases. They also concluded that *N* = *6, 7, 8* gives the similar best fit to the observed distribution. However, for some sites, *N* less than six was also preferable and it was not worthy to go beyond *N* greater than six, as there is the only marginal increase in R^2. Lately, **Masseran, and Razali [190]** classified the data into two seasons, namely, southwest monsoon and northeast monsoon and then again applied the finite mixture of von Mises distribution. They found that on classification the data into seasons, the number of components to define data was reduced to *N* = *2, 3* rather than *6* for all sites.

2.8 Research Gaps

In this study, various probability density functions that are available up till now to model unimodal wind speed, as well as, direction data; the methods adopted to estimate the parameter of these distributions; various analytical models to define the wind turbine characteristics have been discussed. Various mixture distributions and distribution based on maximum entropy principle have also been studied. From the above discussion, following research gaps have been made and mentioned below:

1. There are only three basic methods available that are discussed by most of the researchers to estimate the parameters of any distribution; these are Least Square Method, Method of Moments, and Maximum Likelihood Method. All three methods require an iterative technique for their solution and among them two are having the problem of binning. Therefore, for *W.pdf*, based on these three primary methods other methods have been deduced. However, the secondary methods deduced from these primary methods are an approximate solution for the estimation of Weibull parameters. Therefore, still, this area of research is wide open, and new methods are still required that are simple to use, robust, accurate, and free from binning for the estimation of Weibull parameters.
2. For unimodal data modelling, Weibull is the most popular distribution. However, one cannot be judgmental about its universal acceptance due to its certain limitations. Other distributions can also fit the given data better as compared to *W.pdf*. The better approach is that for a given wind speed data, all the frequently used continuous distributions should be tried and most suitable distribution should be selected.

3. For heterogeneous data such as those with bimodal or bi-tangential distribution, the mixture of unimodal distribution is the suitable alternative to fit the wind speed data. Three commonly employed mixture distributions are *WW.pdf*, *TNW.pdf*, and *GW.pdf*. However, these distributions are not fitted in long-term data by any researchers. Therefore, their applicability to long-term data need to be checked and verified. Moreover, other mixture distributions can also be constituted from unimodal distribution by mixing at different weight proportion. The appropriateness of most suitable mixture distributions in the Indian context is a subject of ample research.
4. Due to the complexity involved in the solution of wind directional data, researchers have rarely studied the parameterization of wind direction data and its validity with the wind flow behaviour. Therefore, there exists wide scope of research in the wind directional analysis.
5. Majority of the researchers have confined their work by merely fitting the wind speed data with the various simulated models. However, the practical applicability of the estimated parameters from wind speed data has rarely been studied. Therefore, there is the broad scope of research in the practical application of parametric study, especially, in the field of energy production.
6. Most of the research was carried out at the locations where wind climatic conditions are stable throughout the year, unlike Indian subcontinent which has varied wind climatic condition. Therefore, the application of all distributions in the Indian context is of utmost importance.

2.9 Objectives

Motivated from the above discussions, the authors found an ample scope of research in the field of wind resources assessment. The present work brings out the following objectives in view of the current research study:

1. The wind data supplied by IMD, Pune is in integer *km/h*. However, they record the wind speed in *knots*. The conversion of wind speed data from *knots* to integer lead to the generation of errors. The presence of these errors in wind speed data result in the biased estimation of parameters. Therefore, unbiased parameter estimation is the challenging task before statistical modelling of wind speed data.
2. Weibull is the most popular distribution, its applicability in Indian subcontinent wherein wind climate is of variable nature, needs to be verified and rechecked. Moreover, the Indian subcontinent has large geographical boundaries where climate conditions are not same at all locations for the same duration of time. Therefore, the Weibull parameters for several sites need to be studied, so that it can be used for practical applications.
3. The methods adopted to estimate the parameters of *W.pdf* is still a matter of research, as there is no single consensus available on methods to be used to estimate the parameters of the distribution.
4. The appropriate selection of speed parameters of the wind turbine that can optimized wind turbine characteristics is a matter of research so that the efficiency of the wind turbine can be maximized.
5. Exploring new distributions that can replace conventional *W-pdf* need further studies.
6. The wind directional modelling has rarely been done, especially, for Indian subcontinent. Therefore need detail studies.

2.10 Contribution of the Current Research Work

In view of the objectives, the present work focuses on the following points:

1. Estimation of an unbiased Weibull parameter that is free from all types of errors
2. Identification of the best method to estimate the parameters of the *W.pdf*
3. Identification of optimum speed parameters of the wind turbine that enables in maximizing the wind energy output
4. Identification of the most suitable distribution amongst mixture distributions and the distribution derived based on maximum entropy principle for fitting the wind speed data, especially, for Indian subcontinent.
5. Identification of design wind speed for tall towers that can withstand wind load
6. Identification of directional distributions, suitable to fit wind direction data and to specify probable zone of wind speed flow using wind rose diagrams