# 7. A Novel Approach for Accurate Assessment of Design Wind Speed for Variable Wind Climate

The ideal locations for the installation of the wind farm are the onshore or offshore sites of any country, including India. India has two ideal zones for the installation of the wind farm these are eastern zone surrounded by the Bay of Bengal, and western zone surrounded by the Arabian Sea. However, the coastal regions are often vulnerable to storms, and the structures located at these locations are subjected to structural damages due to the oncoming of strong winds. These strong winds, inherently having the high turbulence intensity, may cause forced rupture to the structures. The records of the damage indicate that even today structures are not adequately robust against the wind-induced hazards. Therefore, appropriate design methodologies are the prerequisite to ensure the structural safety.

Wind load is directly proportional to the square of the wind speed, and hence the specification of the design wind speed ($V_d$) can be considered as the essential requirement for the stipulation of design wind load. *W.pdf* is the most popular distribution because as wind speeds follow the *W.pdf*, then square of wind speeds also follows the same distribution **[96]**. However, **Castillo [215]** remarked the surprising result about *W.pdf* is that extreme (maxima/minima) sampled from parent distribution including Weibull distribution does not converge on a Weibull distribution as sample size approaches infinity. They suggested *EVD* for modelling extreme sampled form Parent distribution, provided the sample size of the parent distribution approaches towards infinity. Hence, they did not suggest the *W.pdf* to fit a sample of maxima or peaks over a threshold. There should be a threshold wind speed up till which *W.pdf* can be considered as a suitable one and beyond which wind speed data need to be Modelled by extreme value distributions (*EVD*), viz., type I, type II, and type III or Generalized Pareto Distribution.

Since extreme events are the rare events, their non-stationarity should be explained **[216]**. However, after sampling these extreme events through block maxima (*BM*) or peaks over threshold (*POT*) approach, extreme events become independent and identically distributed data (*i.i.d.*) and can be Modelled by different *EVD*. But since wind speed data are highly non-Gaussian, the observed cumulative probability should be properly estimated by the suitable choice of plotting position methods **[217]**.

This study proposes the novel approach for the specification of the $V_d$ of the structures. The specification of the $V_d$ can be achieved by using extreme wind climate modelling. Extreme data was firstly Modelled by **Gumbel [172]** using *EVD*. These distributions are the limiting distributions of extreme (maxima/minima) values and mainly focus on the asymptotic tail of the distribution in the sense that extreme values have larger probability than usual. The *EVD* is classified into three types, viz., type I (Gumbel distribution), type II (Fréchet distribution), and type III (Reverse Weibull distribution). These distributions enable an estimation of the $V_d$ for a given return period. Depending on the classes of the structures the return period is specified by the user. The important structures such as satellite communication towers and nuclear power reactors etc. require a high return period (low-risk level). Therefore, for the better service life of the structures, the high return period is the primary requirement. In this study, the return period of 1000 years has been considered so that the annual exceedance probability becomes 1/1000 and the corresponding non-exceedance probability becomes 0.999.

## 7.1 Selection Criteria for Appropriate Choice of Types of Wind Speed Data

The selection of appropriate type of wind speed data for the specification of $V_d$ is of utmost importance. The type of wind speed data may be the gust wind speed, 10-min mean wind speed or hourly mean wind speed. The proper choice depends on the ability to represent best the intensity or strength of a storm **[218]**. The locations, considered in this study, are situated in the coastal regions of India and they are mainly influenced by cyclones and monsoon gales. For these storm types, both hourly mean and maximum gust wind speeds are able to indicate the intensity of a single storm event. However, the purpose of taking hourly mean wind speed data for this study rather than daily maximum gust wind speed is due to the reason that the structures such as tall buildings, chimneys, latticed towers, cooling towers, turbine tower, etc. are vulnerable to wind-induced oscillations which are the major causes for their damages. The gusts cause an increase in the air pressure, but their effect on the stability of the building may not always cause rupture to the structures. The inertia of the building can take care of short period gusts which may not cause any appreciable increase in stress in main components of the building though it can cause damages to its accessories **[179]**. Therefore, this study focuses on the response of the structures to the dynamic velocity fluctuations i.e., along wind loads on the structures. These along wind loads are calculated using Gust Factor method. Since in this method, the expression for the dynamic wind load on the structure already contains a term like the gust factor G; $V_d$ can best be represented by the hourly mean wind speed considering the generalized extreme value distribution which is a heavy right tail distribution. Therefore, the hourly mean wind speed with very high return period (1000 years) have been taken to estimate design hourly mean wind speed that can ensure structural safety as well as economic viability.

## 7.2 Methods Employed for Identification of Extreme Data

Several researchers **[172, 182, 219-221]** also discussed the statistical methods employed to identify the extreme wind, such as Standard Gumbel (Annual Maxima/ Block Maxima), Modified Gumbel, Method of Independent Storms (*MIS*), and Peaks over Threshold (*POT*). For identifying the extremes the two most common methods employed are the Block maxima (*BM*) and *POT*. These two approaches have certain advantages as well as certain limitations. **Ferreira, and de. [222]** discussed the merits of *BM* method and found this method still works even when the maxima are not *i.i.d*. However, the choice of the block is a challenging task, as too low block size leads to biased estimates, whereas, too high block size results in large estimation variance **[217, 223]**. For getting the reliable result, the block size of 1 year was suggested by **Cook [224]** when the available record is more than 20 years. However, the disadvantage associated with the annual maxima is that the second peaks of one year may be greater than the maxima of another year leading to the false estimation of the $V_d$ **[175, 182, 207]**. Another common method employed for *EVA* is *POT*. The *POT* has an advantage that it is flexible and focuses on the tail of the *EVD*. The size of data records above the threshold is high in numbers. However, the disadvantage of this method is that the data over threshold appears to be correlated.

In this chapter, the new approach has been adopted that take the advantages of these two approaches and avoids their demerits. The reference sites selected to carry out this study is same as that selected for the analysis of wind power potential (see Table 4.1). The entire observation period has been classified into blocks of months, thereby making the total block of 12 in number, instead of taking the block of 1 year.

The purpose of taking block on a monthly basis is that the major extreme events in India occur during some specified months, i.e., when south-westerly monsoon prevails in India. This is the period of occurring of several extreme events.

Therefore, estimating $V_d$ on a monthly basis gives the correct prediction about it rather than on an annual basis in which there are chances that the several low values qualify as extreme values. In each block, *POT* method has been applied to get several maxima. The purpose of practicing such an exercise will have two advantages firstly; no maxima will be left out, and secondly, the data collected by this process are *i.i.d*. Taking month wise data for analysis has an additional advantage too, that it can take into account the varied wind climate of the location as well.

The chapter has been carried out in the following manner. Initially, the average numbers of observations above reference wind speed ($V_{ref}$) that do not follow the *W.pdf* have been graphically demonstrated. Secondly, the extent to which the wind speed data follow the *W.pdf* has been determined, and it has been graphically represented that the supplied data are nonstationary and follow the non-gaussian distribution. Thirdly, the entire observed data have been classified into an equal block of 12 for 12 months. Within each block, an appropriate threshold value has been selected, and entire wind speed data above the threshold value have been fitted by an appropriate extreme value distribution. Fourthly, based on a monthly basis the $V_d$ has been estimated for 1000 years return period with identifying the month that shows highest $V_d$ for three stations Lastly, comparisons of different plotting position formulae based on the criteria of $R^2$ and *RMSE* have been carried out.

## 7.3 Mathematical Model

### *7.3.1 Methodology Adopted for the Determination of Appropriate Threshold: Weibull Quantile Function*

**Fisher [174]** suggested that if the sampled data followed *W.pdf*, then its maximum follows *EVD*. Considering this as the axiom the threshold wind speed up till which *W.pdf* is valid has been determined. As the supplied wind data is a mixture of a parent as well as extreme data, the *W.pdf* is not a suitable candidate to fit the entire wind speed data. Hence, a method is required to detect the threshold beyond which the extreme values can be considered. This can be achieved by following the method suggested by **Tukey [193]** which is based on the quantiles, (see section 4.2.9) .

### *7.3.2 Extreme Value distribution (EVD)*

The non-exceedance probability, *F(v)* of an *EVD* with wind speed *v* is given by **[173]**:

$$F(v) = \exp\left[ -\left( f_1 - \mathrm{sgn}(\tau) \cdot f_2 \cdot \frac{v-m}{\sigma} \right)^{\frac{1}{\tau}} \right] \tag{7.1}$$

The coefficient $f_1$ and $f_2$ depend on the curvature parameter $\tau$ and are given as follows:

$$f_1 = \Gamma(1+\tau) \tag{7.2}$$

$$f_2 = \sqrt{\Gamma(1+2\tau) - f_1^2} \tag{7.3}$$

where $m$ is the mean, $\sigma$ is the standard deviation, $\tau$ is the curvature parameter and $\Gamma(^*)$ is the Gamma function.

For $\tau = 0$, the extreme distribution of type I is obtained. The expression of type I is given as:

$$F(v) = \exp\left[-\exp\left(-\left[\gamma' + \frac{\pi}{\sqrt{6}}\frac{v - m_{gumbel}}{\sigma_{gumbel}}\right]\right)\right] \tag{7.4}$$

where $m_{gumbel}$ and $\sigma_{gumbel}$ are the mean and standard deviation of Gumbel distribution, $\gamma'$ is the Euler constant ($\gamma'$ = 0.5772). Table 7.1 shows the generalized *EVD* types and their properties.

Table 7.1: Generalized Extreme Value distribution type and their properties

| **Distribution** | **Type** | **Curvature parameter (τ)** | **Curvature** | **Lower Bound** | **Upper Bound** |
|---|---|---|---|---|---|
| Gumbel | I | τ = 0 | Linear | None | None |
| Fréchet | II | τ < 0 | Concave | $x_{\min} = m + \sigma\frac{f_1}{f_2}$ | None |
| Reverse Weibull | III | τ > 0 | Convex | None | $x_{\max} = m + \sigma\frac{f_1}{f_2}$ |

*7.3.2.1 Method Employed for Estimating the Parameters of Extreme Value Distribution*

The least-square method is the most commonly used method to estimate the parameters of generalized *EVD* **[173]**. In this study, the least-square method has been adopted to estimate the parameters of *EVD*. The expression for the linearization of the cumulative distribution functions as mentioned in Eqs. (7.5) and (7.6) are given as:

$$(-\ln F(v))^{\tau} = f_1 - f_2 \cdot \left(\frac{v - m}{\sigma}\right) \tag{7.5}$$

and

$$-\ln(-\ln F(v)) = \gamma' + \frac{\pi}{\sqrt{6}}\frac{v - m_{gumbel}}{\sigma_{gumbel}} \tag{7.6}$$

For estimating the parameters, the extreme wind speed data have been sorted into an ascending order viz., $v_1 \leq v_2 \leq v_3 \ldots v_N$. The formula for defining the plotting position is proposed by **Gringorten [225]**, i.e., $\alpha = 0.44, \beta = 0.12$, has been used to determine cumulative probability. The shape parameters ($\tau$) has been varied from -0.45 to 0.45 in a step size of 0.05. For different shape parameters, Eqs. (7.5) and (7.6) have been used to estimate the mean and standard deviation by using the graphical technique. The distribution with minimum sum of square error is the most suitable distribution for *EVA*. The calculation for estimating the sum of square error has been mentioned in the section 7.3.2.2.

Eqs. (7.5) and (7.6) have linear forms such as $y = a_1x+b_1$ and $y = a_2x+b_2$. In Eq. (7.5), $y$ can be equated to $(-\ln F(v))^{\tau}$, whereas in Eq. (7.6), $y$ can be equated to $(-\ln(-\ln F(v)))$. In both cases $x$ can be equated to $v$. The regression coefficients, $a_1$, $b_1$, $a_2$, and $b_2$ can be found by the method of least squares. Once these regression coefficients are known, the parameters, $m$, $\sigma$, and $m_{gumbel}$, $\sigma_{gumbel}$ can be determined from Eqs. (7.7) and (7.8) respectively:

$$m = \frac{f_1 - b_1}{a_1}; \sigma = -\frac{f_2}{a_1} \tag{7.7}$$

$$m_{gumbel} = \frac{\gamma - b_2}{a_2}; \sigma_{gumbel} = \frac{\pi}{\sqrt{6}a_2} \tag{7.8}$$

*7.3.2.2 Estimation of Best extreme Value Distribution*

To evaluate the goodness of fit of three types of *EVD*, the sum of squared error has been calculated using Eq. (7.9) and is given as:

$$SSE = \sum_{i=1}^{N} (Y_{theoretical} - Y_{observed})^2 \tag{7.9}$$

where

$$Y_{observed} = \left(-\ln\left(-\ln F\left(v\right)\right)\right) \tag{7.9a}$$

$$Y_{theoretical} = -\ln\left(\left(f_1 - f_2\left(\frac{v-m}{\sigma}\right)\right)^{1/\tau}\right) \tag{7.9b}$$

The parameters of probability distributions for which this sum of square error has a minimum value should be taken as the best estimator.

### 7.4 Plotting Position Formulae

Plotting position is an important tool to graphically analyze the data records such as rainfall, flood, wind speed data, etc. It has many advantages too viz., it is applied to derive the test statistics of the goodness of fit and is also helpful in selecting the appropriate probability density function. Earlier, several researchers **[225-228]** proposed unbiased plotting position formulae for *EVA*. Since then various attempts to derive the exact plotting position formula for the *GEV* have continued for decades. Recently, many researchers **[229-231]** proposed several other plotting position formulae for *GEV* distribution by taking into account the effect of coefficient of skewness. Table 7.2 shows the name of authors and their proposed formulae of plotting position.

Table 7.2: Plotting position formulae used in this study for comparison.

| Authors | Plotting Position Formulae |
|---|---|
| Hazen (1914) | $F = \frac{i-0.5}{N}$ |
| Gringorten (1963) | $F = \frac{i-0.44}{N+0.12}$ |
| Cunnane (1978) | $F = \frac{i-0.4}{N+0.2}$ |
| Goel and De (1993) | $F = \frac{i-0.02\gamma-0.32}{N-0.04\gamma+0.36}$ |
| Kim (2012) | $F = \frac{i-0.3200}{N+0.0149\gamma^2-0.1364\gamma+0.3225}$ |

where, γ is the skewness coefficient, the third standardized moment. The expressions for estimation the skewness coefficient are given as:

$$\gamma = \frac{skewness}{(variance)^{3/2}} \tag{7.10}$$

where

$$Skewness = \frac{\left(N(N-1)\right)^{1/2}}{N-2} \frac{\frac{1}{N}\sum_{i=1}^{N}(v_i-\bar{v})^3}{\left[\frac{1}{N}\sum_{i=1}^{n}(v_i-\bar{v})^2\right]^{3/2}} \tag{7.11}$$

$$variance = \frac{1}{N-1}\sum_{i=1}^{n}(v_i-\bar{v})^2$$

### 7.5. Results and Discussion

Figure 7.1 shows the plot of the average number of hours above $V_{ref}$ versus $V_{ref}$ for Calcutta. From Figure 7.1 (a) it is difficult to predict the deviation of the average number of hours above $V_{ref}$ as both the axes have been plotted on the linear scale (linear-linear scale). Therefore, to find the deviation between the observed average number of hours above $V_{ref}$ from theoretical *W.pdf*, the axes of the plot have been converted into the logarithmic scale and three more graphs have been plotted. In Figure 7.1 (b) both axes have been converted into the logarithmic scale (log-log scale). Figure 7.1 (b) reveals the deviation of the upper tail from the theoretical *W.pdf*. However, this deviation of upper tail is prominent when ordinate has been converted into the logarithmic scale, and the abscissa is on the linear scale (log-linear scale) as the upper tail is heavily weighted as shown in Figure 7.1 (c). Conversely, when abscissa is converted into logarithmic scale and ordinate on the linear scale (linear-log scale), the lower tail is heavily weighted. The deviation of the observed average number of hours from theoretical *W.pdf* has been clearly observed from Figure 7.1 (d).

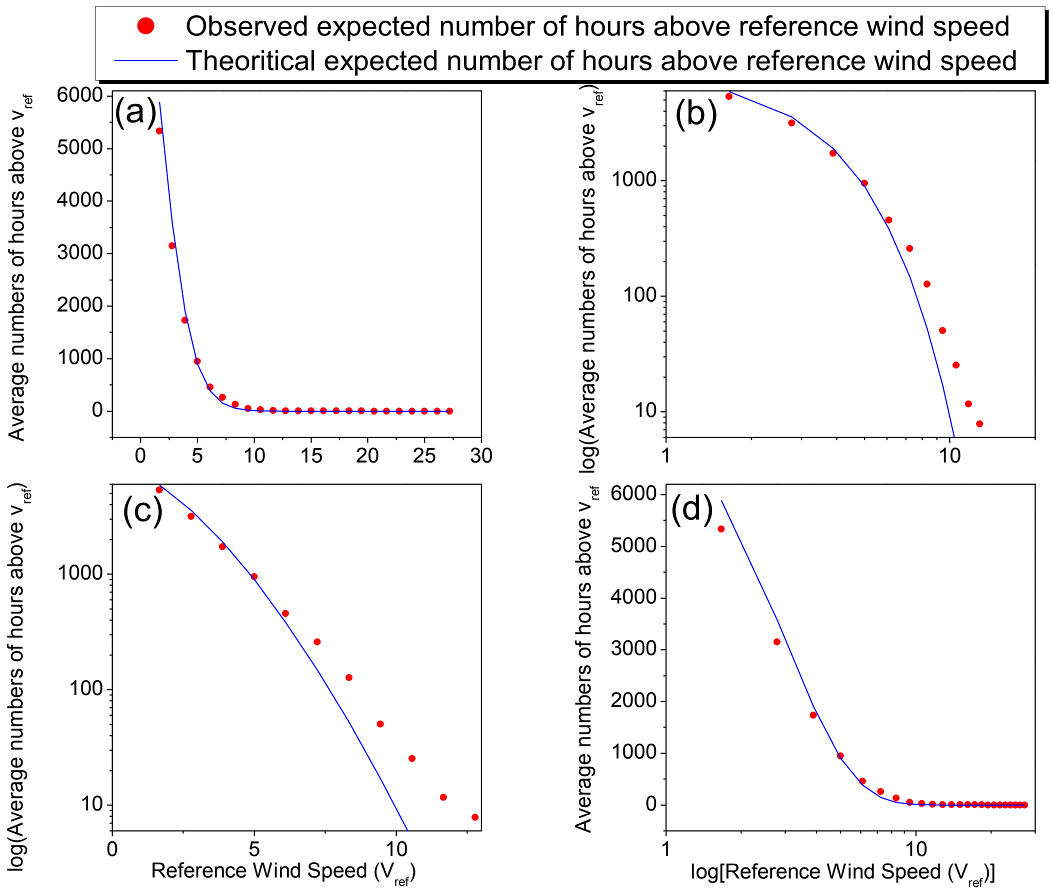


Figure 7.1: Deviation of an average number of hours above reference wind speed for Calcutta on (a) Linear-Linear Scale (b) Log-Log Scale (c) Log-Linear Scale (d) Linear-Scale.

As the focus of this study is on the extreme upper wind speed data, the behaviour of other two stations namely Trivandrum and Ahmedabad have been plotted on log-linear and log-log scale as shown in Figure 7.2.

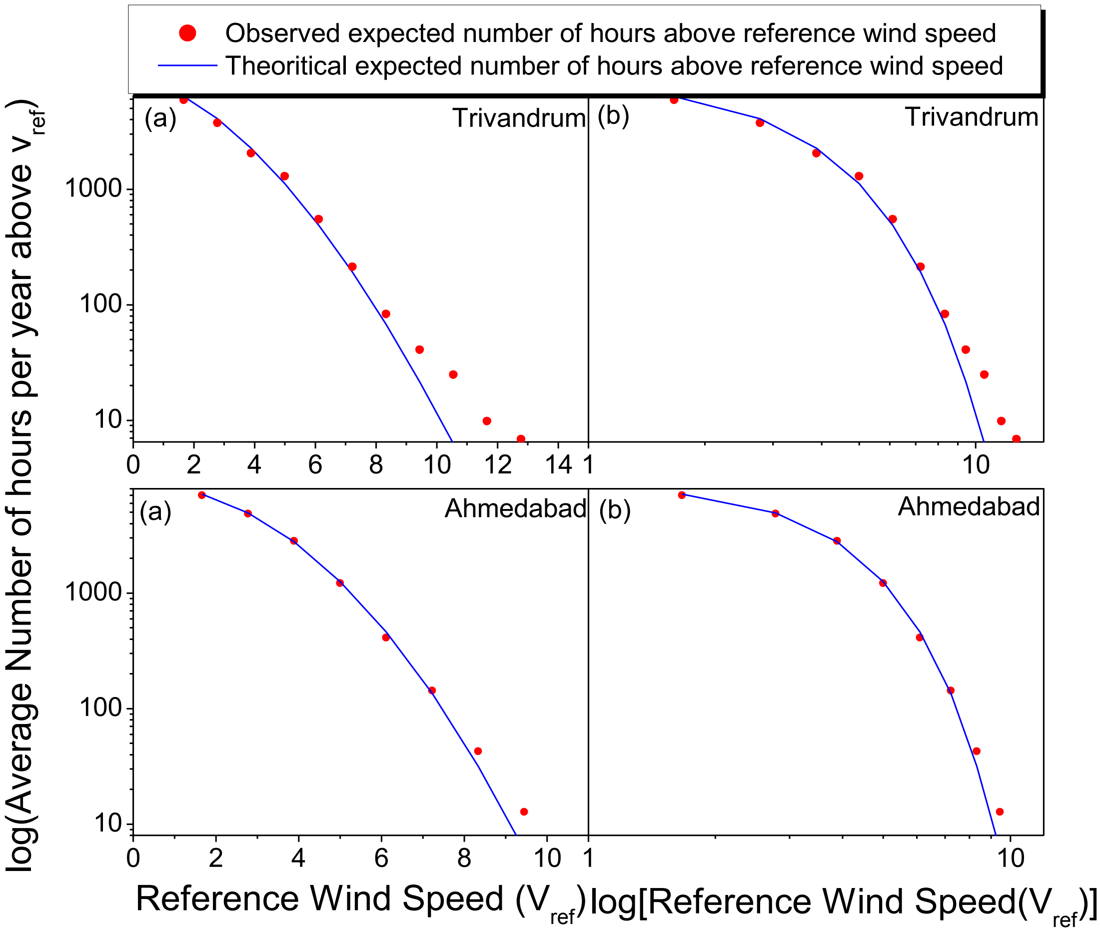


Figure 7.2: Average number of hours above reference wind speed for Trivandrum and Ahmedabad on (a) Log-Linear Scale (b) Log-Log Scale.

It is seen from Figure 7.2 (a) that the upper tail deviation of the observed average number of hours is more pronounced in Trivandrum as compared to that of Ahmedabad. Figures 7.1, and 7.2 reveal that the entire average number of hours did not follow the *W.pdf*. Therefore, from parent data set an upper extreme data have to be separated out for *EVA*. Hence, in this study initially, the data set has been sorted out in ascending order. The appropriate threshold value has been calculated using Weibull quantile function (see section 4.2.9). The dataset below or equal to this threshold value follows the *W.pdf*, whereas, above which the data set follows the *EVD*.

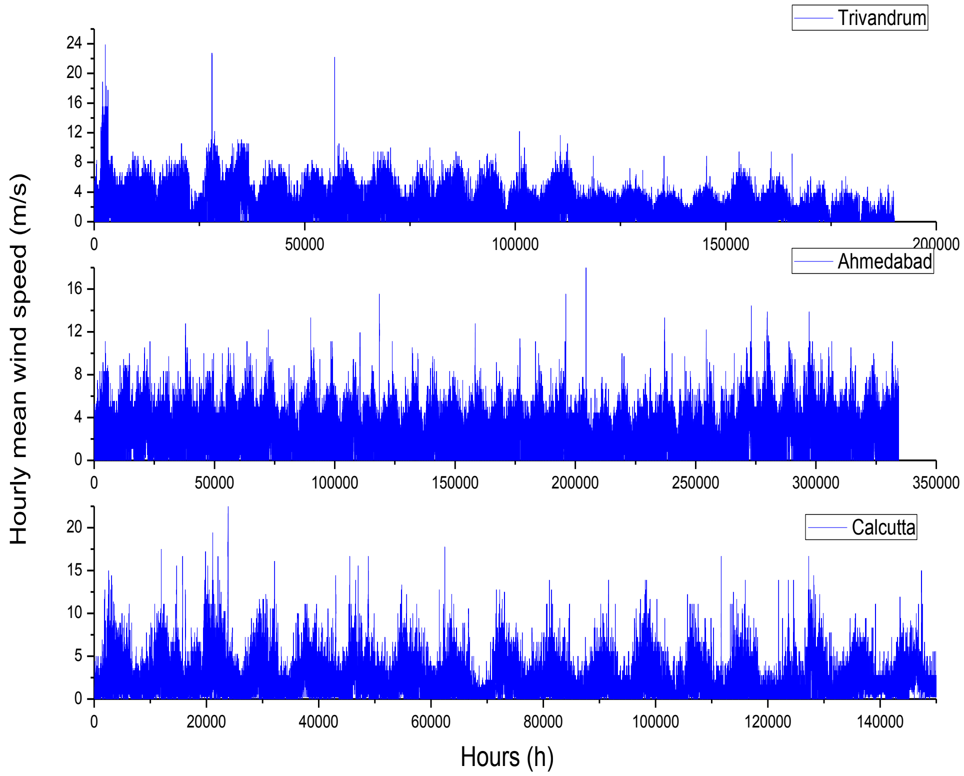


Figure 7.3: Distribution of hourly mean wind speed with some gaps in a year.

Figure 7.3 shows the run sequence plot of hourly mean wind speed for many years of three stations namely Trivandrum, Ahmedabad, and Calcutta respectively. According to this figure, the data are neither stationary nor follow the Gaussian distribution for all three stations. The zero wind speed shows that there is no wind during that period. These data cannot be directly used for probabilistic analysis either it is for parent (Weibull) distribution or for *EVA* (Fisher-Tippet distribution). Therefore, to eradicate the problem of non-stationarity, the data are initially classified into the block of months. From each block, the extreme data has been separated from that of the parent data, so that the chances of correlation among data can be minimized.

Figure 7.4 shows the 3-D plot of Trivandrum for March. This 3-D plot shows wind speed on the z-axis, days on the x-axis and hours on the y-axis. The threshold value for March is 7.222 *m/s* and has been marked with pink colour. The data equal or above the threshold value have been considered as extreme data. The basic reason for choosing March is that higher $V_d$ has been obtained for this month in comparison with other months as given in Table 7.3.

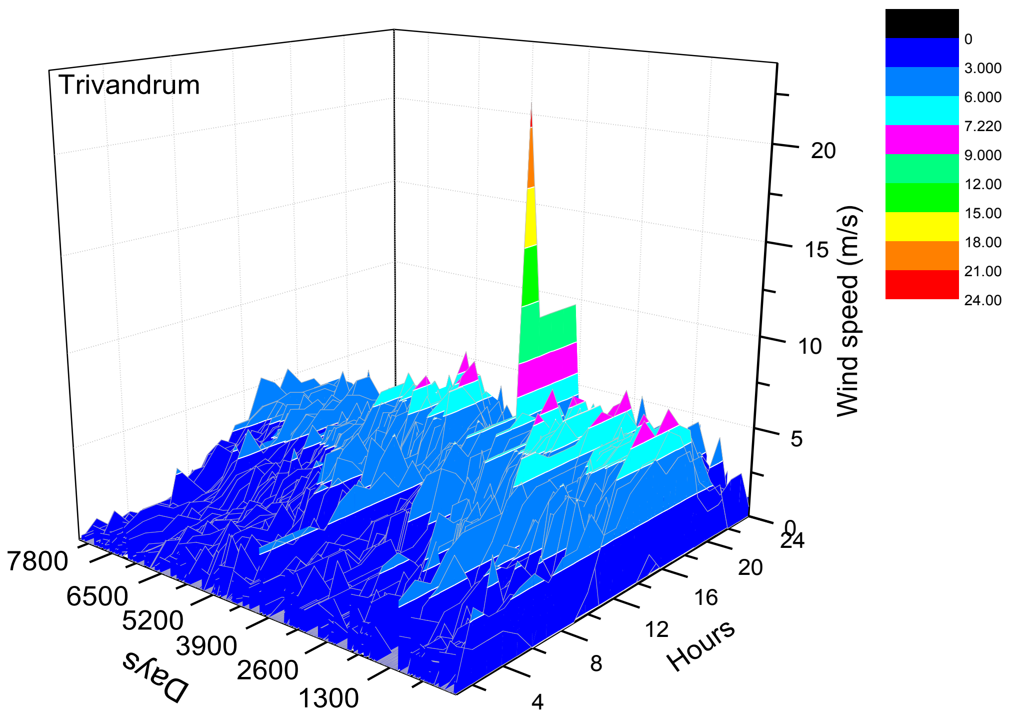


Figure 7.4: 3-D plot of month of March with threshold value is represented by pink colour for Trivandrum station.

It is seen from the Figure 7.4 that there is a wide gap between the extreme data points are shown with pink and other colours, viz., green, yellow, orange, and red which in turn reveals that the extremes are uncorrelated. The size of these extreme data is greater than the numbers of years under consideration. Hence, these extremes can be considered for probabilistic analysis by *EVD*.

Figures 7.5 and 7.6 show the 3-D plot of Ahmedabad and Calcutta for September. For this month the highest wind speed has been obtained for both these stations as given in Tables 7.4 and 7.5. The threshold wind speed values for Ahmedabad and Calcutta are 7.222 and 7.027 *m/s* respectively as shown with pink colour.

The colour other than pink viz., green, yellow, orange and red show the extreme data points. Similar to Figure 7.5, it has been observed that there is a wide gap between extreme data points. Therefore, it is concluded that extreme wind speed data are highly uncorrelated.

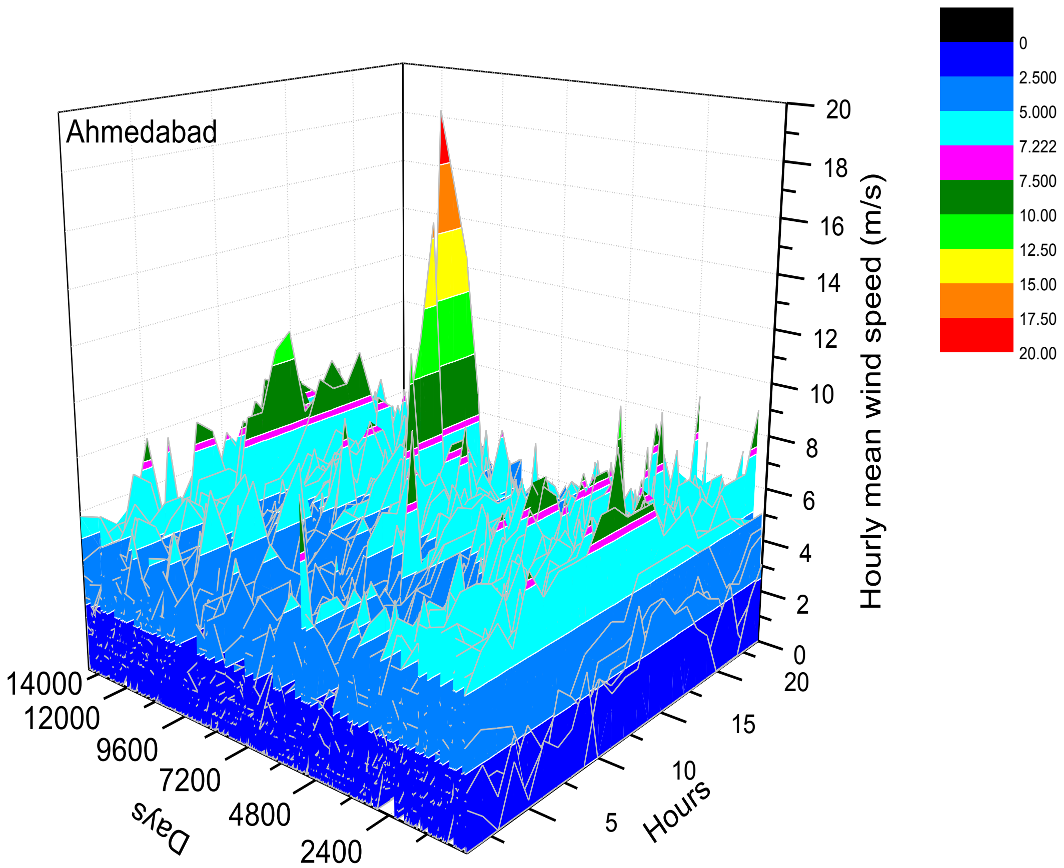


Figure 7.5: 3-D plot of September with the threshold value is represented by pink colour for Ahmedabad station.

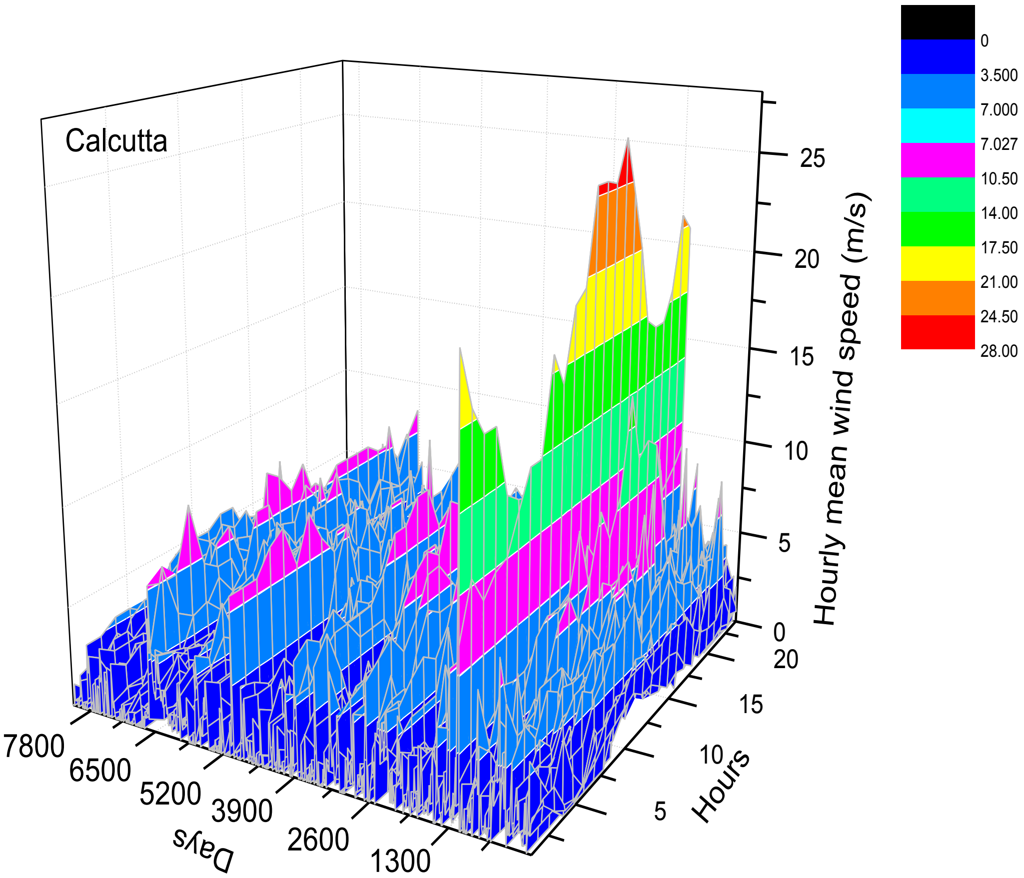


Figure 7.6: 3-D plot of September with the threshold value is represented by pink colour for Calcutta station.

Figure 7.7 shows the non-exceedance probabilities of extreme hourly mean wind speeds for March as well as sub-annual maxima and annual maxima for Trivandrum. Table 7.3 shows the estimated parameters of best-fitted *EVD*, sample size, $V_d$, $R^2$ and *RMSE* for different months as well as sub-annual and annual maxima for Trivandrum.

Table 7.3: Estimation of design wind speed based on fitting of *EVD* on month, Sub-annual and Annual maxima with the goodness of fit for Trivandrum.

| **Month** | **Distrib-ution** | **Curva-ture param-eter (τ)** | **Mean (m) (*m/s*)** | **Stand-ard deviate-on (σ) (*m/s*)** | **No. of data points** | **Design Speed (*m/s*)** | ***$R^2$* *(%)*** | **RMS-E** |
|---|---|---|---|---|---|---|---|---|
| Jan. | Reverse Weibull | 0.05 | 5.899 | 0.451 | 48 | 7.883 | 84.3 | 0.024 |
| Feb. | Gumbel | 0 | 6.371 | 0.544 | 97 | 9.054 | 76.5 | 0.615 |
| March | Fréchet | -0.45 | 8.182 | 5.314 | 51 | 50.189 | 86.3 | 0.458 |
| April | Fréchet | -0.25 | 8.539 | 1.253 | 116 | 19.131 | 92.4 | 0.130 |
| May | Reverse Weibull | 0.05 | 11.257 | 2.148 | 166 | 20.713 | 96.6 | 0.011 |
| June | Fréchet | -0.25 | 10.795 | 2.260 | 199 | 29.896 | 96.1 | 0.094 |
| July | Fréchet | -0.45 | 10.555 | 4.891 | 180 | 49.219 | 88.4 | 0.486 |
| Aug. | Fréchet | -0.05 | 9.574 | 0.751 | 97 | 13.251 | 93.3 | 0.017 |
| Sept. | Fréchet | -0.15 | 8.839 | 0.727 | 112 | 13.887 | 86.6 | 0.086 |
| Oct. | Reverse Weibull | 0.05 | 7.076 | 0.556 | 132 | 9.523 | 85.0 | 0.023 |
| Nov. | Fréchet | -0.25 | 6.161 | 0.939 | 130 | 14.100 | 94.7 | 0.109 |
| Dec. | Fréchet | -0.15 | 5.341 | 0.625 | 126 | 9.678 | 89.3 | 0.077 |
| Annual | Fréchet | -0.35 | 11.182 | 8.204 | 23 | 89.028 | 86.6 | 0.265 |
| Sub-Annual | Fréchet | -0.25 | 9.603 | 5.323 | 46 | 54.581 | 91.3 | 0.134 |

As seen from Table 7.3 that majority of the months along with annual maxima and sub-annual maxima are following the Fréchet distribution except for February which follows the Gumbel distribution and January, May, and October which follow Reverse *W.pdf*. The coefficient of determination ($R^2$) and *RMSE* reveal that the theoretical distribution fit the observed data quite decently. The number of the data points for *EVA* is more than 30 which is a primary requirement for *EVA*. However, in annual maxima the number of the data points is 23 and therefore, chances of fallacious judgment about the estimation of $V_d$ are much pronounced. Hence, sub-annual maxima have been used for EVA with the increase in the number of data points; the estimated $V_d$ is more authenticated. The month of March shows the maximum $V_d$, i.e., 50.189 *m/s* for the 1000 years return period. The probable reason is that Trivandrum has the mixed climatic condition which is influenced by tropical monsoon and a tropical savanna climate. As a result, Trivandrum has similar weather condition throughout the year. However, Trivandrum has an average high temperature in March and April. As a result, the pressure gradient is also high that causes high wind speed as compared to other months of the year. Based on annual maxima the $V_d$ has been estimated as 89.03 *m/s* and based on sub-annual maxima $V_d$ has been found as 54.58 *m/s* (see Figure 7.7).

In this context, it is worthy to mention that Indian standard **[179]** uses daily maximum gust wind speed data for the specification of the $V_d$. The maximum gust wind speed data have been fitted by type I distribution for 50 years return period. According to the recent Atomic Energy Regulatory Board guidelines, important structures, especially, the slender structures in the nuclear industry such as natural draft cooling tower, heavy water drafting tower have return periods of 1000 years.

The probability or risk factor for Trivandrum, as per Indian standard **[179]**, has been found as 1.29 (the calculation to find risk factor for 1000 years has been shown in Appendix C). The basic wind speed of Trivandrum is 39 *m/s*, so the calculated $V_d$ based on maximum gust wind speed for Trivandrum is 50.31 *m/s*. Hence, compared to monthly assessment of hourly mean wind speed data it has been found that there is an overestimation of 0.241% regarding wind speed as well as 0.48% regarding wind load.

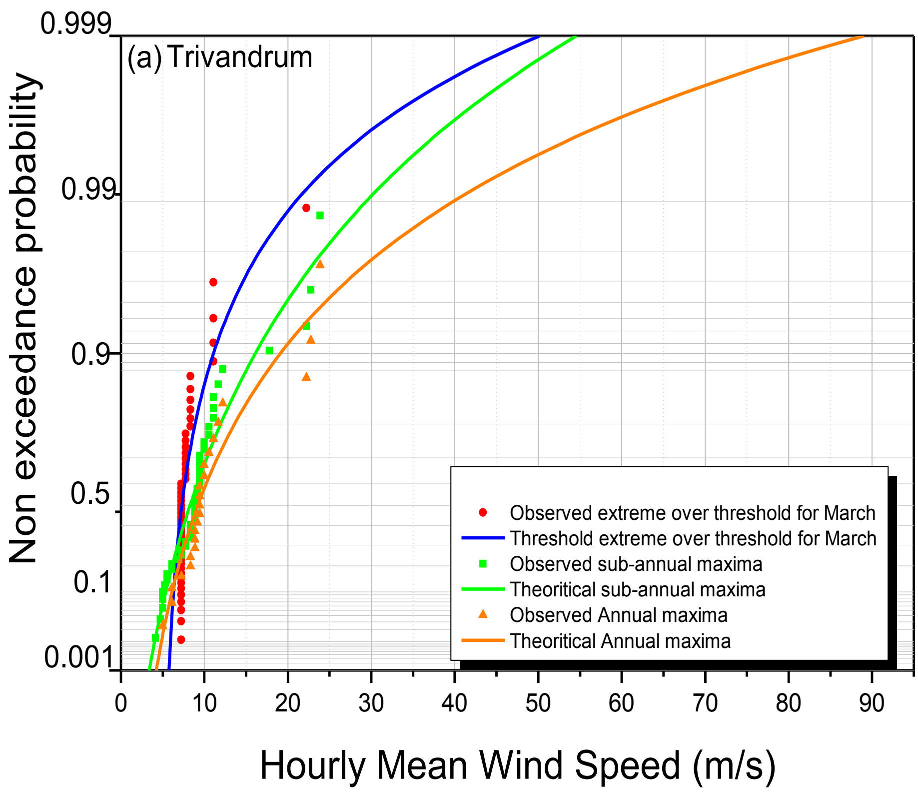


Figure 7.7: Non-exceedance probability of extremes hourly mean wind speed for March, sub-annual maxima and annual maxima for Trivandrum.

Figure 7.8 shows the non-exceedance probability of extreme hourly mean wind speeds for September as well as sub-annual, and annual maxima for Ahmedabad. Table 7.4 shows the estimation of $V_d$ based on fitting of *EVD* on different months including sub-annual and annual maxima with two *GOF* for Ahmedabad.

Table 7.4: Estimation of design wind speed based on fitting of *EVD* on month, Sub-annual and Annual maxima with the goodness of fit for Ahmedabad.

| **Month** | **Distrib-ution** | **Curv-ature para-meter (τ)** | **Mean (m) (*m/s*)** | **Stand-ard deviat-ion (σ) (*m/s*)** | **No. of data points** | **Design Speed (*m/s*)** | $R^2(\%)$ | **RMSE** |
|---|---|---|---|---|---|---|---|---|
| Jan. | Fréchet | -0.40 | 6.617 | 1.2558 | 95 | 18.3213 | 91.02 | 0.3132 |
| Feb. | Fréchet | -0.15 | 7.598 | 0.6830 | 53 | 12.3375 | 92.60 | 0.0634 |
| Mar. | Fréchet | -0.15 | 7.597 | 0.6097 | 62 | 11.8272 | 90.85 | 0.0707 |
| Apr. | Fréchet | -0.30 | 8.218 | 0.9691 | 55 | 17.0360 | 93.11 | 0.1587 |
| May | Fréchet | -0.25 | 8.834 | 1.5284 | 173 | 21.7494 | 91.77 | 0.1377 |
| June | Gumbel | 0 | 9.887 | 1.0145 | 206 | 14.8945 | 95.88 | 0.2594 |
| July | Fréchet | -0.35 | 9.860 | 0.9437 | 77 | 18.8149 | 95.62 | 0.1675 |
| Aug. | Fréchet | -0.35 | 9.418 | 1.0946 | 91 | 19.8053 | 95.89 | 0.1643 |
| Sept. | Fréchet | -0.45 | 8.078 | 2.9876 | 169 | 31.6937 | 96.11 | 0.2803 |
| Oct. | Fréchet | -0.10 | 7.371 | 0.9552 | 133 | 13.3028 | 95.10 | 0.0323 |
| Nov. | Gumbel | 0 | 8.007 | 2.4740 | 78 | 20.2219 | 90.62 | 0.3906 |
| Dec. | Fréchet | -0.10 | 6.484 | 0.5982 | 74 | 10.1988 | 94.48 | 0.0341 |
| Annual | Fréchet | -0.10 | 11.480 | 2.4407 | 39 | 26.6368 | 98.26 | 0.0190 |
| Sub-Annual | Fréchet | -0.10 | 10.349 | 2.2166 | 78 | 24.1137 | 98.15 | 0.0198 |

From Table 7.4, it has been observed that the extreme wind speed data for all the months including annual and sub-annual follow Fréchet distribution except for June and November. The measure of goodness of fit, namely, $R^2$ ( > 90%) and *RMSE* reveal that good fitting agreement exists between the theoretical distribution and the observed data. The $V_d$ has been calculated for all the months as well as for sub-annual and annual maxima.

It has been observed that among the months, September shows the highest $V_d$, i.e., 31.69 *m/s* as also shown in Figure 7.8. The probable reason for having the highest wind speed in September is that the September is the month of the end of South-West monsoon season. The high pressure prevails during this month causing a heavy thunderstorm associated with the withdrawal of the monsoon season. As the South-West monsoon ends during this month bringing this month in a category of unsettled weather. The $V_d$ for sub-annual, and annual maxima are approximately same, i.e., 26 *m/s*. This may be due to the reason that the total number of data points for annual maxima is 39, which is quite a decent number for *EVA.* Hence, for both the sample size viz., sub-annual maxima and annual maxima the design wind is approximately same (see Figure 7.8). However, from Indian standard **[179]**, the basic wind speed for Ahmedabad is 44 *m/s*, multiplying this basic wind speed with risk coefficient ($k_1$) for the 1000 year return period the $V_d$ has been found to be 58.40 *m/s*. Hence, there is an overestimation of $V_d$ by 45.73*%*. The high value of overestimation is due to the reason that Ahmedabad is situated on the west coast where the storms are comparatively much lesser than that of the east coast. However, Indian standard **[179]** unanimously uses type I distribution for both coastal regions of the country without considering the fact that the types of extremes for these two coastal regions are entirely different due to the difference in turbulence intensities of storms. For example, the east coast is more prone to the cyclone which has much higher turbulence intensity than monsoon gales which frequently occur on the west coast. As a result, the $V_d$ as specified in Indian code for a station in the west coast like Ahmedabad is much higher.

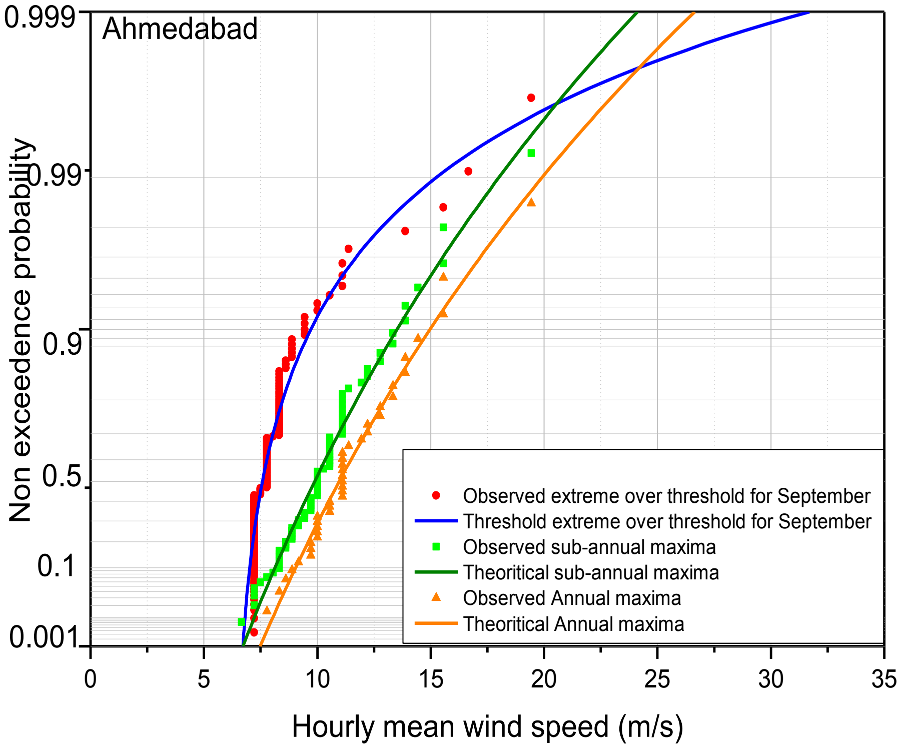


Figure 7.8: Non-exceedance probability of extremes hourly mean wind speed for September, sub-annual maxima and annual maxima for Ahmedabad.

Figure 7.9 shows the non-exceedance probability of extreme hourly mean wind speeds for September as well as for sub-annual and annual maxima for Calcutta. Table 7.5 shows the estimation of $V_d$ based on fitting of *EVD* on different month as well as sub-annual and annual maxima with two *GOF* for Calcutta.

Table 7.5: Estimation of design wind speed based on fitting of *EVD* on month, Sub-annual and Annual maxima with the goodness of fit for Calcutta.

| **Month** | **Distrib-utions** | **Curva-ture param-eter (τ)** | **Mean (m) (*m/s*)** | **Stand-ard deviat-ion (σ) (*m/s*)** | **No. of data points** | **Design Speed (*m/s*)** | ***R*$^2$*(%)*** | ***RMSE*** |
|---|---|---|---|---|---|---|---|---|
| Jan. | Fréchet | -0.05 | 5.017 | 0.7517 | 198 | 9.181 | 91.93 | 0.0194 |
| Feb. | Fréchet | -0.30 | 6.046 | 1.5459 | 239 | 20.112 | 96.23 | 0.1267 |
| Mar. | Fréchet | -0.30 | 7.869 | 1.7630 | 257 | 23.911 | 97.75 | 0.0981 |
| Apr. | Fréchet | -0.15 | 10.855 | 1.6842 | 132 | 22.540 | 96.77 | 0.0426 |
| May | Fréchet | -0.20 | 11.002 | 1.3225 | 150 | 21.186 | 96.05 | 0.0689 |
| June | Fréchet | -0.20 | 9.424 | 1.5368 | 351 | 20.087 | 94.32 | 0.0841 |
| July | Fréchet | -0.25 | 8.806 | 1.3584 | 170 | 20.284 | 97.22 | 0.0800 |
| Aug. | Fréchet | -0.25 | 9.180 | 1.8639 | 250 | 24.930 | 92.23 | 0.1355 |
| Sept. | Fréchet | -0.45 | 9.434 | 6.9732 | 266 | 64.554 | 84.09 | 0.5926 |
| Oct. | Fréchet | -0.30 | 6.989 | 2.4507 | 289 | 29.288 | 91.59 | 01907 |
| Nov. | Fréchet | -0.35 | 6.463 | 2.4855 | 197 | 30.048 | 91.84 | 0.2434 |
| Dec. | Fréchet | -0.35 | 6.923 | 3.6485 | 118 | 41.544 | 88.43 | 0.2806 |
| Annual | Fréchet | -0.10 | 14.618 | 4.1001 | 23 | 40.078 | 90.77 | 0.0437 |
| Sub-Annual | Fréchet | -0.05 | 12.883 | 3.9303 | 46 | 34.653 | 96.52 | 0.0126 |

As seen from this table that all the months including annual and sub-annual data follow the Fréchet distribution. September has the highest $V_d$ for Calcutta, i.e., 64.55 *m/s*. However, the $V_d$ has been found as 40. 0789 *m/s* based on annual maxima and the same has been found as 34.653 *m/s* based on sub-annual maxima (see Figure 7.9).

This difference in the $V_d$ may be caused due to the low observation period for available wind speeds in Calcutta. From Indian standard **[179]**, the basic wind speed for Calcutta is 50 *m/s* and multiplying this basic wind speed with risk coefficient ($k_1$) for the 1000 year return period the $V_d$ has been found as 68.35 *m/s*. Hence, there is a slight overestimation of 5.56% in regard to $V_d$ and 11.429% in regard to design wind load. Calcutta lies in the east coast of India which is more prone to tropical cyclone and is having much higher turbulence intensity. As a result, there is a slight difference between the estimated $V_d$ and the one that is mentioned in the Indian code.

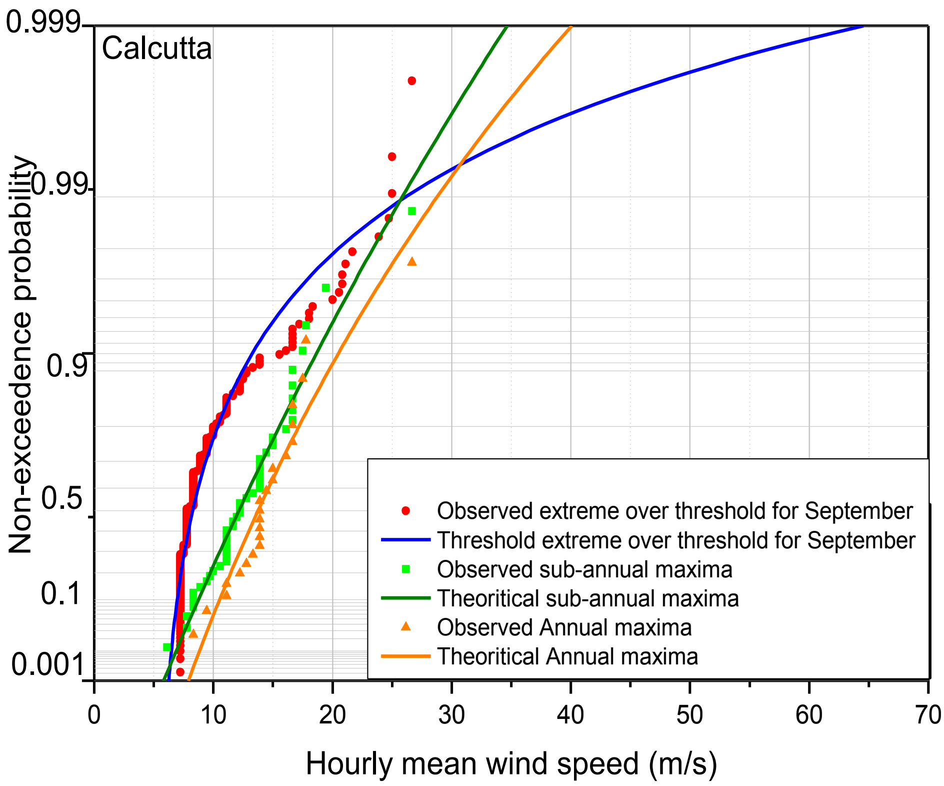


Figure 7.9: Non-exceedance probability of extremes hourly mean wind speed for September, sub-annual maxima and annual maxima for Calcutta.

From above discussions, it has been found that Fréchet distribution is suitable to model extreme wind speed data among three types of extreme value distribution for all three stations. However, on the other hand, the exceedances over a threshold which tends to the upper end follow a generalized Pareto distribution (*GPD*) **[215]**. Therefore, on comparing *GPD* and Fréchet distribution on the selected data above the selected thresholds as mentioned above for March for Trivandrum, September for Ahmedabad and also September for Calcutta, it has been found that based on Akaike Information Criteria (*AIC*) as shown in Table 7.6.

Table 7.6: Akaike Information Criteria for *GPD* and Fréchet distributions

| **Stations** | **Akaike Information Criteria** | |
|---|---|---|
| | **Fréchet** | ***Generalized Pareto Distribution**** |
| Ahmedabad | 356.96 | 203.9384 ($\alpha$ = -0.4467;$\beta$ = 0.5109 *m/s*) |
| Trivandrum | 163.0108 | 94.9904 ($\alpha$ = -0.0237; $\beta$ = 1.3441 *m/s*) |
| Calcutta | 1021.1 | 943.0597 ($\alpha$ = -0.2289; $\beta$ = 1.9505 *m/s*) |

* $\alpha$ is the shape and $\beta$ is the scale parameter of the *GPD*

The *GPD* fits the observed data better than the Fréchet distribution. To estimate the design wind speed for the 1000 years return period, the eligibility of the *GPD* to extrapolate outside the range of given data set is of utmost importance. Therefore, extreme data set for these three locations have been fitted with both distributions namely *GPD*, and Fréchet distribution on Gumbel probability paper to evaluate the design wind speed (see Figure 7.10).

It is seen from the Figure 7.10 that for Ahmedabad both the distributions perform equally well and the design wind speed for 1000 years is approximately 31.5 *m/s*. However, while applying the similar approach for Trivandrum and Calcutta, it has been found that the design wind speed is approximately 50 *m/s* for Fréchet distribution and approximately 18 *m/s* for *GPD* for Trivandrum and on the similar way, the design wind speed obtained from Fréchet distribution is approximately 63 *m/s* and the same has been obtained approximately 39 *m/s* from *GPD* for Calcutta. The design wind speed estimated by *GPD* is far less than that estimated by Fréchet distribution. It is well known that Calcutta lies on the cyclone-prone zone of India, such less estimated design wind speed as predicted by *GPD* has not been practically feasible.  The similar reason is also valid for Trivandrum. Although the *GPD* has been found to be a better-fitted distribution to the given data set, its eligibility to extrapolate outside the range of the given data set is under serious ambiguity. Similar conclusions were also drawn in by **[232]**. He clearly stated that the *GPD* approach could not define design loads unambiguously. This is simply not acceptable as a basis for codes of practice and safety standards. Therefore, the estimated design wind speed for 1000 years return period does not justify its use for the unbiased estimation of the design wind speed.

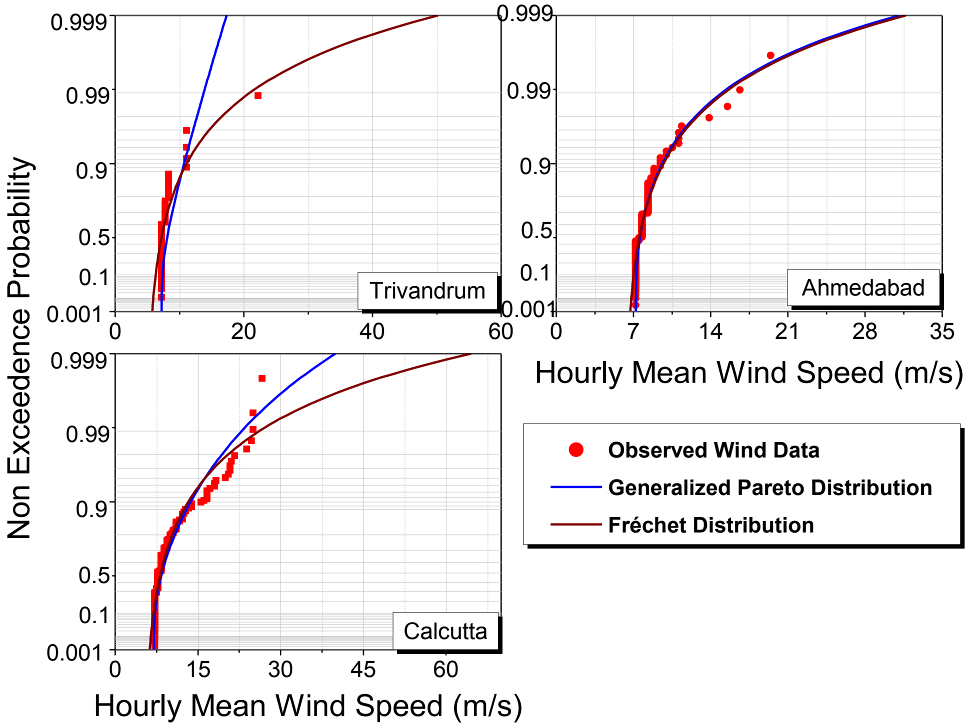


Figure 7.10: Fitting of *GPD* and Fréchet distribution on extremes hourly mean wind speed data for March, September for Trivandrum, Ahmedabad, and Calcutta respectively.

In this study, a comparison has been carried out to determine the suitability of four other plotting position formulae viz., **[227-230]** with respect to **[225]** in terms of $R^2$ and *RMSE* for all three stations for the months that show maximum $V_d$. Table 7.7 shows the comparison of five formulae with the estimated parameters and their *GOF* on March for Trivandrum station and September for both Ahmedabad and Calcutta stations. As seen from Table 7.7 that three new formulae based on skewness coefficient do not help in improvement of $R^2$, and *RMSE* and their performances have been found to be lower than **Gringorten [225]** in terms of $R^2$ and *RMSE*. The other two unbiased plotting position formulae also do not help in significant improvement of $R^2$ and *RMSE* as compared to the one proposed by **Gringorten [225]**. From the above comparisons, it can be understood that these new plotting positions do not enable significant improvement of the *GOF* as compared to the one proposed by **Gringorten [225]**.

Table 7.7: Goodness of fit for six different plotting position formulae

| **Cumulative relative frequency formulae** | **Trivandrum (March)** | | **Ahmedabad (September)** | | **Calcutta (September)** | |
|---|---|---|---|---|---|---|
| | ***$R^2$* (%)** | ***RMSE*** | ***$R^2$* (%)** | ***RMSE*** | ***$R^2$* (%)** | ***RMSE*** |
| $F(v)=\frac{i-0.44}{N+0.12}$ | 86.3 | 0.458 | 96.1 | 0.280 | 84.0 | 0.592 |
| $F(v)=\frac{i-0.4}{N+0.2}$ | 85.4 | 0.462 | 96.2 | 0.272 | 84.9 | 0.566 |
| $F(v)=\frac{i-0.5}{N}$ | 87.7 | 0.450 | 95.8 | 0.298 | 82.5 | 0.637 |
| $F(v)=\frac{i-0.02\gamma-0.32}{N-0.04\gamma+0.36}$ | 83.9 | 0.467 | 96.2 | 0.264 | 86.5 | 0.522 |
| $F(v)=\frac{i-0.3200}{N+0.0149\gamma^2-0.1364\gamma+0.3225}$ | 85.7 | 0.460 | 95.9 | 0.295 | 85.7 | 0.546 |

### 7.6 Summary

The aim of this study is to propose a new approach for extreme value analysis (*EVA*) of hourly mean wind speed data. This new approach is a most suitable for the region of varied wind climate as it takes the advantages of both the block size and the peaks over the threshold and avoids their limitations. The hourly mean wind speed data enables in an accurate estimation of design wind speed ($V_d$) and thereby making the structure to be more economically viable. This study focuses on the estimation of the appropriate threshold value, as entire observed hourly mean wind speed data set do not follow the *W.pdf*, which is widely accepted distribution for wind speed data modelling. *W.pdf* is suitable to model wind speed data only up to a certain threshold, and after that, the wind speed data should be Modelled by extreme value distribution (*EVD*). However, the types of *EVD* (type I, II, and III) suitable for Indian wind speed data modelling is also important for different stations. As Indian code of standard has identified type I distribution as the appropriate one for all stations of India without considering the geographical location and climatic condition of that station which may lead to the fallacious judgment regarding the $V_d$. Therefore, in this study, three coastal locations of India namely Trivandrum, Ahmedabad and Calcutta have been considered for the analysis. The month wise data have an advantage that there is a wide gap between the extreme data points that make the data uncorrelated, and its non-stationarity does not have much effect, and the data are ideal for probabilistic analysis. While analyzing extreme data, it has been found that Fréchet distribution is more appropriate for most of the months as well as for annual and sub-annual maxima for all three stations.

As compared to the type I distribution which is suggested in Indian code, the sum of the squared error between the Fréchet and observed distribution on Gumbel probability paper has been found to be minimum. March is the month that shows the highest $V_d$ of 50.189 *m/s* for Trivandrum, and for both Ahmedabad and Calcutta the highest $V_d$ have been found in September, i.e., 31.69 and 64.55 *m/s* respectively. The percentage difference between estimated $V_d$ and the one that is mentioned in Indian code is high for Ahmedabad as compared to Trivandrum and Calcutta and the structures situated in Ahmedabad are not at all economically viable from the viewpoint of design wind load. The comparison between the plotting position formulae reveals that the plotting position formula as proposed by **Gringorten [225]** still acts as a useful tool for the parameter estimation of generalized *EVD*. The plotting position formulae recently developed based on skewness coefficient does not enable in significant improvement of the *GOF* for generalized *EVD*.