3. Flaws in Wind Speed Sampling Technique: Methodology for Its Eradication using Weibull Distribution

Indian Meteorological Department, Pune supplied the long-term hourly mean wind speed data in *km/h*. The data have been recorded at 10 *m* height above ground level measured with Dyne Pressure Tube Anemograph. The wind speed data are available in time series format where every data point represents hourly mean wind speed from each of 24 hours of a day. These data have been sampled at 10-minute mean values taken from the last 10-minute period before the full hour. This last 10-minute mean value represents the hourly mean wind speed. They record the wind speed data in the fractional *knot* up to first decimal place **[191]** and convert it into *km/h* by multiplying with the conversion factor of 1.852, after that it will be rounded to nearest integer. Moreover, if the decimal part of the wind speed in *km/h* is 0.5, it will be rounded to the nearest odd integer. For example, if wind speeds have the values of 10.5 *km/h* and 11.5 *km/h*, both will be rounded to 11 *km/h*, since 11 is the nearest odd integer to both values. The technique of sampling induce flaws in the data, and the histograms made from these data are not at all stable as can be evident from the histograms shown in Figure 3.1 of Ahmedabad. It is seen from the Figure 3.1 that the double trend (see Figure 3.1 (a), and (c)) as well as uneven bars, exist in the histograms. The double trend in the histograms is called 'rounding error,' and the uneven bars in the histograms is called 'sampling error.'

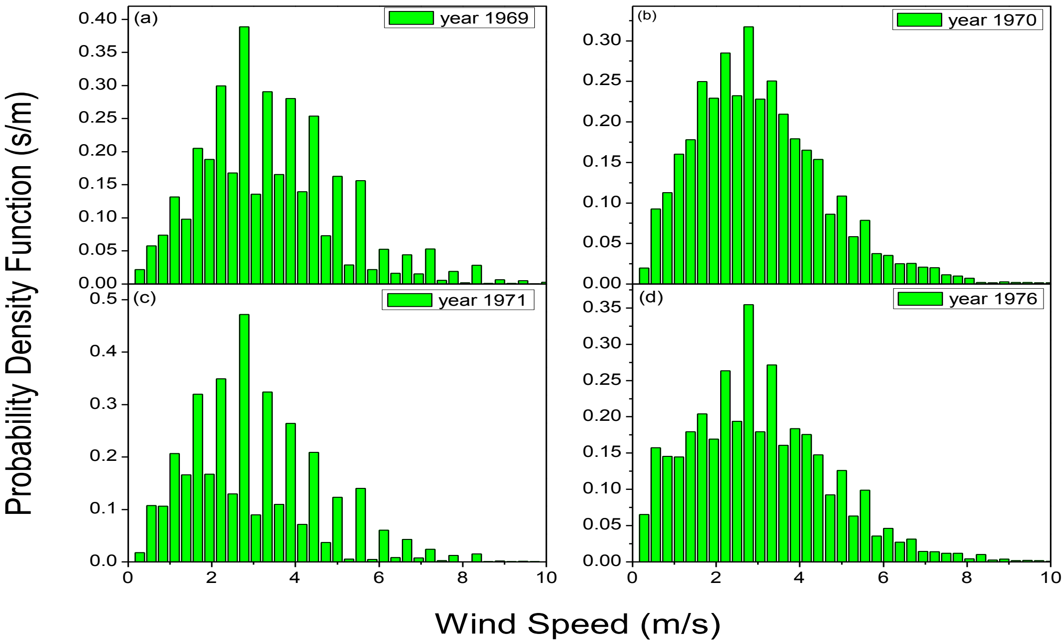

Figure 3.1: Histogram for the years 1969, 1970, 1971 and 1976

The double trend can be removed by increasing the class width of the histograms, so that the observed histograms become much stable. However, the extent of increase in the class width is a challenging task, otherwise, an excess increase in class width leads to information loss. Therefore, the suitable choice of class width is the prime requirement, so that the observed histograms become stable as well as not much information lost from the observed data. The simulated study for calculating the suitable choice of class width is given in section 3.1.

3.1 Choice of the Class Width of Histograms

Simulated wind speed data (order of 10^6) have been generated by inverse cumulative distribution function technique with known Weibull parameters, ($k_{act.}$ = 1.5, 2, 3 and $s_{act.}$ = 4), using the simulated wind speed, the Weibull parameters have been estimated after classifying the simulated wind speed data at different class width (Δv). The histograms of percentage error between estimated *k* ($k_{est.}$) and actual *k* ($k_{act.}$) for fixed '*s*' have been plotted with respect to varying class width (Δv) to get the optimum value of the class width.

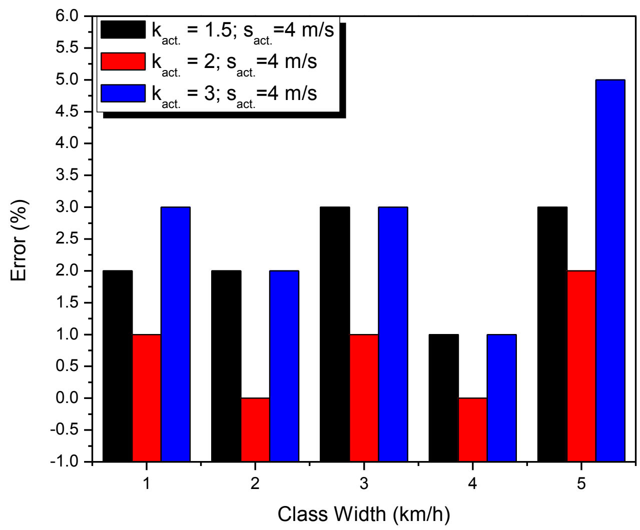

Figure 3.2: Determination of optimum value of the class width

Figure 3.2 shows the histograms of percentage error between estimated *k* ($k_{est.}$) and actual *k* ($k_{act.}$) is minimum and almost zero for Δv = 2 *km/h* and 4 *km/h*. To get stable histograms without not much loss of information 2 *km/h* class width is a suitable choice. Thus it can be concluded that the suitable choice of the class width for achieving the stable histograms is 2 *km/h* which will enable in the removal of a double trend from the histograms. Figure 3.3 shows the observed histogram of two years, i.e., 1977, and 1980 with 2 *km/h* class width.

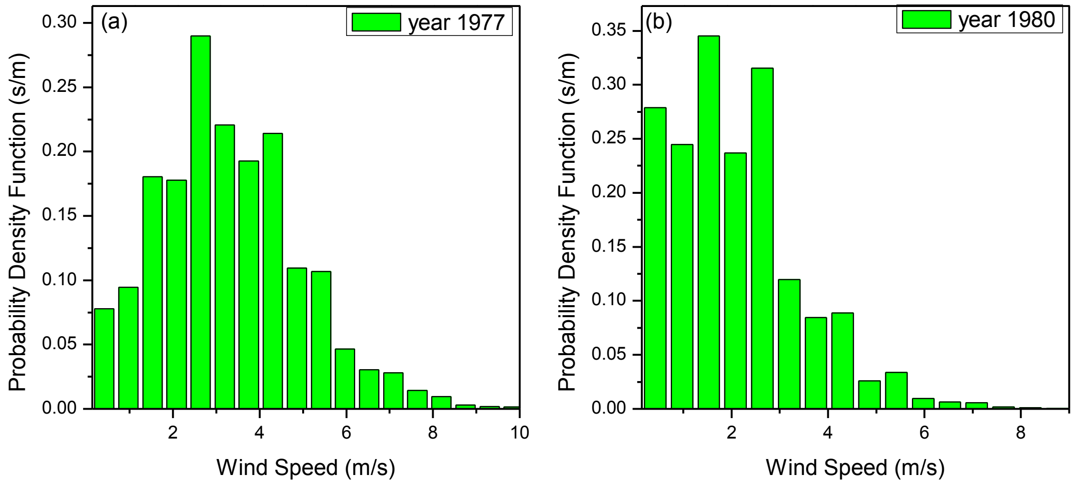

Figure 3.3: Histogram for 1977 and 1980 taking 2 *km/h* class width

As can be seen from the Figure 3.3 that the double trend from the histograms are completely removed. However, the problem of uneven bars still exists in the histograms. They are similar to the erroneous histograms of 1970 and 1976 (plotted with 1 *km/h* class width) as shown in Figures 3.1 (b), and (d). The probable reason for the unevenness in the bars of the histograms is due to the conversion of decimal *knot* to integer *km/h*, as discussed in section 3.0. However, to be sure about above-mentioned argument, the wind speed data using Monte Carlo simulation (order of 10^6) have been simulated by taking arbitrarily $k = 2$, and $s = 9.72$ *knots* (5 *m/s*). The data have been initially sampled at 0.1 *knots* and then converted to integer *km/h*. The histograms of the simulated data with and without conversion are shown in Figure 3.4. It can be seen from the Figure 3.4 (a) that the uneven bars do exist in the histograms of converted data, whereas, it is absent in the actual histograms of wind speed data sampled in *knot*, as shown in Figure 3.4 (b). Therefore, it can be affirmed that the sampling error is due to the conversion of fractional *knot* to integer *km/h*.

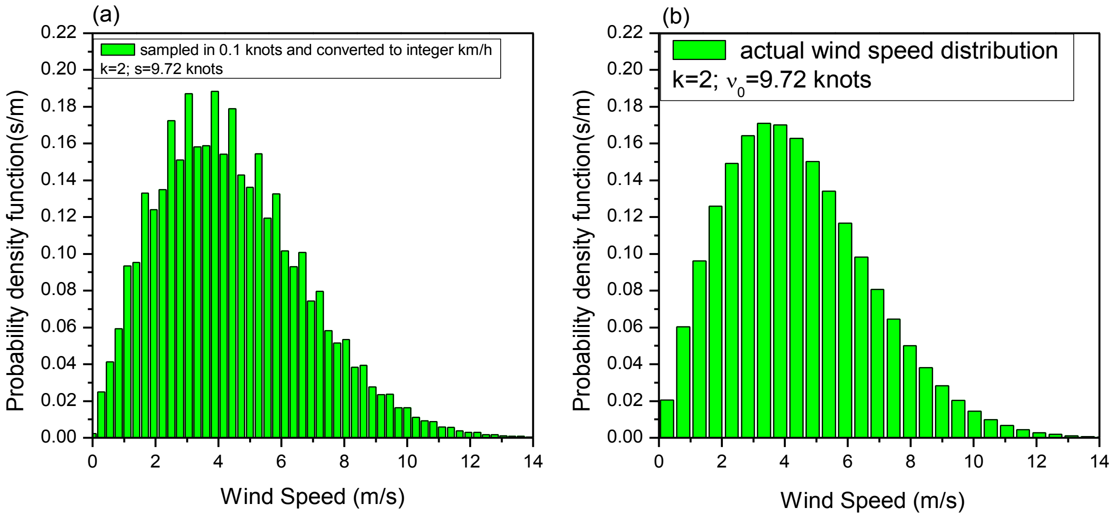

Figure 3.4: Histogram sampled in 0.1 *knot*

For studying the influence of sampling error in the hourly mean wind speed dataset of India, wind speed data have been simulated by inverse cumulative distribution function technique with arbitrarily Weibull parameter (k = 1.5 and s = 3 *m/s*). The wind speed data are initially generated in *knots* and sampled at 0.1 *knots* class width and later converted into integer *km/h*. The initial histogram with 0.1 *knots* class width as well as the histogram with 1 *km/h* class width is shown in Figure 3.5. It is clear from the Figure 3.5 that the *W.pdf* perfectly fit the histograms sampled at 0.1 *knots*. However, it shows a certain degree of biasness while fitting the histograms sampled with integer *km/h*. The extent of biasness needs to be studied, to calculate the unbiased Weibull parameters, which has been discussed in details in section 3.2.

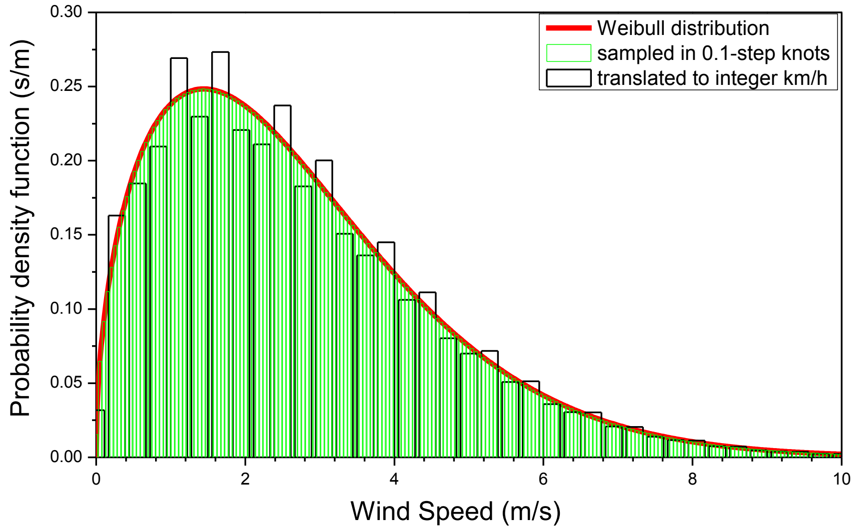

Figure 3.5: Plot of sampling error

3.2. Biased Estimation of Weibull Parameters

For estimation of biased Weibull parameters, initially, with Weibull parameters k = 0.5 to 2 and s = 3.88 (2 *m/s*) to 17.49 (9 *m/s*) *knot* (hereafter, these actual Weibull parameters ($k_{act.}$,$s_{act.}$) called as unbiased Weibull parameters ($k_{unbiased}$,$s_{unbiased}$)) in a step size of 0.1 wind speed of the order of 10^6 have been generated. These wind speed data have been truncated to first decimal places and then converted into integer *km/h*. The Weibull parameters have been estimated using these converted wind speed. The Maximum Likelihood Method has been employed to estimate these Weibull parameters. Hereafter, these estimated Weibull parameters ($k_{est.}$,$s_{est.}$) called as biased Weibull parameters (k_{biased},s_{biased}) as these are estimated from biased wind speed data. The ratio of biased to unbiased Weibull parameters have been estimated, and a contour plot have been plotted. The contour plots are the ratio of biased to unbiased Weibull parameters with unbiased shape and scale parameters on the y-, and x- axes of the contour plots (see Figure 3.6). The contour plots have been used to show the extent of biasness in estimating the Weibull parameters.

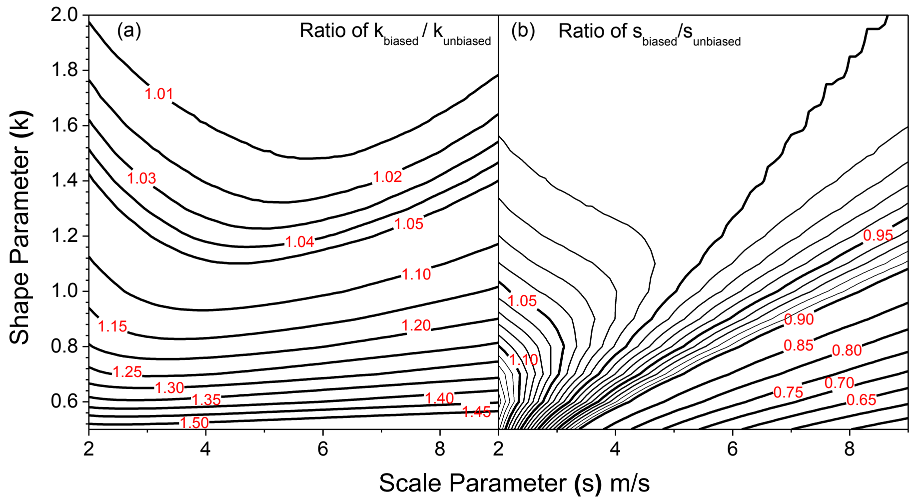

Figure 3.6: Contour Plot of Ratio of biased to unbiased Weibull Parameters

Figure 3.6 (a) reveals that the ratio of biased to unbiased shape parameters have a wide variation from 1.15 to 1.45 when $k_{unbiased}$ is less than 1 and as the $k_{unbiased}$ increases the variation decreases, and as $k_{unbiased}$ reaches to 2, the variation is limited to only 1%. Figure 3.6 (b) shows diagonal profile at a particular ratio of biased to unbiased scale parameters, i.e., at 1. At this diagonal line, s_{biased} is same as that of $s_{unbiased}$. At the right of this diagonal line the ratio is less than 1 that reveals that s_{biased} underestimate the $s_{unbiased.}$ However, on the left of the diagonal line, the s_{biased} overestimate the $s_{unbiased.}$

To infer the effect of biased Weibull parameters on pdf, an example has been studied with a certain value of Weibull parameters (k = 1 and s = 4 *m/s*). These are the unbiased Weibull parameters. Using unbiased Weibull parameters and contour plot as shown in Figure 3.6, biased Weibull parameters have been estimated. Using biased and unbiased Weibull parameters pdfs' have been estimated. Figure 3.7 shows the plot of biased and unbiased Weibull pdfs' and the ratio of biased to unbiased pdf versus mean wind speed.

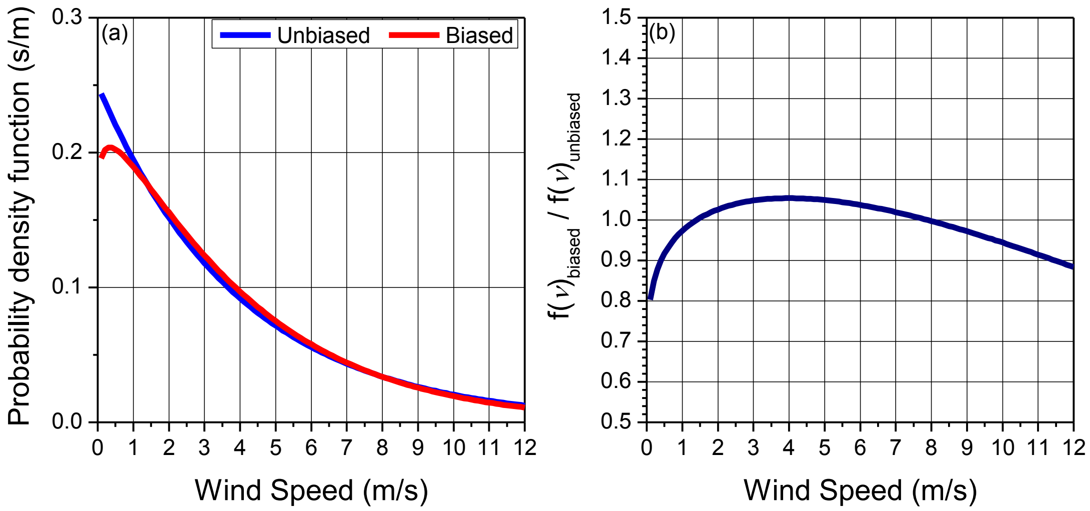

Figure 3.7: Comparison of biased and unbiased Weibull pdf and the plot of the ratio of biased to unbiased Weibull pdf versus wind speed.

Figure 3.7 (a) shows that the curve of biased pdf does not follow the curve of unbiased pdf. The extent of deviation has been studied, and the plot of the ratio of biased to unbiased pdf versus mean wind speed is shown in Figure 3.7 (b). According to this figure the mean wind speed range between 2 to 8 *m/s*, biased pdf overestimate the unbiased pdf. However, the overestimation is less than 10%. The wind speed outside this range, i.e., below 2 *m/s* or greater than 8 *m/s*, the biased pdf underestimate the unbiased pdf by 20%.

3.3. Unbiased Estimation of Weibull parameter

3.3.1 Inverse Contour Plot

As the wind data supplied by IMD, Pune is in integer *km/h*, the parameters estimated from these wind speed data are biased. Therefore, it is important to know the unbiased Weibull parameters. To estimate unbiased Weibull parameters from the biased one, an inverse contour plots have been plotted. It is the ratio of unbiased to biased Weibull parameters. To plot the inverse contour, initially biased Weibull parameters have been estimated from contour plot.

The 3-D plot has been constructed, with biased Weibull parameters are kept on x- and y-axes, whereas unbiased Weibull parameters (k = 0.5 to 4, s = 2 to 9 m/s; step size 0.1) on the z-axis. The six degree polynomial has been fitted to the 3-D plot, to achieve the $R^2 \geq 99.0\%$. This fitted polynomial function has been used to estimate the unbiased Weibull parameters. The ratio of unbiased Weibull parameters to biased Weibull parameters is plotted as the inverse contour plot. Figure 3.8 shows the inverse contour plot of unbiased to biased Weibull parameters.

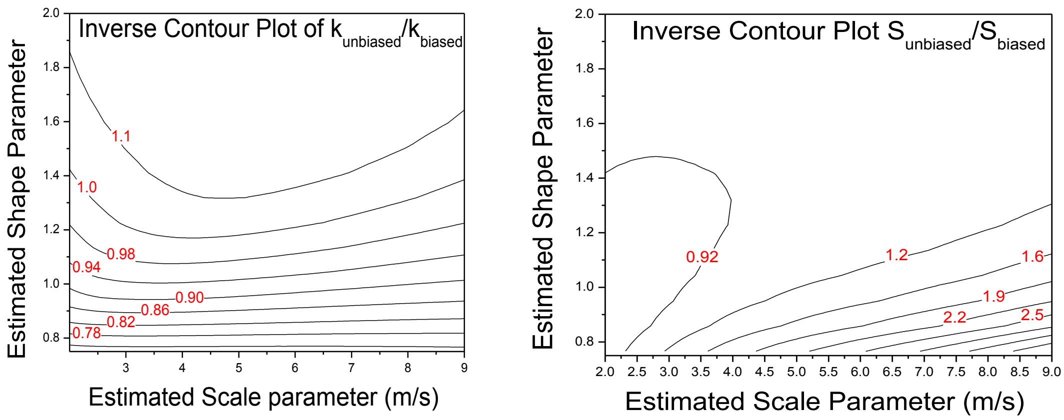

Figure 3.8: Inverse contour plot of the ratio of unbiased to biased Weibull parameters.

Figure 3.8 (a) reveals that the ratio of $k_{unbiased}$ to k_{biased} shows the wide variation from 0.38 to 0.75 below shape parameter 1. The variation of ratio decreases from 0.75 to 0.97 between shape parameter 1 to 2. From inverse contour plot, it is revealed that k_{biased} always overestimate the $k_{unbiased}$. The cause of overestimation may be due to the sampling error. The analogous behaviour can also be predicted from contour plot as shown in Figure 3.6 (a). Figure 3.8 (b) reveals that the s_{biased} underestimate the $s_{unbiased}$. The analogous behaviour can also be predicted from contour plot (see Figure 3.6 (b)). These inverse contour plots have further been utilized to enable the estimation of unbiased Weibull parameters for different locations of India.

3.4. Determination of Unbiased Weibull parameter for Different Stations

Figure 3.9 shows the histogram of probability density function of wind speed for Ahmedabad. The class width of these histograms is 1 *km/h* as IMD Pune supplied the wind speed data in integer *km/h* as shown in Figure 3.9 (a). Figure 3.9 (a) reveals that the pdf of observed wind speed, i.e., the wind speed supplied by IMD Pune show rounding error. This rounding error has been removed by increasing the class width to 2 *km/h* as shown in Figure 3.9 (b). However, to eradicate the sampling error, the biased Weibull parameters have been obtained from the observed wind speed data using Maximum Likelihood Method. From these biased Weibull parameters, unbiased Weibull parameters have been obtained using inverse contour plot using Figure 3.8.

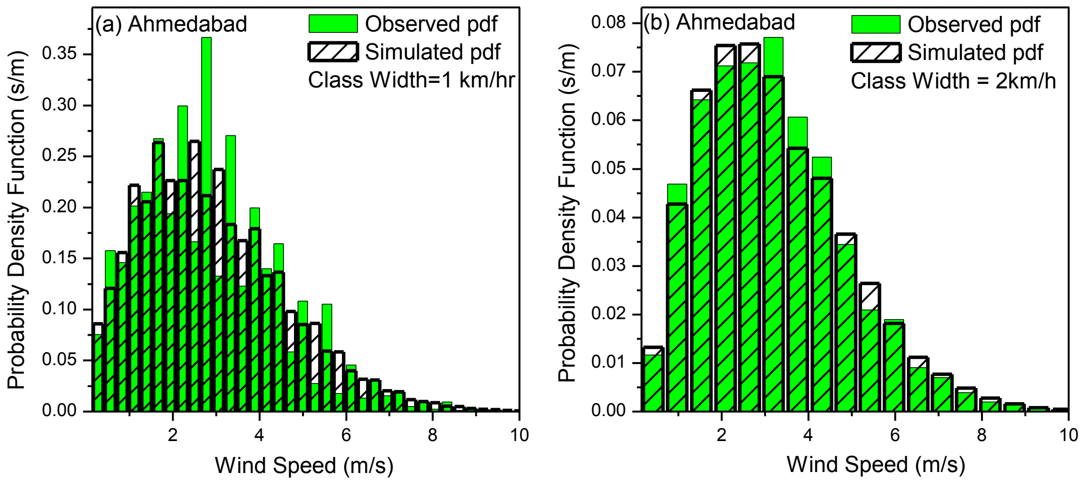

Figure 3.9: Probability density histogram of observed and simulated pdf at a class width of 1*km/h* and a 2 *km/h* for Ahmedabad.

The unbiased Weibull parameters are converted into *knot* and wind speed data equal to the length of observed wind speed data have been generated using inverse *CDF* method. These wind speeds are truncated to 1st decimal places and then converted into *km/h* by multiplying with the factor of 1.852. The obtained wind speeds are then rounded to integer *km/h* and are called simulated wind speed. As these simulated wind speeds replicate the observed wind speed, the size of simulated wind speed is kept equal to the size of observed wind speed data so that both the data are bought at a common platform. The simulated wind speed is again classified into a class width (bins size) of 1 *km/h*, and 2 *km/h* histogram. The simulated histogram is superimposed on the observed histogram as shown in Figure 3.9. It can be seen from the Figure 3.9 (b) that the simulated histogram is exactly following the observed histograms at the parent side. However, it is not at all understood whether the upper extreme tails are also following the *W.pdf* or not. Therefore, to estimate the extent of wind speed data following *W.pdf*, a plot *1-F(v)* versus wind speed has been plotted on the log scale as shown in Figure 3.10. If the observed distribution followed the simulated distribution completely, the observed wind speed follow *W.pdf*. However, the point of deviation of the observed distribution is the point on the abscissa up till that the observed distribution follows the *W.pdf*.

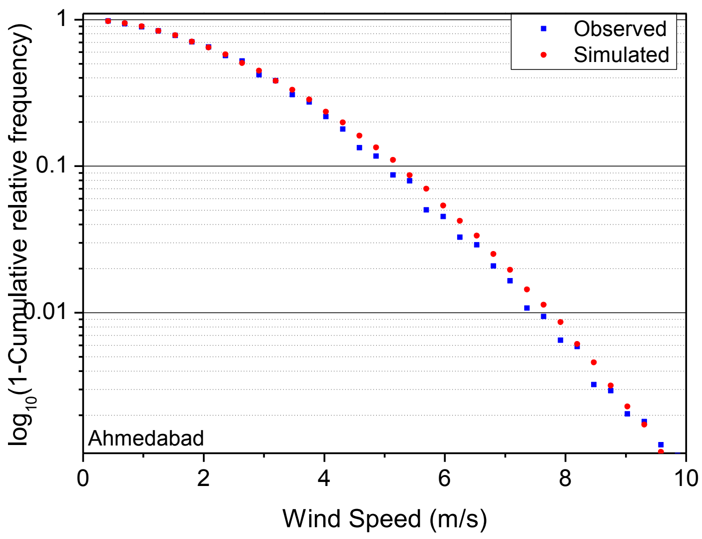

Figure 3.10: Probability distribution of wind speed for Ahmedabad.

Figure 3.10 reveals that the upper tail does not follow the *W.pdf*. However, the exact range of wind speed which can be described by the *W.pdf* varies from station to station. Table 3.1 shows the 40 meteorological stations of India their geographical locations as well as an observation period. Their biased, unbiased Weibull parameters and range of wind speed for *W.pdf* are listed in Table. 3.2.

Table 3.1: Different meteorological stations of India.

Stations	Latitude (^{0}N)	Longitude (^{0}E)	Altitude (m)	Observation Period
Madras Harbour	12.99	80.18	16	January 1969-January 1988
Tuticorin H.P.	8.72	78.03	39	January 1969-June 1983
Amritsar	31.63	74.87	234	January 1969-December 2005
Palam (A)	28.57	77.10	237	January 1969-December 2005
New Delhi / Safdarjung	28.61	77.21	216	January 1969-December 2005
Jaipur / Sanganer	26.92	75.82	432	January 1969-December 2002
Lucknow / Amausi	26.76	80.89	125	January 1969-December 2000
Baghdogra (A)	26.68	88.33	126	January 1975-December 1987
New Kandla	23.03	70.22	3	January 1969-October 1991
Ahmedabad	23.03	72.58	53	January 1969-October 1998
Bhopal Bairagarh	23.25	77.42	427	January 1969-September 1990
Jamnagar (A)	22.47	70.01	21	July 1969-September 1984

Table 3.1: Continued

Baroda	22.30	73.19	39	January 1969-November 1991
Indore	22.42	75.54	553	June 1972-December 1998
Kalaikunda (A)	22.34	87.21	61	January 1969-December 1982
Nagpur / Sonegaon	21.07	79.27	310	January 1969-December 1998
Raipur	21.23	81.63	298.15	October 1969-January 1993
Veraval	20.90	70.37	0	January 1969-October 1998
Ahmedabad / Santacruz	19.09	72.84	14	January 1969-September 2006
Jagdalpur	19.07	82.03	552	January 1969-July 1979
Gopalpur	19.27	84.92	1	January 1969-December 1980
Ahmedabad	18.98	72.83	14	January 1969-December 2006
Pune	18.53	73.85	560	January 1969-December 2006
Pune (A) / Lohagaon	18.58	73.92	592	January 1969-December 1982
Hyderabad (A)	17.37	78.48	536	January 1969-October 1998

Table 3.1: Continued

Hakimpet (A)	17.55	78.52	616	April 1973-December 1984
Vishakhapatnam (A)	17.72	83.22	5	January 1969-May 1990
Mormugao	15.25	73.98	2	January 1969-January 2005
Madras / Minambakkam	13.09	80.27	6	January 1969-December 2005
Mangalore H.P. / Panamb	12.87	74.88	22	January 1973-December 2000
Bangalore	12.97	77.56	920	January 1969-December 2000
Bangalore (A)	13.20	77.71	915	January 1969-December 2000
Port Blair	11.67	92.76	0	January 1969-December 2000
Kodaikanal	10.24	77.49	2133	January 1969-December 2005
Tiruchirapalli (A)	10.77	78.71	88	January 1969-December 2005
Cochin (N.A.S.)	9.97	76.22	0	January 1969-May 1997
Trivandrum/ Tiruvananthapuram	8.5	76.9	5	January 1973-December 2003
Trivandrum (A)	8.48	76.85	4	January 1973-October 2002

Table 3.1: Continued

Tuticorin	8.81	78.14	9	July 1983-December 2005
Calcutta / Dum Dum	22.62	88.42	11	January 1969-December 2000

Table 3.2: Numerical values of Weibull parameters for hourly mean wind speed data for 40 locations in India.

Stations	**Biased**	**Unbiased**	**Biased**	**Unbiased**	**Wind Speed Range**
	Shape Parameter *(k)*		**Scale Parameter *(s) m/s***		**Weibull Distribution Validity**
Madras Harbour	1.86	1.84	5.00	5.01	0-10
Tuticorin H.P.	2.33	2.34	6.49	6.59	0-14
Amritsar	1.29	1.13	2.64	2.23	0-8
Palam (A)	1.68	1.62	3.41	3.39	0-10
New Delhi / Safdarjung	1.46	1.35	2.68	2.51	0-10
Jaipur / Sanganer	1.37	1.22	2.14	1.94	0-6
Lucknow / Amausi	1.47	1.38	3.25	3.05	0-10
Baghdora (A)	1.55	1.48	3.26	3.15	0-8

Table 3.2: Continued

New Kandla	1.99	1.98	6.41	6.40	0-10
Ahmedabad	1.83	1.79	3.22	3.28	0-6
Bhopal Bairagarh	1.79	1.75	4.31	4.30	0-10
Jamnagar (A)	1.88	1.86	5.37	5.36	0-8
Baroda	1.62	1.53	2.72	2.66	0-6
Indore	2.02	2.00	5.75	5.77	0-8
Kalaikunda (A)	1.95	1.94	3.99	4.09	0-8
Nagpur / Sonegaon	1.40	1.26	2.23	2.01	0-6
Raipur	1.33	0	1.31	0	0-8
Veraval	1.72	1.67	5.05	4.93	0-14
Ahmedabad Santacruz	1.76	1.71	3.19	3.24	0-9
Jagdalpur	1.51	0	1.86	0	0-7
Gopalpur	1.61	1.54	4.62	4.46	0-8
Ahmedabad	1.51	1.41	2.83	2.64	0-10
Pune	1.63	1.54	3.05	2.96	0-11
Pune (A) Lohagaon	1.69	1.64	4.78	4.68	0-14
Hyderabad (A)	1.46	1.36	3.03	2.83	0-12

Table 3.2: Continued

Hakimpet (A)	1.79	1.75	4.77	4.74	0-8
Vishakhapatnam (A)	1.58	1.51	4.14	4.00	0-7
Mormugao	1.34	1.17	2.45	2.05	0-11
Madras / Minabakkar	1.47	1.37	3.01	2.82	0-12
Manglore H.P. Panamb	1.25	0	1.94	0	0-8
Banglore	1.87	1.84	2.84	2.99	0-8
Banglore (A)	1.81	1.76	3.81	3.83	0-10
Port Blair	1.64	1.56	3.97	3.83	0-6
Kodaikanal	1.84	1.79	3.66	3.69	0-11
Tiruchirapalli (A)	1.74	1.69	4.89	4.78	0-12
Caochin	1.33	1.17	2.76	2.37	0-12
Trivandrum/ Thiruvananthapuram	1.42	1.29	2.84	2.54	0-10
Trivandrum (A)	1.89	1.86	4.52	4.57	0-4
Tuticorin	1.97	1.96	5.33	5.38	0-14
Calcutta / Dum Dum	1.25	1.09	2.30	1.92	0-10

3.5 Influence of Sampling Error for Estimation of Wind Power Density

Figure 3.11 shows the influence of sampling error on the wind power density. According to this figure, there is a significant difference between biased *WPD* to unbiased *WPD* when shape parameter is less than 1. The biased *WPD* maximum underestimates the unbiased *WPD* by 90% and this underestimation decreases as the shape parameters increases. However, when the shape parameters greater than 1.5. The percentage difference of estimation is limited to only 1%.

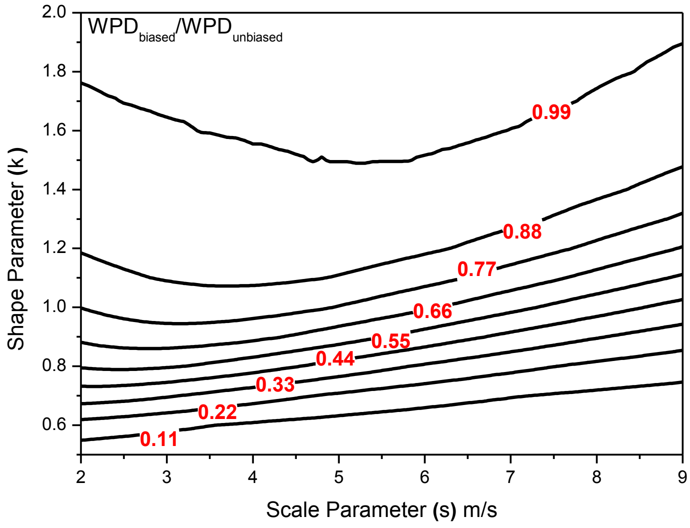

Figure 3.11: Contour plot of wind power density.

3.6 Summary

The wind data provided by IMD Pune are biased, the reason for their biasness is that the wind speed data are recorded in decimal *knots* up to Ist place and then converted into integer *km/h*, this sampling technique induces two types of errors namely rounding and sampling error in the wind speeds. The rounding error can be removed by increasing the size of the class width. The optimum class width has been found to be 2 *km/h*. This 2 *km/h* class width enables the removal of the double trend from the histograms, without much information loss from the data. The sampling error is the cause of uneven bars in the histograms. As a result, the Weibull parameters estimated from the histograms are biased. In the present scope of work, the extent of biasness and the contour plot to estimate the unbiased Weibull parameters has been discussed in details. Using these contour plot, the unbiased Weibull parameters can be estimated. The contour plot of Weibull parameters reveals that the sampling error has the significant impact on Weibull parameters when the shape parameters are less than unity. However, its effect decreases as the shape parameters increase and almost negligible for shape parameter 2 and above.