Proceedings of the Pakistan Academy of Sciences 50 (4): 327–333 (2013)

ISSN: 0377 - 2969 (print), 2306 - 1448 (online)

Pakistan Academy of Sciences

Research Article

Statistical Analysis of Cloud Cover at Pakistan Coastal Regions

Faisal Ahmed Khan Afridi[1], M. Ayub Khan Yousuf Zai[2*] and Khusro Mian[1]

[1]Institute of Space and Planetary Astrophysics (ISPA), University of Karachi, Karachi-75270, Pakistan
[2]Department of Applied Physics, Solar-Terrestrial & Atmospheric Research Wing and Institute of Space & Planetary Astrophysics (ISPA), University Karachi, Karachi-75270, Pakistan

Abstract: This study presents a descriptive statistical analysis of cloud cover using monthly data for Sindh and Makran coasts recorded by the Pakistan Meteorological Department from 1980 to 2004. For Sindh coast, we are considered the data for Karachi and for Makran, we have taken the case of Quetta. For this purpose, Exploratory Data Analysis (EDA) approach was utilized. Through this approach, analysis of the data illustrated variability in the behavior of the cloud dynamics for both the regions. The degree of variation in the computed parametric values at Karachi region was remarkable as compared to the Quetta region. Interpretations of the temporal character depict that the fluctuations in the cloud concentration of Quetta region is different from Karachi due to coastal location.

Keywords: Cloud, cloud cover, atmosphere, climatology, descriptive statistics, exploratory data analysis

1. INTRODUCTION

It has been noticed that the clouds manifest an important role to maintain life on the Earth's biosphere. Moreover, cloud density, overall size, altitude, shapes, area and other microphysical properties have significant impact on lower altitude atmosphere. Because of their low altitude, where temperature is comparatively high, these clouds reflect long-wave radiation towards the upper space. On the other hand, higher altitude clouds are comparatively thin, reflect lesser incoming radiations, and create a warming effect with emission of a small amount of infrared radiations

It has been observed that variations in the cloud cover certainly change the balance in Earth's energy budget. Cloud climatology describes the diurnal, seasonal, and inter-annual variations. On an annual average basis, clouds coverage area is between 60–65% of the Earth. Normally it is considered that cirrus clouds contribute globally to a warming of the atmosphere due to their contribution to downward transfer of long wave radiation, like a greenhouse agent [1-5].

A systematic research work for each month's data on cloud cover for Karachi and Quetta region for a relatively long period has never been reported. In the present work, a systematic study of the variability of cloud cover at Karachi and Quetta regions, using descriptive statistics and exploratory data analysis (EDA), has been carried out for the period 1980–1984. The significance of this study is enormous because cloud cover has an apparent effect on the temperature forecast and atmospheric circulation.

2. EFFECTS OF CLOUD COVER

Thick and continuous cloud cover forms a significant barrier to the penetration of solar radiation. How much radiation is actually reflected by clouds depends on the amount of cloud cover and its thickness. It also serves to retain much of the heat that would otherwise be lost from the earth by long-wave radiations. Earth-space transmissions

Received, June 2013; Accepted, September 2013
*Corresponding author: M. Ayub Khan Yousaf Zai; Email: ayubzai@yahoo.com

propagate through two important atmospheric regions: the troposphere and the ionosphere. The troposphere contains most of the Earth's weather effects in terms of clouds, e.g. rainfall and snow. The dual effects of the troposphere and the ionosphere can give rise to the communication signal attenuation caused by atmospheric gases, clouds, precipitation, sand and dust. Attenuation of radio wave signals by clouds is similar to the attenuation by rainfall droplets Attenuation by clouds depends on their characteristics such as, type, thickness and coverage. The clouds affect the radio wave propagation by absorbing and scattering the wave: a process described by Rayleigh scattering. Transmission of solar radiation through cloud coverage is measure of its layers and density [6-8].

3. METHODOLOGY

The cloud cover data was obtained by Pakistan Meteorological Department, Government of Pakistan by in-situ observations recorded in accordance with the World Meteorological Organization (WMO) standards. This study comprised of data analysis for cloud cover of Exploratory Data Analysis (EDA). Histogram is one of the oldest and most frequently used tools of the data analyst for the investigation of the overall distribution of a sample. Skewed histograms are not symmetric and can be either positively or negatively skewed. When a histogram is either positively or negatively skewed, the mean and the standard deviation may not be the appropriate measures of the center and spread. Use the median as the measure of the center and the inter-quartile range as the measure of the spread [9-10].

A normal probability plot is a scatter plot that compares observed and theoretical values much in the same manner. On the vertical axis, we find percentile values for a theoretical normal distribution sharing the same mean and standard deviation as the empirical data. The actual data values appear on the horizontal axis. If the two distributions matched perfectly, plotted points would fall on a straight line with a slope of 1.0, a line rising from the origin at a 45^{o} angle [11].

A probability plot indicates how the distributions differ from each other. In the literature there are two types of probability plots, P-P plots and Q-Q plots. These plots can be best explained as cumulative distribution functions (CDF). The P-P plot is a simple & informative method for drawing such a comparison. It is used to determine how well a specific distribution fits to the observed data. By using a univariate or bivariate scaling of multivariate data, a P-P plot can also be used for comparing two multivariate samples with each other [12, 13].

The attraction of using the mean is that it is a single summary measure which is easy to calculate and readily understood. The median is particularly attractive when the data set is skewed, when the observations tend to be concentrated at one end of the range of values. The most important measures of variability are range and modified ranges, mean deviation, standard deviation, variance and coefficient of variation. The measure of variability appropriate to a problem depends on which measure of central location is used. The standard deviation, mean deviation and variance would all be used in association with the mean, while the inter-quartile range would be used in association with the median. [14, 15].

The difference between the third and first quartiles is called the inter-quartile range (IQR). In some sense, the standard deviation, the range and the inter-quartile range provide measures of spread of the sample. We can calculate the inter-quartile range

$$\text{Inter-quartile Range} = q_3 - q_1$$

Where q_3 is the 3rd quartile and q_1 is the 1st quartile [16].

It has been said that the coefficient of variation (CV) is the measure that describes the scatter of the distribution relative to the size of the estimated mean and the diversity of the data from normality. The coefficient of variation is a measure of relative dispersion. It is appropriate for comparing the variability within different data sets. It determines the stability and consistency of the data and is given by

$$CV = (\text{Standard Deviation} / \text{Mean}) \times 100$$

It has been indicated that the extremely low value of the calculated CVs about 7.0% that confirms a good degree of normality. From the above analysis it is shown that the standard error of the mean, is a method used to estimate the standard deviation of a sampling distribution. The error in the mean computed from a sample set of measured values that result because all measured values contain errors.

The standard error of the mean is computed by:

$$\text{Standard Error of Mean} = S / \sqrt{n}$$

where S is the sample standard deviation and n is the number of observations of the sample [14, 17].

Since the mean is arrived at by using all the values being examined, it can be distorted by extreme values (or so-called outliers). For this reason it is often appropriate to calculate a trimmed mean. The trimmed mean gives a robust measure of central location which is therefore not affected by a few extreme values. Trimmed means are often very useful and many times are used with a smaller trimming percentage. The mean, median and the trimmed mean are measures of the middle of the sample. We can calculate trimmed mean by the following equation:

$$\text{Trimmed Mean} = (q_1 + 2q_2 + q_3) / 4$$

Where q_1 is the 1st quartile, q_2 is the 2nd quartile and q_3 is the 3rd quartile [15-17].

Skewness is a measure of the degree of asymmetry of a distribution. When a distribution is symmetrical about the mean, it is equal to zero. A measure of skewness is defined by

$$E[(X-\mu)^3] / \sigma^3$$

4. RESULTS AND DISCUSSION

In this section the various parameters have been computed for this data at 1200 UTC during 1980 to 2004 for both the regions. Analyzing data with the use of descriptive approach show clear distinction between the behavior of data sets for both the regions. Table 1 exhibits comparison of descriptive parameters of monthly mean total cloud covers of Karachi and Quetta regions during 1980 to 2004. It is to be noted that, an "Okta" is a unit used in meteorology indicating the presence of clouds, ranging from 0 to 8 Oktas.

Table 1. Monthly mean total cloud cover at 1200 UTC (Oktas), at *(A) Karachi region, *(B) Quetta region.

S.No.	Parameter	* (A)	*(B)
01	Total No. of Observation	300	300
02	Mean	2.582	2.800
03	Median	2.150	2.890
04	Stand. Dev.	1.766	1.457
05	Variance	3.119	2.125
06	Quartile (Q1) 25%	1.300	1.600
07	Quartile (Q3) 75%	3.500	3.875
08	Range	7.500	7.900
09	Interquartile Range	2.200	2.275
10	Semi-interquartile Range	1.100	1.137
11	Minimas	0	0.1
12	Maximas	7.5	8.0
13	Coefficient of Variation	68.40	52.22
14	Standard Error Mean	0.102	0.084
15	Trim Mean	2.490	2.764
16	Skewness	0.82	0.20

The calculated value of mean of monthly total cloud cover during 1980 to 2004 at Karachi region is 2.582, which is smaller value than the calculated value of mean of Quetta region i.e., 2.800. The computed value of median of the monthly total cloud cover at Karachi region is 2.15, which is smaller value than the calculated value of median of Quetta region, i.e., 2.89. It is interesting that the calculated value of mean (2.80) and median (2.89) is approximately same for the monthly total cloud cover of Quetta region during 1980 to 2004, while Karachi region has different values of mean (2.58) and median (2.15). So the data of monthly total cloud cover of Quetta region has a symmetrical distribution.

The data set of cloud cover at Karachi region has a larger standard deviation, i.e., 1.766 than the data set of cloud cover at Quetta region, i.e., 1.457. Thus, the data set for Karachi region was less precise than the data set for Quetta region for symmetrical distribution. We calculated the variance of cloud cover for Karachi region which was 3.049 while the variance of cloud cover of Quetta region which was

2.10. This also revealed that the data of cloud cover of Karachi region is more dispersed than the data of Quetta region.

It can be seen from Table 1 that the data set could be described as having either a mean of 2.58 and a standard deviation of 1.766 or a median of 2.15 and an interquartile range of 2.2 or semi-interquartile range of 1.1. This means that approximately 50% of the observations lie within the range 2.2 ± 1.1 of the data of cloud cover at Karachi region. The values of median and interquartile range at Karachi region were very close to each other.

The minimum value for monthly total cloud cover during 1980 to 2004 at Karachi region was zero, while the minimum for monthly total cloud cover during 1980 to 2004 at Quetta region was 0.1. Thus, during the 30-year period, there was no cloud cover during certain days of a few months at Karachi whereas Quetta region had a minimum 0.1 Okta value, which indicates that a minimum amount of cloud cover was always present there during the 1980 to 2004 period.

The calculated value of coefficient of variation of monthly total cloud cover during 1980 to 2004 at Karachi region is 68.4% while Quetta region has 52.22%, which clearly shows that the standard deviation is quite high relative to the mean and thus the data show considerable variation around the mean of monthly total cloud cover at Karachi region. The Standard Error of the Mean of monthly total cloud cover at Quetta region is 0.084 whereas Karachi region has 0.102, therefore the standard errors of Mean of both regions are nearly zero.

We have calculated the trimmed mean of monthly total cloud cover at Karachi region is 2.49 while calculated mean is 2.58. The results show that the distribution of the data is positively skewed as the mean is greater than the median. Trimming the mean removes the extreme values and in this case gives a value equal to the mean, so we can say that the data of Karachi region has no extreme values or outliers.

The value of skewness of cloud cover at Quetta region is 0.20 that is small as compared to the Skewness value (0.82) of cloud cover at Karachi region. Both skewness values are positive.The skewness of Quetta region is near to zero. For the descriptive parameters regarding cloud cover data (1980 – 2004) of Pakistan coastal zones (Karachi region and Quetta region) and the corresponding histogram, as pointed out in Fig. 1 and Fig. 2, respectively. The histogram illustrates that the data of cloud cover at Karachi region are skewed to the right or positively skewed. While the other histogram illustrates that the data of cloud cover at Quetta region are symmetric about the center but has a steeper slope than Karachi region of Fig. 1, with a higher peak for its central value.

The vertical axes represent the normal quartiles and the observed observations are given on the horizontal axes. The normal probability plot for cloud cover data during 1980 to 2004 of Karachi region and Quetta region are depicted in Fig. 3 and Fig. 4, respectively. Normal probability plot exhibits that the data is half Gaussian. The normal probability plot of Karachi region indicates that the distribution has shorter (lighter) tails than the normal distribution. On the other hand, the normal probability of Quetta region is normal which signify that the distribution is almost linear in nature. It is noticeable in the plot that there is minor bend at the ends, representing some departure from normality.

The P-P plots for cloud cover data during 1980 to 2004 of Karachi region and Quetta region are represented in Fig. 5 and Fig. 6, respectively. A graphical representation of P-P plot can help us to decide whether our data sets are following Normal distribution or not. The P-P plot of Quetta region is approximately linear, so it follows normal distribution. The P-P plot of Karachi region has an apparent variation as shown in the curve. The Q-Q plot of Karachi region exhibits half Gaussian as depicted in Fig. 7 while Quetta region illustrates almost half Gaussian with an outlier. The reason of this point will be investigated further.

Our data are illustrated in Fig. 7, and Fig. 8 that depict a set of data appears to come from a particular probability distribution or not. The illustration clarifies probability of each value of the depth of cloud cover in its calculated volume. It is apparent that the Q-Q plot of Karachi region is periodic in nature. It is clear that the data for Q-Q plot for the cloud cover at Quetta region are

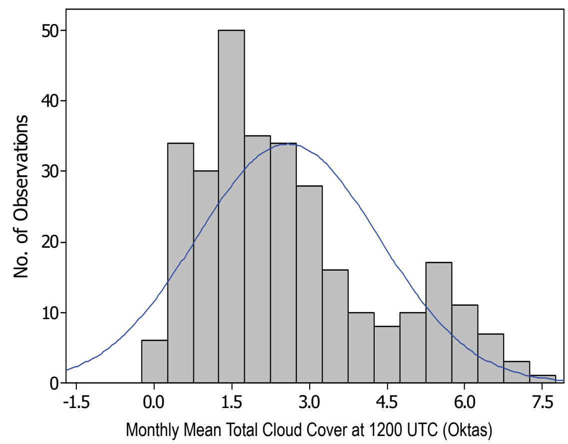

Fig. 1 Cloud cover at Karachi region during 1980 to 2004.

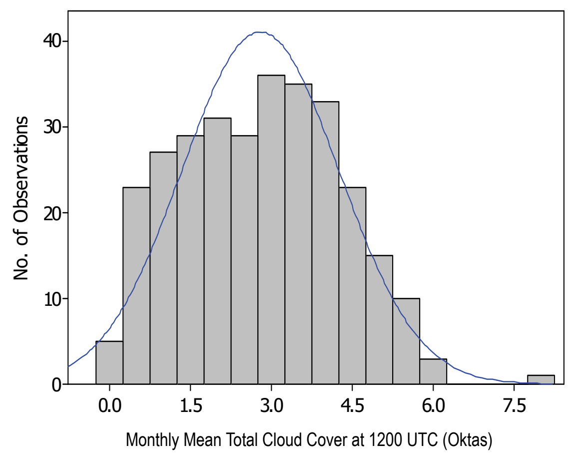

Fig. 2 Cloud cover at Quetta region during 1980 to 2004.

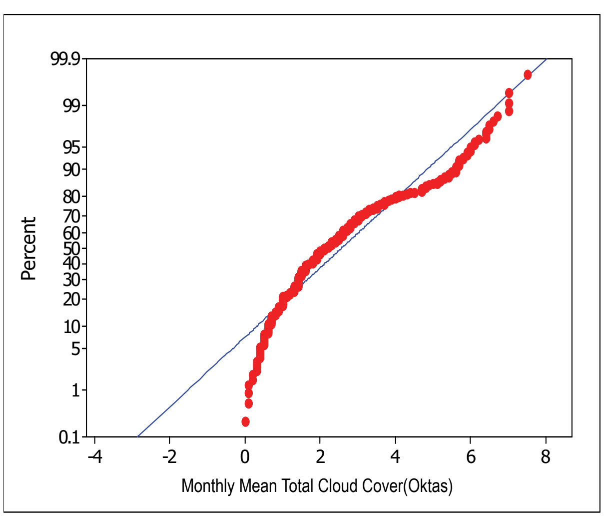

Fig. 3 Normal probability plot of cloud cover at Karachi region during 1980 to 2004.

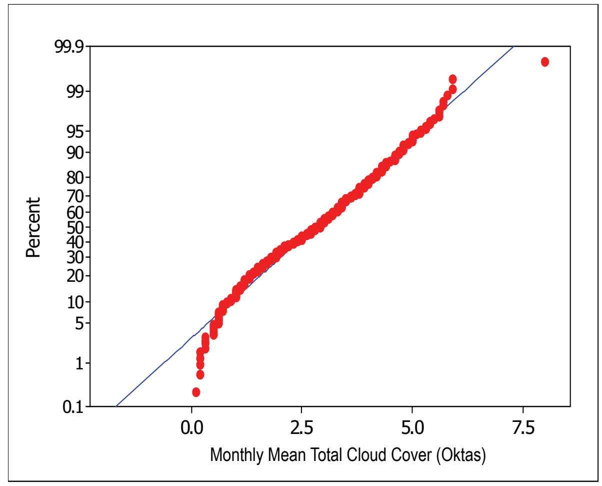

Fig. 4 Normal probability plot of cloud cover at Quetta region during 1980 to 2004.

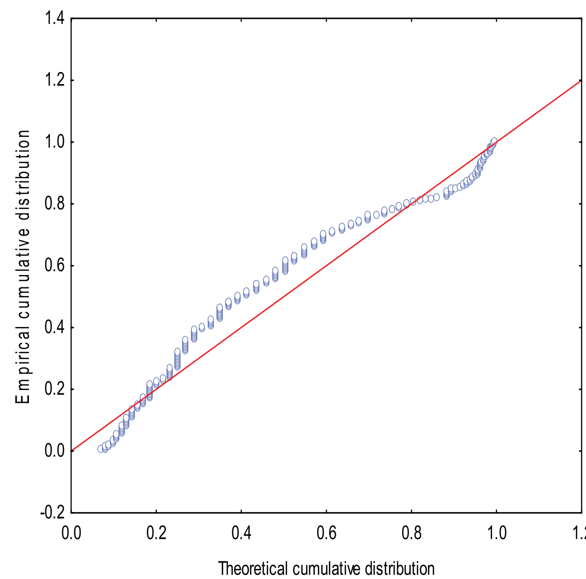

Fig. 5 Probability-probability plot of cloud cover at Karachi region during 1980 to 2004.

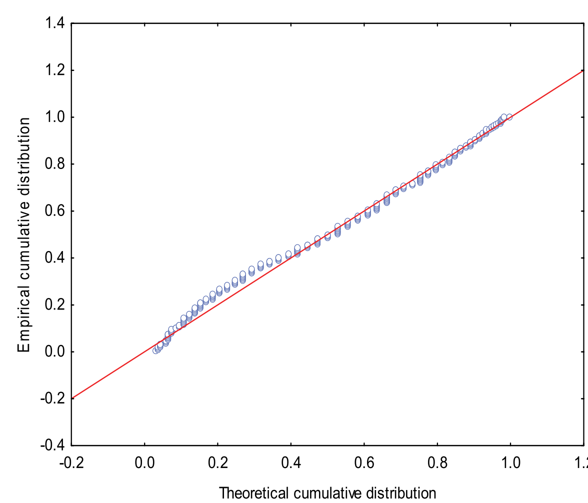

Fig. 6 Probability-probability plot of cloud cover at Quetta region during 1980 to 2004.

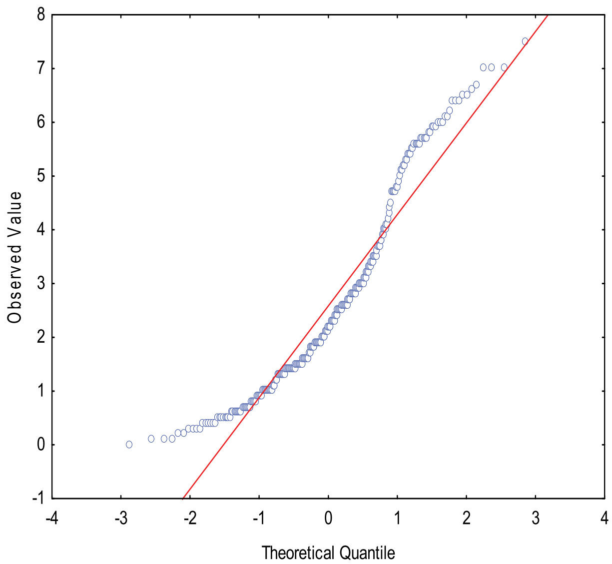

Fig. 7 Quantile-quantile plot of cloud cover at Karachi region during 1980 to 2004.

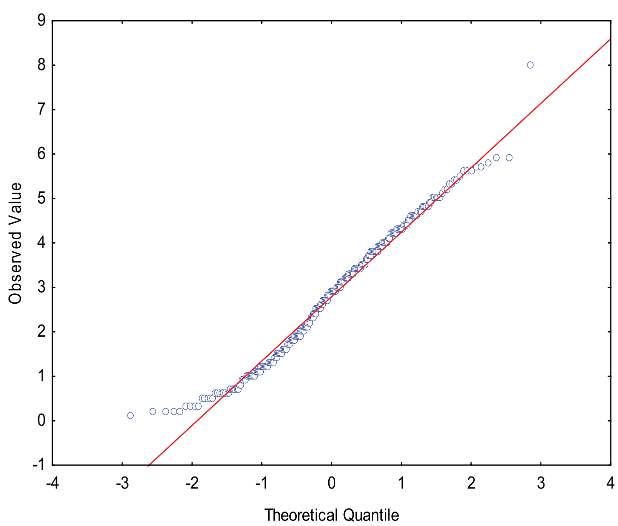

Fig. 8 Quantile-quantile plot of cloud cover at Quetta region during 1980 to 2004.

normally distributed because almost all points lie on a straight line. We can see in the plot of Quetta region that there is slight curvature at the ends, indicating some departure from normality.

5. CONCLUSIONS

The monthly data of total cloud cover for Quetta region revealed a symmetrical distribution, showing almost a linear behavior. However, the total cloud cover data for Karachi region indicated considerable variation, compared with the Quetta region. Comparative analysis for the 1980–2004 period also revealed no cloud covers in Karachi region during the month of October 1987, whereas a minimum amount of cloud cover always existed in Quetta region during the 1980 to 2004 periond. The standard error of mean of both regions was negligible. The distribution of the data was positively skewed as the mean was greater than the median. Skewness of the data was more for Karachi region than for Quetta region.

The data of cloud cover at Karachi region are skewed to the right or positively skewed, while the data of cloud cover at Quetta region are symmetric but has a reasonable slope than Karachi region. The normal probability plot of Karachi region indicates that the distribution has shorter (lighter) tails than the normal distribution, while the normal probability of Quetta region is normal which signify that the distribution is almost linear in nature. It is noticeable in the plot that there is minor bend at the ends, representing some departure from normality. The P-P plot of Quetta region is approximately linear. Therefore, it will be an appropriate thus admits normal distribution. The P-P plot of Karachi region has an apparent variation as shown in the curve. The Q-Q plot of Karachi region exhibits half Gaussian while Quetta region illustrates almost half Gaussian with an outlier. The reason of this point will be investigated further. The information attained by this analysis can be utilized for the researchers, personnel of Pakistan Meteorological Department, Pakistan Space and Upper Atmosphere Research Commission (SUPARCO) and also in the flight operations by Civil Aviation Authority including Pakistan Air Force.

6. ACKNOWLEDGEMENTS

Authors of this communication immense gratified to the technical staff members of Pakistan Meteorological Department, Karachi, Pakistan, who provides us the climatological data and other parameters. The authors also express thanks the anonymous referees of the paper for their supportive suggestions.

7. REFERENCES

1. Salby, M. L. *Fundamentals of Atmospheric Physics*. Academic Press, CA, USA, p. 36-39 (1996).
2. Horst, M. Comparison of mean cloud cover obtained by Satellite photographs and ground-based observations over Europe and the Atlantic. *Monthly Weather Review.* 101(12): 893–897, doi: 10.1175/1520-0493(1973)101<0893:COMCCO>2.3.CO;2 (December 1973).
3. http://www.climate4you.com/climateAndclouds.htm
4. Upgren, A. & J. Stock. *Weather: How it works and why it matters.* Perseus Publishing, MA, USA, p. 28-29 (2001).
5. Heintzenberg, J. & R. J. Charlson. *Clouds in the Perturbed Climate System.* The MIT Press, Cambridge, USA, p. 357-359 (2009).
6. Barry, R. G. & R. J. Chorley. *Atmosphere, Weather and Climate, 8th ed*. Routledge, NY, USA, p. 39-99 (2003).
7. Jones, R. W. *Handbook on Satellite Communications, I.T.U., 3rd ed*. John Wiley & Sons Inc., p. 939-943 (2002).
8. Prasad, A. A. *Investigation of the Effects of Cloud Cover on Ku-Band Satellite Link and Earth's Potential Gradient*, M.S Thesis, Division of Physics, School of Engineering and Physics, Faculty of Science, Technology and Environment, University of the South Pacific, Suva, Fiji Islands, p. 3-4 (2009).
9. Murphy, A. H. & R. W. Katz. *Probability, Statistics and Decision Making in the Atmospheric Sciences*. West view Press, Colorado, USA, p. 4-13 (1985).
10. Brightman, H. J. *Data Analysis in Plain English with MS Excel.* Brooks/Cole Publishing Company, Canada, p. 32-33 (1999).
11. Carver, R. H. *Doing Data Analysis with Minitab 14.* Thomson Brooks/Cole Publishing Company, Canada, p. 109-110 (2004).
12. Murphy, A. H. & R. W. Katz. *Probability, Statistics and Decision Making in the Atmospheric Sciences*. West View Press, Colorado, USA, p. 4-13 (1985).
13. Toit, S. H. C. d., A. G. W. Steyn & R. H. Stumpf. *Graphical Exploratory Data Analysis.* Springer-Verlag, New York, NY, USA, p. 283-285 (1986).
14. Ghilani, C. D. & P. R. Wolf. *Adjustment Computations: Spatial Data Analysis.* 4th ed. John Wiley & Sons Inc., NJ, USA, p. 13-21 (2006).
15. Fleming, M. C. & J. G. Nellis. *Principles of Applied Statistics.* Routledge, London, U.K., p. 40-60 (1994).
16. Hogg, R. V., E. A. Tanis & J. M. Rao. *Probability and Statistical Inference, 7th ed*. Dorling Kindersley (India) Pvt. Ltd., New Delhi, India, p. 91-474 (2007).
17. Walpole, R. E., R. H. Myers, S. L. Myers & K. Ye. *Probability & Statistics for Engineers & Scientists, 9th ed*. Pearson Education Inc., Boston, USA, p. 12-277 (2012).