

**Performance probability distribution function for modelling
Solar Radiation in South southern Nigeria:
A Case Study of Yenagoa.**

Abstract

This study has attempted to assess the performance of the most suitable statistical distribution function for modelling solar radiation over Yenagoa, Bayelsa State in Nigeria. The probability distribution functions are tested based on eleven years (2007-2017) solar radiation data obtained from National Aeronautics and Space Administration (NASA). Six probability distribution functions are tested to ascertain the most appropriate one based on four different statistical tools and fitting accuracy. The associated parameters of the most appropriate fitted probability distribution function are calculated and the trends in the characteristic of the solar radiation are deduced. The result shows that logistic distribution presents the most suitable probability distribution function for modelling solar radiation over the selected environment with RMSE of 1.500 KW-hr/m²/day, MAE of 1.260 KW-hr/m²/day, MAPE of 22.000 % and R² of 0.880. When compared with the other five distribution functions, the same trend could be seen although with different values of RMSE, MAE, MAPE and R². The estimated distribution location and scale parameters of the model vary with month and season. The overall result will be useful for predicting future solar radiation over the studied environment. It will also be a good reference point for the design of large solar power projects in Yenagoa in particular and south southern Nigerian environments at large.

Keywords: probability distribution functions, solar radiation, logistic distribution, Yenagoa

1. Introduction

Knowing the available solar radiation data of an area is one of the essential requirements for assessing the potential use and designing of a solar energy system such as solar thermal system, passive solar design and photovoltaic farms [1]. Solar radiation data are fundamental

requirement for the design of solar energy conversion systems.

The planning and development of local solar energy system is dependent on the availability of solar radiation data which could be used to assess the probable long term solar systems performance and economic viability. Solar radiation data should be continuously and

accurately measured over long period of time for the purposes of feasibility studies. Unfortunately, in some parts of the world, solar radiation measurement system are not readily available, this could be as a result of technical, financial and institutional limitations. The solar radiation received at the surface of the earth varies daily, seasonally and annually depending on the climatic condition of the region. Hence, many years of measurement observation is required in order to ensure accurate prediction and distribution. However, many developing and underdeveloped countries do not have the required facilities for measurement of continuous and accurate solar radiation in many locations. It is then imperative to develop empirical methods to estimate the solar radiation data of a potential location.

Various empirical models have been developed to model solar radiation from measured temperature, relative humidity, atmospheric pressure and relative duration of sunshine. Also, some statistical distribution functions have been used to model solar radiation values, for example, Ayodele, [2] had earlier investigated the most probable probability distribution model for modelling solar radiation in Ibadan, South western Nigeria. He concluded that logistic distribution function performs better than the other models considered. In the same vain, Authuret *al.*, [3] had shown that different distribution functions could be used to model different monthly solar radiation, judging from the solar radiation measurement performed in Kumasi, Ghana.

Chang, [4] worked on frequency distribution of solar radiation of Taiwan and he concluded that lognormal distribution function is better in modelling the solar radiation of Taiwan.

Abdulazeezet *al.*, [5] in their investigation, used factorial design in modelling solar radiation data measured from eight chosen locations in Nigeria. Other empirical methods and distribution functions that had been used in solar radiation modelling include: beta distribution [6], Angstrom model [7, 8], Page model [9] and Garcia model [10]. Although, solar radiations have been estimated for some part of Nigeria using different empirical and distribution methods but non to the best knowledge of the authors has modelled solar radiation of South southern Nigeria using probability distribution functions. Having the knowledge of the probability distribution function and distribution parameters of solar radiation of a location, helps in generating solar radiation data that will exhibit the same characteristics as the actual values of the considered location in the future. This is essential because it is the starting point for analysing the design of any solar power project. Presented in this study is the most appropriate probability distribution function that models the solar radiation data over Yenagoa, South southern Nigeria.

The rest of the paper is structured as in the following manner: section 2 presents information about the site and data acquisition method. In section 3, the methodology adopted was discussed while section 4 presents the

results and discussion and conclusions are summarised in section 5.

2. Site and Data Acquisition

The selected site for the study is Yenagoa, the capital city of Bayelsa State, South Southern Nigeria. Yenagoa lies between latitude 4.921162 and longitude 6.274773 the elevation is 8 metres with a population of about 2.6 million people. It is located within the coastal region and has two major seasons: the dry season which runs normally from November to March and the wet season, which normally runs from April to October. Most of the wet season months are associated with heavy rainfall, as a coastal region, sometimes rainfalls are also experienced in dry season months [11].

The solar radiation values used for analysis in this study were measured over a period of eleven years (2007-2017) and were collected from National Aeronautics and Space Administration (NASA) insolation data for Yenagoa. The long observed period of data will allow a proper understanding of the variation of solar radiation from month to month and season to season.

3. METHODOLOGY ADOPTED

It is impractical to report all the probability distribution functions as they are too many to be accommodated in the paper. Nevertheless, several probability distribution functions have been tested, however, the best six that are appropriate in terms of curve fitting were repeated in this study. The statistical distribution functions adopted for this study

are: normal, lognormal, gamma, logistic, extreme value and Weibull distributions.

3.1. Normal Distribution Function

The probability density and cumulative distribution functions of normal distribution are respectively expressed as:

$$f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], -\infty < x < \infty \quad (1)$$

$$F(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x (\exp\left[-\frac{(t-\mu)^2}{2\sigma^2}\right]) dt \quad (2)$$

The parameters of distribution μ and σ of normal distribution are calculated using [12]:

$$\mu \approx \frac{1}{n} \sum_{i=1}^n (x_i), -\infty < \mu < \infty \quad (3)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}, \sigma > 0 \quad (4)$$

where x is the measured solar radiation data, μ and σ are the parameters of the distribution

3.2. Lognormal Distribution Function

The lognormal distribution is applicable only when the quantities of interest are all positive, since $\log(x)$ can only exist when x is positive. The probability density function (pdf) and cumulative distribution function of the lognormal distribution are given as [12]:

$$f(x|\mu, \omega) = \frac{1}{x\omega\sqrt{2\pi}} \exp\left\{-\frac{(\ln x - \mu)^2}{2\omega^2}\right\}; x \geq 0 \quad (5)$$

$$F(x|\mu, \omega) = \frac{1}{x\omega\sqrt{2\pi}} \int_0^x \frac{\exp\left\{-\frac{(\ln t - \mu)^2}{2\omega^2}\right\}}{t} dt \quad (6)$$

where x is the measured data, ω is the lognormal shape parameter, μ is the lognormal scale parameter. The parameters (mean and

variance) of the lognormal distribution can be evaluated using (7) and (8) [12]:

$$\mu = \exp\left(\mu + \frac{\omega^2}{2}\right), \quad -\infty < \mu < \infty (7)$$

$$\sigma^2 = \exp(2\mu + \omega^2) [\exp(\omega^2) - 1], \sigma > 0 (8)$$

3.3. Gamma Distribution Function

The gamma distribution probability density function (pdf) and the cumulative distribution function are expressed as [12]:

$$f(x|a, b) = \frac{x^{a-1}}{b^a \Gamma(a)} \exp\left[-\left(\frac{x}{b}\right)\right], \quad a, b > 0 (9)$$

$$F(x|a, b) = \frac{1}{b^a \Gamma(a)} \int_0^x t^{a-1} \exp\left[-\left(\frac{t}{b}\right)\right] dt (10)$$

where x is the measured solar radiation values, a is the scale parameter and b is the shape parameter of the gamma distribution. The equations for mean and standard deviation of gamma distribution function are shown in (11) and (12) [12]:

$$\mu = ab (11)$$

$$\sigma = \sqrt{ab^2} (12)$$

3.4. Logistic Distribution Function

The probability density function of a logistic distribution with the location parameter μ and scale parameter s is expressed as [12]:

$$f(x|\mu, s) = \frac{\exp\left(\frac{x-\mu}{s}\right)}{s(1+\exp\left\{\frac{x-\mu}{s}\right\})^2}, \quad -\infty < x < \infty (13)$$

The cumulative distribution function of the distribution is given by:

$$F(x|\mu, s) = \left[1 + \left\{-\left(\exp\left(\frac{x-\mu}{s}\right)\right)\right\}\right]^{-1} (14)$$

$$\mu = E(x), \quad -\infty < \mu < \infty (15)$$

$$s = \sqrt{\frac{3 \text{var}(x)}{\pi^2}}, \quad s > 0 (16)$$

3.5 Extreme Value Distribution Function

The probability density function and cumulative distribution function for extreme value distribution function is expressed as [13]:

$$f(x|\alpha, \beta) = \frac{1}{\beta} \exp\left(-\frac{x-\alpha}{\beta}\right) \exp\left(-\exp\left(-\frac{x-\alpha}{\beta}\right)\right), \quad x \geq \alpha, \beta > 0 (17)$$

$$F(x|\alpha, \beta) = \exp\left[-\exp\left(-\frac{x-\alpha}{\beta}\right)\right] (18)$$

where α is the location parameter and β the scale parameter and can be determined by (17) and (18) [12]:

$$\mu = \alpha + 0.5772\beta (19)$$

$$\sigma = \frac{\beta\pi}{\sqrt{6}} (20)$$

3.6. Weibull Distribution Function

For positive values of the scale parameter a , and shape parameter c , the density and cumulative functions are given as [13]:

$$f(x|a, c, m) = \frac{c}{a} \left(\frac{x}{a}\right)^{c-1} \exp\left[-\left(\frac{x}{a}\right)^c\right], \quad x > m (21)$$

$$F(x|a, c, m) = 1 - \exp\left[-\left(\frac{x}{a}\right)^c\right] (22)$$

The Weibull mean and variance equations are expressed as [14]:

$$\mu = a\Gamma\left(1 + \frac{1}{c}\right), \quad a > 0, c > 0 (23)$$

$$\sigma^2 = a^2\Gamma\left(1 + \frac{2}{c}\right) - \left[\Gamma\left(1 + \frac{1}{c}\right)\right]^2 (24)$$

3.7. Assessing the performance of the distribution functions

To check how accurately a theoretical distribution function fits with the measured solar radiation data, four different statistical goodness of fits were selected as benchmark. The metric measures are: the Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and Coefficient of Determination (R^2). The closer the values of RMSE, MAE and MAPE of a distribution to zero the better the curve fitting. In the same vain, R^2 is a statistical tool used to determine to what extent a prediction can be deduced from a model. The condition for evaluation is $0 \leq R^2 \leq 1$, with 1 as the perfect fit. The closer the value of R^2 is to 1, the better the fit to the measured variables.

The expressions for the evaluation of the statistical tools are given as [1]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (x_m(j) - x_p(j))^2} \quad (25)$$

$$MAE = \frac{1}{n} \sum_{j=1}^n [|x_m(j) - x_p(j)|] \quad (26)$$

$$MAPE = \sum_{j=1}^n \left[\left| \frac{(x_m(j) - x_p(j))}{x_m(j)} \right| \right] \times 100 \quad (27)$$

$$R^2 = \left[\frac{n(\sum_{j=1}^n (x_m(j)x_p(j)) - (\sum_{j=1}^n x_m(j))(\sum_{j=1}^n x_p(j)))}{\sqrt{[n\sum_{j=1}^n (x_m(j))^2 - (\sum_{j=1}^n x_m(j))^2][n\sum_{j=1}^n (x_p(j))^2 - (\sum_{j=1}^n x_p(j))^2]}} \right]^2 \quad (28)$$

where $X_m(j)$ is the j^{th} measured solar radiation, $x_p(j)$ is the j^{th} predicted solar radiation and n is the number of observed data.

4. Results and Discussion

Fig. 1 presents the comparison of the mean monthly solar radiation of Yenagoa for each of the considered years. The figure shows that the month of January exhibited the least solar radiation in most of the considered years, followed by the month of December and then February. The average values of solar radiation for these months are 1.05, 11.75 and 1.90 KW-hr/m²/day respectively with standard deviation of 0.81, 1.53 and 1.10 KW-hr/m²/day.

However, the month of June recorded the highest average solar radiation value of around 6.82 KW-hr/m²/day with standard deviation of 0.86 KW-hr/m²/day, this is followed by the month of May with average value of 6.72 KW-hr/m²/day and standard deviation of 1.27 KW-hr/m²/day, then the month of October with average value of 6.63 KW-hr/m²/day and standard deviation of 0.94 KW-hr/m²/day. Similar trend could be observed in the other months, although with different average values of solar radiation and standard deviations. The summary of the average mean values of solar radiation for the different years under study as well as the standard deviation are presented in Table 1.

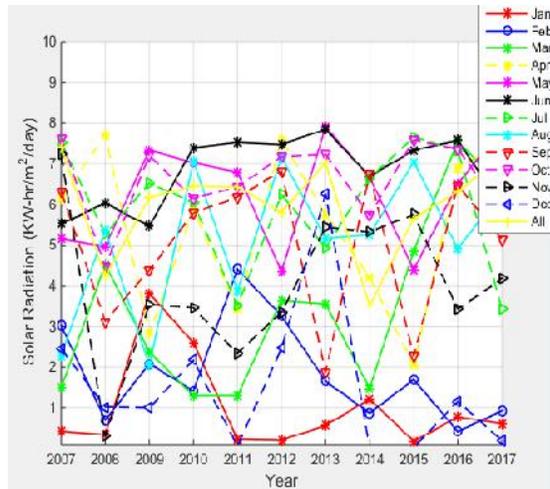


Fig. 1. Comparison of the monthly average global solar radiation for each of the year. (2007-2017)

Table 1. Average Value of Global Solar Radiation of Yenagoa Over Eleven Years Period

Parameter (KW-hr/m ² /day)	Jan	Feb	Mar	Apr	May	June	Jul	Aug	Sep	Oct	Nov	Dec
Mean	1.05	1.90	3.51	5.50	6.72	6.82	6.00	5.14	5.01	6.63	4.04	1.75
Std	0.81	1.10	1.41	1.70	1.27	0.86	1.41	1.71	1.74	0.94	1.78	1.53

To determine the most appropriate probability distribution function that can accurately model the solar radiation of Yenagoa, several statistical distribution functions were tested. However, six out of the tested distributions that best curve fit the actual data are presented in this paper and the results are depicted in Fig. 2 with logistic distribution exhibiting a closer match with the measured data followed by normal distribution function. The metric evaluation of the distribution functions and order of good fit are also depicted in Table 2. It can be observed that logistic distribution function presents the best goodness of fit for modelling the solar

radiation of Yenagoa with RMSE of 1.500 KW-hr/m²/day, MAE of 1.260 KW-hr/m²/day, MAPE of 22.000 % and R² of 0.880. When compared with the other distribution functions, logistic distribution function is followed competitively by normal distribution with RMSE of 1.533 KW-hr/m²/day, MAE of 1.280 KW-hr/m²/day, MAPE of 27.000 % and R² of 0.845. This validates the accuracy of the statistical test.

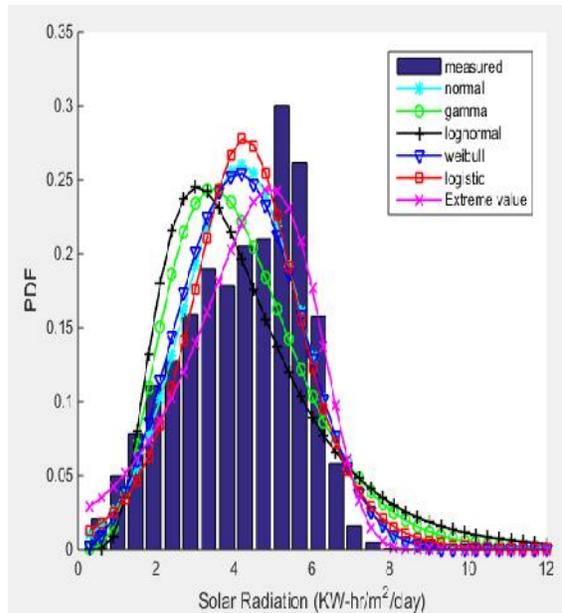


Fig. 2. Fitting of Probability distribution functions to the solar radiation data for Yenagoa.

Table 2. Performance evaluation of statistical distribution for modelling solar radiation for Yenagoa

Statistical Distribution	RMSE (KW- hr/m ² /day)	MAE (KW-hr/m ² /day)	MAPE (%)	R ²	Order of fitness
Normal	1.533	1.280	27.000	0.845	2 nd
Extreme value	1.734	1.355	51.070	0.713	5 th
Lognormal	1.544	1.300	32.000	0.740	3 rd
Gamma	2.090	1.630	59.800	0.688	6 th
Logistic	1.500	1.260	22.000	0.880	1 st
Weibull	1.720	1.350	41.000	0.730	4 th

Fig. 3 depicts the monthly variation in logistic probability distribution of solar radiation over the period of eleven years considered. It can be inferred from the

figure that there is wide variation of solar characteristic in Yenagoa, as such, many years of solar radiation observation are required to make a reasonable conclusion

on implementation of any solar power project.

Table 3 presents the estimated location parameter (μ) and the scale parameter (s) of the best distribution function (logistic distribution) for the months of the years under consideration (2007-2017). From the result of the table, it can be seen that both the location parameter and scale parameters vary over a wide range of values showing wide variations in solar radiation of Yenagoa.

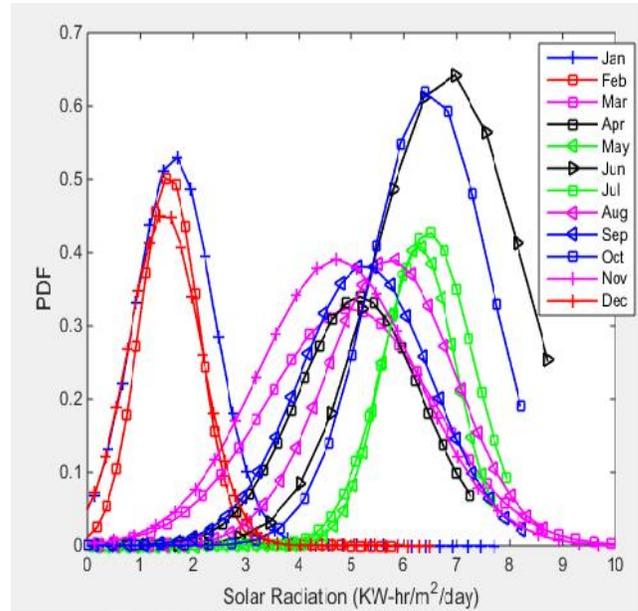


Fig.3. Monthly variation in solar radiation for Yenagoa.

Table 3. Monthly Logistic location parameter (μ), and scale parameters (s) for the considered years.

Parameter KW-hr/m ² /day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
μ	1.64	1.73	3.34	5.61	6.35	6.87	6.05	5.28	5.22	6.73	4.06	1.26
S	0.40	0.43	0.68	0.94	0.43	0.38	0.46	0.56	0.18	0.36	0.56	0.50

The location and scales parameters of logistic distribution function are calculated for two different seasons (wet and dry seasons) observed in Nigeria for the period under study, the results are shown in Table 4. The results from the table show that the dry season has higher value of location parameter when compared to the wet

season. This is because some months in wet season, the weather is always cloudy as a result of constant rainfall. The variation in the seasonal distribution are presented in Figs. 4 and 5 respectively.

Table 4. Seasonal logistic location and scale parameters for Yenagoa

Parameter KW- hr/m ² /day)	Wet	Dry
μ	3.81	4.37
S	0.87	0.83

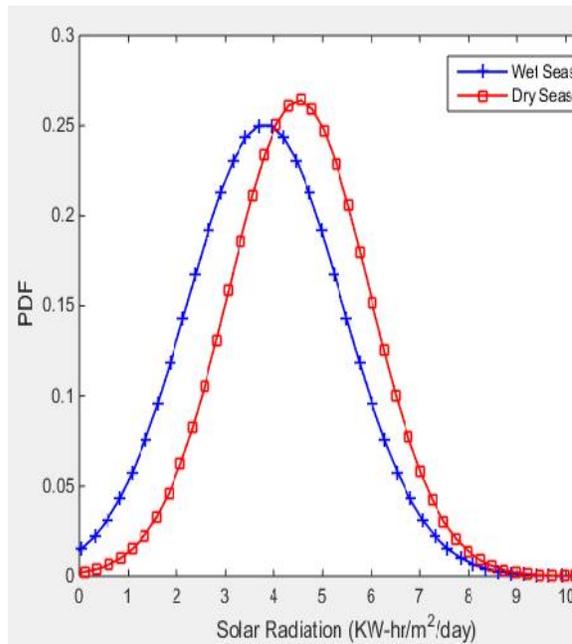


Fig. 4. Seasonal variation in logistic probability distribution function for Yenagoa.

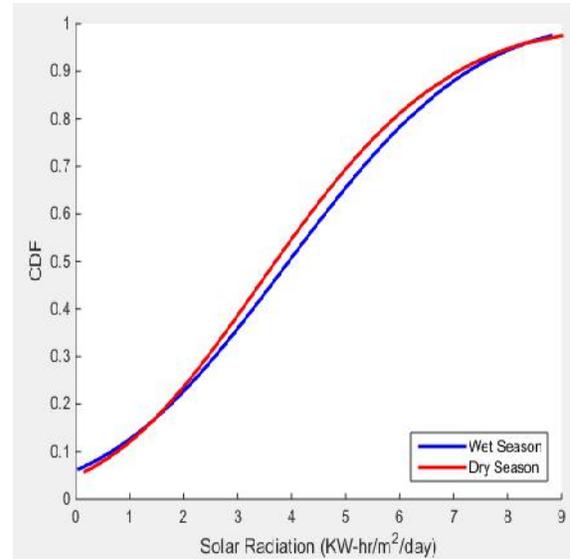


Fig. 5. Seasonal variation in logistic probability cumulative distribution function for Yenagoa

5. Conclusions

In this study, investigation of the most appropriate statistical distribution function that models the solar radiation data of Yenagoa has been carried out and parameters of the distribution function determined. From the study, it was thus found that logistic probability distribution function is the most probable distribution function for modelling the solar radiation of the studied environment out of the six probability distributions considered based on some metric measures. The logistic distribution exhibited RMSE of 1.500 KW-hr/m²/day, MAE of 1.260 KW-hr/m²/day, MAPE of 22.000 % and R² of 0.880.

The monthly and seasonal variations in the location and scale parameters of logistic distribution have been estimated with the month of June having the highest location parameter 6.87 KW-hr/m²/day. Also, the dry season of the years under consideration has a higher location parameter than the wet season, this is as a result of cloudiness of the sky during some months in wet season as a result of rainfall.

The results of this study is useful as a first-hand information to the project engineers, individuals and researchers who are interested in solar power project in Yenagoa, Bayelsa State. Future research emanating from this study will also assist in the predicting of future solar radiation of the study locations using logistic distribution function.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

References

1. Bulut, H and Büyükalaca, O. Simple model for the generation of daily global solar radiation data in Turkey. *Applied Energy*, 84, 5 (2007), pp. 477-491.
2. Temitope, R. A. Determination of Probability Distribution function for modeling global solar radiation: Case study of Ibadan, Nigeria. *International Journal of Applied Science and Engineering*, 13 (3), (2015), 233-245.
3. Arthur, Y.D, Gyamfi, K.B and Appiah, S.K. Probability distribution of hourly solar irradiation in Kumasi-Ghana *International Journal of Business and Social Research*, 3, 3 (2013) pp. 63-75
4. Chang, T.P. Investigation on frequency distribution of global radiation using different probability density functions. *International Journal of Applied Science and Engineering*, 8, 2 (2010) pp.99-107.
5. Abdulzeez, M.A., Ahmed, A and Burari, F.W. The use of factorial design in the analysis of global solar radiation in Nigeria. *Archives of Applied Science Research*, 2 (2010), 5 pp.36-44.
6. Ettoumi, F., Mefti, A., Adane, A and Bouroubi, M.Y. Statistical analysis of solar measurements in Algeria using beta distributions. *Renewable Energy*, 26, 1 (2002) 47-67.
7. Musa, B, Zangina, U and Aminu, M. Estimation of global solar radiation in Maiduguri, Nigeria using angstrom model. *ARP Journal of Engineering and Applied Sciences*, 7, 12 (2012) 8. 1623-1627.
8. Medugu, D.W and Yakubu, D. Estimation of mean monthly global solar radiation in Yola, Nigeria using angstrom model. *Advances in Applied*

- ScienceResearch*, 2, 2 (2011) 414-421.
9. Olayinka, S. Estimation of global and diffuse solar radiations for selected cities in Nigeria. *International Journal of Energy and Environmental Engineering*, 2, 3 (2011) 13-33.
10. Isikwue, B.C. Amah, A.N and Agada, P.O. Empirical model for the estimation of global solar radiation in Makurdi, Nigeria. *Global Journal of ScienceFrontier ResearchPhysics & SpaceScience*, 12, 1 (2012) 59-61.
11. Ojo, J.S. Falodun, S.E. and Odiba, O. 0 °C isotherm height distribution for Earth-space communication Satellite links in Nigeria. *Indian Journal of Radioand Space Physics*, Vol. 43, (2014). Pp. 225-234.
12. Krishnamoorthy, K. Handbook of Statistical distributions with applications. *Taylor & Francis Group, 6000 Broken Sound Chapman & Hall/CRC, Parkway NW, Suite 300 Boca Raton. FL,334487 2742*, (2006).
13. Johnson Kotz N.L.S and Balakrishnan, N. Continuous univariateinterscience distributions. *Vol. 1, Hoboken, NJ: Wiley* (1993).
14. Salo, J. Vuokko, L and Vainikainen, P.
- Why is shadow fading lognormal?
Proceedings of the International Symposium on Wireless Personal

MultimediaCommunications, Aalborg, Denmark (2005), pp. 522–526.