Empirical Validation of Regression Models for Forecasting Global Solar Irradiance: A Case Study of Abeokuta, Nigeria

*1Martins O. Osifeko, 1Olufunmilayo I. Sanusi, 1Olatilewa R. Abolade, 1Sonia Iwieriebor, 2Gbenga D. Obikoya
1Department of Computer and Electrical Engineering, Olabisi Onabanjo University, Ago-Iwoye, Nigeria
2Department of Electrical and Electronics Engineering, Federal University Oye – Ekiti, Nigeria
(osifeko.martins|sanusi.olufunmi|abolade.raphael}@oouagoiwoye.edu.ng)
iwierieborsonia@gmail.com|gbenga.obikoya@fuoye.edu.ng

Abstract—The performance of solar radiation distribution models is highly dependent on the location of use. This makes the need for location dependent validation necessary. An empirical validation of three solar radiation distribution models was carried out in this work using empirical data obtained from the Nigeria Meteorological Agency (NIMET). The models include the Angstrom-Prescott, Hargreaves and Garcia model. Four standard statistical tools of correlation ratio (R), mean bias error (MBE), mean percentage error (MPE), root mean square error (RMSE) were used for the validation. Results from the work reveal that the Angstrom-Prescott model had the least MBE with an average value of 0.1 making it the most reliable among the three models. In terms of MPE values, the Angstrom-Prescott model also performed best with an average MPE value of 0.1 while the Hargreaves and Garcia model had an average MPE performance of 0.2 and 0.15 respectively. In terms of the RMSE and correlation ratio, the Angstrom-Prescott model also performed better than the other two models with an average RMSE and correlation ratio of 0.3 and 0.65. This work thus reveals that, for a medium term solar radiation forecasting goal, the best regression based model that accurately predict the solar radiation distribution in Abeokuta is the Angstrom-Prescott model.

Keywords—Angstrom-Prescott model, Garcia model, Solar Hargreaves model, Radiation Distribution, Solar irradiance

1. INTRODUCTION

The dwindling fossil fuel reserves in many countries and the need for a cleaner source of energy is motivating researchers to continually seek for better ways of generating energy (Park, Das and Park, 2015). One of such ways is using energy from the sun. However, the variability and intermittent nature of solar radiation are major factors that have hindered advancement in the use of solar energy as a reliable alternative source of energy. It is believed that the knowledge of solar radiation distribution at a geographical location would enhance the development of solar energy methods. Thus, forecasting has been proposed as a solution to the variability nature of solar radiation (Sansa et al, 2014). This forecast can be carried out on a short-term, medium-term or on a long-term basis. Solar forecasting is therefore a technology that aids the integration of solar energy into the grid. The twin benefit of better energy quality and operational costs derived from solar forecasting has motivated researchers to continually develop forecast techniques that accurately predict solar radiation pattern in a location (Inman et al, 2013). Several models have been proposed and attempts made to validate proposed models.

2. RELATED WORK

The estimation of daily global solar radiation has been reviewed in most of the researches based on the duration of sunshine, identifying the best model and determining different coefficients for several locations. Some parameters that include the average daily global solar radiation, maximum and minimum temperature, cloudiness index and relative humidity were used to develop some correlation equations (Olayinka, 2011). The combination of sunshine hours, maximum and minimum temperature, cloudiness index and relative humidity revealed results which indicate that the proposed model can predict solar radiation. It was also demonstrated by Etuk et al., (2016) that the availability of daily photosynthetic active radiation (PAR) depends on a horizontal surface for specified locations using clearance index. Since there are limited weather stations to measure this radiation component, it is necessary to develop a model for predicting PAR using meteorological parameters. Simulations were conducted to correct these problems. It was observed that for a wide variety of climatic conditions, the ratio of PAR to global solar radiation is constant between 0.4 and 0.7. A quadratic model was derived from this and used to forecast the monthly average daily PAR. Popular linear regression methods were evaluated and the models used for the estimation of monthly and annual average global solar radiation from specified locations in France are the Linear model, Quadratic model and Cubic model (Jemaa et al., 2013). The result revealed that the values obtained from the three models and the measured data were similar for all the days of the years with the linear model being better in estimating solar parameters where the errors between the measured and calculated values are negligible.

The solar radiation data in Ankara, Turkey (lat. 39°95′N, long. 32°88′E, alt. 891 m) based on 6 years of global solar radiation data measured on a horizontal surface was analyzed in Yorukogu and Celik (2006). The yearly and monthly optimum tilt angles were computed and considered the optimal results for Turkey and places with similar climatic conditions. An analysis of 10 arithmetic models made by Ulgen, (2004) was used to calculate diffuse solar irradiance on inclined surfaces in Valladolid, Spain. The data readings were taken on hourly and daily basis from the 1st of August 1998 until the 15th of March 2000. Three statistical methods were used to confirm the results, taking into consideration the area’s feature which is important for diffuse irradiance on inclined surfaces. They concluded the Muneer model and the Reindl model gave the best results for hourly and for daily values respectively.
From the literature reviewed, it can be observed that most of the work only used one statistical tool for the evaluation of the forecast methods. However, this current research work aims to carry out an empirical validation of three regression based solar radiation forecasting methods using four statistical tools.

3. METHODOLOGY

Three models would be validated with empirical data obtained from the Nigeria Meteorological Agency (NIMET) with an aim of determining the model that accurately predicts solar radiation distribution in Abeokuta, Nigeria. Based on their popularity and use in literature, three popular regression models for forecasting solar radiation were identified and selected for use in this research work. The models are provided in the following sub-sections.

3.1 Angstrom-Prescott Model

This model, developed in 1940 by Prescott, has been widely used and described by various authors in the study of global solar radiation. Its variations as given by Gairaa and Bakelli, (2013), are as shown below:

i. Angstrom-Prescott Linear Model:

\[ \frac{H}{H_0} = a + b \left( \frac{ss}{sso} \right) \] (1)

ii. Angstrom-Prescott Quadratic Model:

\[ \frac{H}{H_0} = a + b \left( \frac{ss}{sso} \right) + c \left( \frac{ss}{sso} \right)^2 \] (2)

where, 

- \( H \): Monthly average daily global solar radiation on horizontal surface (MJ/m²/day)
- \( H_0 \): Monthly average daily extraterrestrial radiation (MJ/m²/day)
- \( ss \): Monthly Daily Sunshine Duration
- \( sso \): Maximum possible monthly average daily length
- \( w \): the mean sunrise hour angle for the given month
- \( a, b = \) empirical constants

3.2 Hargreaves and Samani Model

This model was first used by Hargreaves et al. 1985 to estimate the global solar radiation by using the difference between the daily maximum and minimum air temperature and extra-terrestrial radiation (Gairaa and Bakelli, 2013). This model is represented below:

\[ \frac{H}{H_0} = a + b(T_{max} - T_{min})^2 \] (3)

where,

- \( H \): Monthly average daily global solar radiation on horizontal surface (MJ/m²/day)
- \( H_0 \): Monthly average daily extraterrestrial radiation (MJ/m²/day)
- \( T_{max} \): Maximum temperature
- \( T_{min} \): Minimum temperature
- \( a, b = \) empirical constants

3.3. Garcia’s Model

The Garcia model is a slight modification of the Angstrom-Prescott model. Equations for its variations are given below (Gairaa and Bakelli, 2013):

i. Garcia Linear Model:

\[ \frac{H}{H_0} = a + b \left( \frac{\Delta T}{N} \right) \] (4)

ii. Garcia Quadratic Model:

\[ \frac{H}{H_0} = a + b \left( \frac{\Delta T}{N} \right) + c \left( \frac{\Delta T}{N} \right)^2 \] (5)

where,

- \( N = \) Maximum possible monthly average daily length
- \( \Delta T = \) Difference between maximum and minimum temperature values
- \( H, H_0, a \ and \ b = \) as defined in the Angstrom model

3.4. Statistical Tools

This section gives a brief description of the tools used in this work. The four statistical tools that we used in evaluating the performance of the models of solar radiation estimations include Mean bias error (MBE), root mean square error (RMSE), mean absolute percentage error (MAPE), and the R SQUARE. We describe them briefly in the following subsections:

3.4.1. Correlation Ratio

The correlation coefficient reflects the quality of the model. The closer the correlation coefficient to 1, the better is the quality displayed by the model (Beyer et al, 2009):

\[ R = \frac{(\Sigma_{k=1}^n (y_k - \bar{y})(x_k - \bar{x}))^2}{(\Sigma_{k=1}^n (y_k - \bar{y})^2)(\Sigma_{k=1}^n (x_k - \bar{x})^2)} \] (6)

where, \( y_k \) = kth calculated value, \( x_k \) = kth measured value and \( n \) = total number of observations.

3.4.2. Root Mean Square Error

The root-mean-square error (RMSE) is a frequently used measure of the differences between values predicted by a model or an estimator and the values observed from the modeled item. RMSE is a good measure of precision and provides information on the short-term performance of a model (Beyer et al, 2009). The value of RMSE is always positive, equating to zero in the ideal case. A low RMSE is however desirable. The RMSE may be computed by using the equation below:

\[ \text{RMSE} = \sqrt{\frac{1}{n} \Sigma_{k=1}^n (x_k - \bar{x})^2} \] (7)

where, \( y_k \) and \( x_k \) are as defined in the correlation ratio.

3.4.3. Mean Absolute Percentage Error

The Mean Absolute Percentage Error (MAPE) value provides the percentage deviation between the calculated and measured data. The ideal value of MAPE is zero. The Mean Absolute Percentage Error is given by (Beyer et al, 2009):

\[ \text{MAPE}\text{=}\sum_{k=1}^{n} \left( \frac{|y_k-x_k|}{x_k} \right) \] (8)

where \( y_k \) and \( x_k \) are as defined in the correlation ratio.

3.4.4. Mean Bias Error

The mean bias error (MBE) provides information on the long-term performance of the correlations by allowing a comparison of the actual deviation between calculated and measured values term by term. The ideal value of the MBE is zero. The MBE is given by (Beyer et al, 2009):

\[ \text{MBE}\text{=}\sum_{k=1}^{n} (y_k-x_k) \] (9)

where \( y_k \) and \( x_k \) are as defined in the correlation ratio.
3.5 Data Analysis
As stated in section 3, the monthly mean, sunshine hours and daily data of Global Solar Radiation of Abeokuta with longitude 07.01N, latitude 03.20E and Altitude 104° for year 2016 were collected from the Nigeria Meteorological Agency. The following were calculated from the data set;

i. the monthly mean sunshine duration, calculated in hours as

\[ s = \frac{2w}{12} \]

ii. Solar radiation incident outside the earth's atmosphere called extraterrestrial radiation, calculated using the following equation (Gairaa and Bakelli, 2013):

\[ H_0 = \frac{24}{\pi} G_{SC} \left( 10.033 \frac{360 \cdot \Delta t}{365} \cdot \left( \cos \phi \cos \delta \sin \omega + \frac{2w}{360} \omega \sin \phi \sin \delta \right) \right) \]  

\( \text{(10)} \)

where,

\( \omega = \text{set hour angle} \)
\( \omega = \cos^{-1}(-\tan \phi \tan \delta) \)
\( G_{SC} = \text{Solar constant taken to be equal to 1367 (W/m}^2) \),
\( \phi = \text{Latitude of the location under consideration,} \)
\( \Delta t = \text{Number of the year starting from the first of January (i.e. Julian day number), and} \)
\( \delta = \text{Declination angle given below as (Gairaa and Bakelli, 2013):} \)

\[ \delta = 23.45 \sin \left( \frac{360 (284+\Delta t)}{365} \right) \]  

\( \text{(11)} \)

A MATLAB program is used to determine the empirical constants a, b and c for each of the described model.

4. RESULTS AND DISCUSSIONS
The average monthly solar parameters used for the solar radiation regression models of Abeokuta with longitude 07.01N, latitude 03.20E and Altitude 104°, where constant global solar radiation is 1366Wm-2 is shown in the table 1. Figure 1 shows the solar radiation values calculated from the three models and empirical values obtained from NIMET. The results of the standard statistical tests carried out on each of the three models using MATLAB are shown in Figure 2 to Figure 5. The standard statistical tests are used as the yardstick in determining the best model for a given period.

| Month | Wind Speed (Km/hr) | Solar Radiation (MJ/m2) | MIN TEMP (° C) | MAX TEMP (° C) | SUNSHINE HOURS (hrs) |
|-------|-------------------|-------------------------|---------------|---------------|---------------------|
| Jan   | 3.9               | 22.3                    | 22            | 32            | 180.43              |
| Feb   | 3.7               | 21.4                    | 22.5          | 32            | 211.21              |
| Mar   | 3.7               | 13.3                    | 23            | 34.5          | 212.75              |
| Apr   | 1.9               | 10.9                    | 21.5          | 35            | 175.70              |
| May   | 4.6               | 22.9                    | 21            | 32            | 170.8               |
| Jun   | 2.5               | 11.2                    | 19.5          | 27.5          | 157.5               |
| Jul   | 4.3               | 17.5                    | 20.2          | 30.2          | 191.9               |
| Aug   | 4.5               | 7.01                    | 22.2          | 30            | 179.93              |
| Sep   | 4.1               | 19.1                    | 21            | 30.2          | 186.7               |
| Oct   | 2.6               | 17.8                    | 20.5          | 30.5          | 182.9               |
| Nov   | 3.6               | 18.8                    | 22.5          | 31.5          | 213.9               |
| Dec   | 5.5               | 17.7                    | 23            | 31            | 211.9               |
In terms of the root mean square error values shown in Figure 4, the Angstrom-Prescott model also fared better than the other two models with an average RMSE value of 0.3. The Hargreaves model had an average value of 0.4 while the Garcia model had an average value of 0.35. From the correlation ratio values in Figure 5, it is observed that the Angstrom-Prescott model had the best performance with an average correlation of 0.65 when compared with the Hargreaves and Garcia model with an average value of 0.55 and 0.5 respectively.

5. CONCLUSION
An empirical validation of three regression based solar radiation models, using data obtained from NIMET, was carried out in this research. From the results obtained in this research, it can be concluded that for a medium term solar radiation forecasting goal, the best regression based model that accurately predicts the solar radiation distribution in Abeokuta is the Angstrom-Prescott model based on the consistent performance of the model with respect the four statistical test methods used in the evaluation. It is recommended that a long term empirical validation be carried out to determine the performance of the models on a long-term basis.

REFERENCES
Beyer, H. G et al (2009), “D 1.1. 3 report on benchmarking of radiation products. Report under contract,” (038665).
Etuk, E. S., Nwokolo, S. C., Okechukwu, A. E., & John-Jaja, S. A. (2016), “Analysis of photosynthetically active radiation over six tropical ecological zones in Nigeria,” Journal of Geography, Environment and Earth Science International, 7(4),1-15.
Gairaa, K., & Bakelli, Y. (2013), “A Comparative Study of Some Regression Models to Estimate the Global Solar Radiation on a Horizontal Surface from Sunshine Duration and Meteorological Parameters for Ghardaïa Site, Algeria,” ISRN Renewable Energy.
Inman, R. H., Pedro, H. T., & Coimbra, C. F. (2013). Solar forecasting methods for renewable energy integration. Progress in energy and combustion science, 39(6), 535-576.
Jemaa, A. B et al, (2013), “Estimation of global solar radiation using three simple methods,” Energy Procedia, 42, 406-415.
Olayinka, Sunday. (2011). “Estimation of global and diffuse solar radiations for selected cities in Nigeria,” International Journal of Energy and Environmental Engineering, 2(3), 13-33.
Park, J. K., Das, A., & Park, J. H. (2015), “A new approach to estimate the spatial distribution of solar radiation using topographic factor and sunshine duration in South Korea,” Energy Conversion and Management, 101, 30-39.
Sansa, I et al, (2014), “PV power forecasting using different artificial neural networks strategies,” In Green Energy, 2014 International Conference on (pp. 54-59), IEEE.
Ulgen, K.H., (2004), “Solar radiation models. Part 2: Comparison and Developing New Models,” Energy Sources 26: 521-530.
Yorukogu, M., and Celik, A. N. (2006), “A critical review on the estimation of daily global solar radiation from sunshine duration,” Energy Conversion and Management, 47(15), 2441-2450.