INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT

ISSN 2278 – 0211 (Online)

Modeling and Simulation of Global Solar Radiation in Warri, Nigeria Using Gamma Test and Artificial Neural Network Algorithms

Asagha, Emmanuel Nkoro
Department of Physics, Federal College of Education, Obudu, Nigeria
Sunday O. Udo
Department of Physics, University of Calabar, Calabar, Nigeria
Idugba M. Echi
Department of Physics, University of Agriculture, Makurdi, Nigeria

Abstract:
A few papers on studies of global solar radiation estimation for some locations in Nigeria using artificial neural network have been reported. This paper is the first report of gamma test (GT)-driven artificial neural network (ANN) approach adopted for nonlinear data model development and simulation of global solar radiation (G_{sr}) within a coastal region and tropical location, Warri-Nigeria ($5^0 31^1 N$, $5^0 45^1 E$). Gamma test-run with seven(7) years(2003-2009) data consisting of 2201 unique data points contributed by maximum temperature (T_{mn}), minimum temperature (T_{mn}), rainfall (R_f), Relative humidity at 9hours(Rh_{09}) and extraterrestrial radiation (G_0) is adopted for the ANN modelling and simulation of G_{sr}. Using near neighbour number($p_{\max}$) of 10, the results of the GT scatter plots for all-input embedding dimension gives a gamma statistic(Γ) of 0.0005 and an average correlation coefficient of determination (R^2) of 0.9887. G_{sr} is simulated using data models developed from three different ANN algorithms (Two Layer back-propagation, Conjugate gradient and Bryoden Fletcher Golfab-Shanno) with feed-forward two layer network topology (5-7-7-1) which learnt at a rate of 0.25 The results obtained for various embedding dimensions and data length choices using both training and validation data sets shows the root mean square error (RMSE) and R^2 to range between 0.2373 to 13.1570 and 0.3545 to 0.9994 for the different algorithms. The significant contribution of G_0 for all masks is revealed by the results of mask 11110 where critically low values are obtained when G_0 is excluded and its also noted not to be adequate in predicting G_{sr} for this location and other similar locations as exhibited by results for mask 00001

Keywords: *Modelling, Global Solar Radiation, Gamma Test, Mask, Artificial Neural Networks, Algorithm*

1. Introduction

The use of artificial intelligence techniques like artificial neural network (ANN) in modelling global solar radiation (G_{sr}) has been widely reported in literature (Krishnaiah, *et al*, 2007: Barbero, *et al.*, 2006) and for a few locations in Nigeria (AbdulAzeez, 2011: Ewona, 2011 : Ibeh *et al.*, 2012) before the emergence of the artificial neural network approach, the prediction of global solar radiation had been evolving, beginning from the classical empirical sunshine duration models of Angstrom (1924) and Prescott (1940) to satellite imagery procedure (Dagestad, 2005) and even more complex hybrid models (Yang *et al.*, 2001) for estimating G_{sr}. Efforts to address challenges presented by effects like latitude and elevation have been made (Gopinathan, 1988). Even for the seamless empirical models of the Prescott type, accurate determination of the coefficient in the parametric correlation equations has engaged the thoughts of atmospheric physics scholars for quite a long time. Yeboah-Amankwah and Agyeman (1990) develop a differential form of the Angstrom model while considering the coefficients as time –dependent indices. The work of Ninomiya (1994) on snow and rainfall relationship with the coefficient is relevant developing data model for a location like Warri in Nigeria with high records of rainfall. Model like those of Leckner (1978) considered the physical process in details such that the effect of latitude elevation and other factors that contribute in most G_{sr} modelling have been taken into account (Yang *et al.*, 2001). The use of complex geographical and metrological parameters in the modelling of G_{sr} by means of radiative transfer equations or models has been found to provide considerable precision in the resulting outputs (AbdulAzzez, 2011). However, there exist limitations due to calculation difficulty and the unavailability of specific input parameters required for computation which are often scarce (Dogniaux, 1973). For a reliable prediction and analysis of G_{sr} data for a given location like the one for Fort-Collins Colorado (Conley, 1993), long term and reliable data prediction method have to be developed. It is a challenge however, that for a coastal town in a tropical location like Warri in Nigeria, there is dearth of reliable and consistent records of global solar radiation. Interestingly, aside the few empirical efforts to predict global solar radiation in Warri as reported in Ewona (2011) and Ibeh *et al.*, (2012). The artificial neural network approach to modelling G_{sr} which appear to accommodate the complexity of the relationship between solar radiation and meteorological factors affecting its surface measurements is gradually gaining ground even for other locations in Nigeria (AbdulAzeez, 2011, Ewona, 2011 and Ibeh *et al*, 2012). The growing use of ANN for data analysis is gradually being emphasized with replication of the application of this technique for estimating G_{sr} at different locations across the world. In Nigeria, different ANN architecture and algorithms have been used to estimate global solar radiation in Warri-Nigeria (Ibeh *et al*, 2012) and Ewona, 2011). Despite the advantage of being able to include superfluous input variables to the data model, there is the challenge of handling issues like the increasing complexity of the defining function, highly noisy data and optimal data length for model training and validation. The arbitrary choices of the training and validation subset of data are issues yet to be addressed in studies for this and similar locations in Nigeria. Elsewhere, studies have been made to address these challenges using Bayesian method of automatic relevance determination (ARD) for multilayer feed forward perceptron network (Lopez, *et al.*, 2005) and Gamma test analysis for the Brue catchments in the United Kingdom, a temperate location (Remesan, *et al*, 2008).

In this work, we adopt for the first time for any location in Nigeria, the Gamma test (GT) analysis approach in modelling global solar radiation. Three different artificial neural network learning algorithms are used in the model construction and development process .GT analyses of a set of data(*warri-solar.csv)* for the location, with 2201 data points is performed and the development of ANN data models for predicting global solar radiation, in Warri are constructed, developed and analyzed.

2. Gamma Test Analysis

Using any continuous model fitting method like the LLR or artificial neural network for an unknown function, the use of Gamma test (GT) is imperative given the noise in data as they affect smooth model building and development from inputs to output. GT estimates the best mean square error (MSE) that can be achieved due to the presence of statistical noise on the output. The analysis is implemented using WinGamma with the limiting accuracy being directly linked to measurement noise or insufficient data. While predicting the geo-temporal variation of crime and disorder using GT, Jonathan *et al.*, (2003) noted that given a least square equation as seen in equation (1),

$$y = f(x) + \varepsilon \qquad (1)$$

Where y is the given output of an unknown smooth function $f(x)$, x is a vector of inputs and ε the noise term for an input/output data set $\{(x_i, y_i)/1 \le i \le M\}$. The GT is derived from the Delta function ($\partial_M(k)$) of the input vectors and the Gamma function ($\gamma_M(k)$) of the output values which is based on $N[i,k]$, which are the $k^{th}(1 \le k \le p)$) nearest neighbours $x_{N[i,k]}(1 \le i \le M)$ for each vector $x_i(1 \le i \le M)$ and $y_{N[i,k]}$ of the corresponding output value y_i. This is mathematically expressed thus:

$$\partial_M(k) = \frac{1}{M}\sum_{i=1}^{M}\left|x_{N(i,k)} - x_i\right|^2 \qquad (1 \le k \le p) \qquad (2)$$

$$\gamma_M(k) = \frac{1}{2M}\sum_{i=1}^{M}\left|y_{N(i,k)} - y_i\right|^2 \qquad (1 \le k \le p) \qquad (3)$$

A plot of the $\gamma_M(k)$ Vs $\partial_M(k)$ will give a least square regression fit defined by equation(4).

$$\gamma = A\partial + \Gamma \qquad (4)$$

Where A and Γ are the gradient and noise variance values respectively. The GT algorithm will use the points $\gamma_M(k)$ and $\partial_M(k)$ to produce a 2-D scatter graph, from which a set of coordinates ($\partial_M(k), \gamma_M(k)$) ,$(1 \le k \le p)$ can be identified for the p nearest neighbour. The plot can also be represented as a 3D histogram (see figure 1). The vertical intercept Γ derivable from equation (4), offers an estimate of the MSE achievable when using a modelling technique for any unknown smooth functions of continuous variable. The gradient (A) gives a measure of the complexity of the unknown function with the preferred scenario being a low MSE and shallow gradient (Jonathan et al, 2003).

M-test: Given a near neighbour value $p_{\max}$ for the estimated best MSE, the analysis of the variation of Γ for increasing data points (M) to determine the minimum number of data points ($M_1...M_m$) required to model a smooth unknown underlying function is established through a process referred to as $M-test$. The theory behind the $M-test$ analysis is such that, for a particular $p_{\max}$ the output is believed to exhibit large variation in Γ at small M but gradually stabilizes at higher value of M by producing an asymptote at a given M and a steady or insignificant variation within a given M-range. The asymptotic point indicates the true noise variance inherent in the data and equally gives an insight into the minimum amount of data required for accurate determination of good data model(s) for the prediction of the output variable.

Model Identification: The next step is the feature selection of all possible input combinations derivable from available inputs which are referred to as the embeddings. In Gamma analysis, an embedding is designed by a string of '1's and '0' called a mask. The '1's indicates an inclusion of a given input and '0's indicates an exclusion of a particular input parameter. The full embedding search and the increasing embedding search heuristics are used in this paper to reveal possible combinations of inputs and on the basis of the result returned by the gamma analyses. Judgement is made on the suitability of a particular combination for the production of data models using ANN algorithm.. Given m number of inputs, the full embedding returns $2^m - 1$possible embedding (Jones *et al.*, 2001)

3. Artificial Neural Networks (ANN)

The power of neural network in modelling complex mappings and system identification as has been done by Kohonen and Ito have encouraged many researchers to explore the use of ANN model in many real world applications (Krishnaiah, *et al,* 2007). ANN are simply mathematical techniques configured as a parallel–distributed processing structure systematically arranged in three sets of layers consisting of element called neurons. The input layer receives data, the middle layer which processes the data by performing mathematical operation on the inputs and the Output layer sends information to users or external devices. Like the brain, ANN consists of a collection of processing elements that are highly interconnected and fashioned to transform a set of inputs to a set of desired outputs. The weights associated with the connection and the specific characteristic of the element often determine the results of the transformation (Abdelhay, 2002). The result obtained at a given time is evaluated and the configuration of the system is refined until the analysis of the data used for training reaches a target level, gaining experience with time as it carries out analysis on the data used for the study. The applications of ANN are based on their ability to mimic the human mental and neural structure to construct a good approximation of functional relationship between past and future values of a series (Remezan, *et al.*, 2008). Several algorithms exist for the analysis and evaluation of data using ANN, however, the backpropagation algorithm which uses gradient decent and gradient decent with momentum are often used despite the fact that they require low learning rate for stable learning. Algorithms like conjugate gradient (CG) and Bryoden-Fletcher-Goldfarb-Shanno(BFGS) embedded in WinGamma which is used for this analyses are faster and they make use of numerical optimization techniques. Practical applications have shown that it most convenient and effective to use two-layer architecture (Jones, 2004) given the weakness of single layer perception revealed by Minsky and Papert (1969).

4. Data and Methodology

The data for analyses which include gun-bellani radiation(G_{Gb}), maximum temperature T_{mx}, minimum temperature (T_{mn}), rainfall(R_f), relative humidity measured at 9:00 hours (Rh_{09}) and the derived global solar radiation(G_{sr}) and the extraterrestrial radiation (G_0) used for this study were obtained from data records at the Nigeria Meteorological Agency (NIMET) located in Oshodi, Lagos, Nigeria. As in Falayi, EO. *et al*., (2008)., G_{Gb} is converted to G_{sr} using relationship expressed in equation (5).

$$G_{sr} = (1.35 \pm 0.176)G_{Gb} \;\; MJ/m^2 day \qquad (5)$$

The geographical location of Warri is latitude 5^0 31'N and longitude 5^0,45'E at an altitude of about 5 metres. The daily data records covered a period of seven years (2003-2009) having many days of missing data which were excluded leaving a total of 2201 unique data points in a comma separated variable format (csv) data named *warri-solar.csv*. Gamma test analysis was performed using the WinGamma software to determine the noise variance levels for the different sets of input combination using a near neighbour number, $p_{\max}$ of 10. The analysis on the variation of Γ for increasing data points for the different mask is done to ascertain the threshold point of the noise stability range and hence the data length to be used in developing the artificial neural network data models. As reported in Evans and Jones (2002) and using $k = p$ and $p = p_{\max}$ in equation (2) and (3), an analogy of the basic heuristic governing the noise determination and simulation process by means of the gamma test technique is exemplified algorithmically as follows;

Procedure Gamma (or Near neighbour Test) (data)
{data is an array of points $x(i), y(i)$ *for* $(1 \le i \le M)$ *where* $x(i)$ *is a real vector of dimension* m *and* y *is a real scalar *)*
For i=1 to M (compute x-nearest neighbour list for each data point in O(MlogM) time)
For $p = 1$ *to* $p_{\max}$ ($p_{\max}$ =10)
Compute $N[i, p]$ *where* $x(i)$ *is the* p^{th} *nearest neighbour to* $x(i)$.
end for p
end for i
For $p = 1$ *to* $p_{\max}$
Compute $\partial_M(p)$
Compute $\gamma_M(p)$
End for $p_{\max}$
Perform least square fit on coordinates $\{\partial_M(p)\ \ \gamma_M(p), 1 \le p \le p_{\max}\}$
Obtaining (say) $\gamma = A\partial + \Gamma$
Return (Γ, A*)*

All embeddings for the 5-input/1-output system are automatically generated using the full embedding search. Analyses of the effects of increasing the inputs in succession from 1-5 starting with the extraterrestrial radiation G_0 were also carried out using the increasing embedding search algorithm. After identifying the different input combination reflective of the possible data models, the ANN suite embedded in WinGamma is then deployed with a 5-7-7-1 network topology and learning rate of 0.25 to built data models using the Two Layer Back propagation (2L BP), CG and BFGS learning algorithms.
The performances of the three algorithms are compared using three statistical indicators: correlation coefficient of determination (R^2) root mean square error (RMSE) and the gradient (A).

$$R^2 = \frac{\left[M(\sum G_{sr(m)} G_{sr(p)}) - (\sum G_{sr(m)})(\sum G_{sr(p)})\right]^2}{M\sum G^2_{sr(m)} - (\sum G_{sr(m)})^2 * M\sum G^2_{sr(p)} - (\sum G_{sr(p)})^2} \tag{6}$$

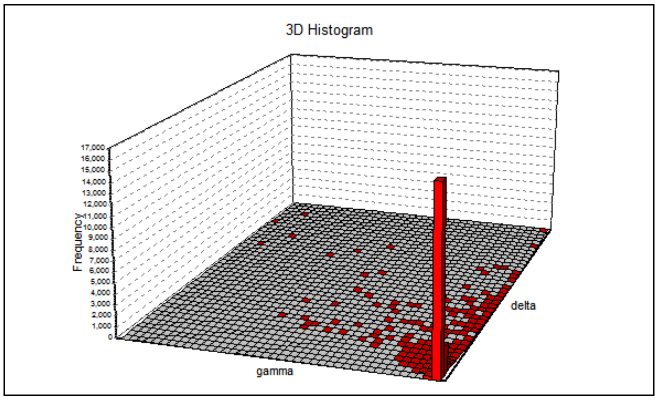
$$RMSE = \sqrt{\frac{\sum_{i=1}^{M}\left(G_{sr(p)} - G_{sr(m)}\right)^2}{M}} \tag{7}$$

Where $G_{sr(m)}$ and $G_{sr(p)}$ are the measured and predicted values of global solar radiation, respectively. M is the number of data point used in the evaluation process. The MSE is critical in highlighting the extent to which the $G_{sr(p)}$ is predictable by the climatological inputs used in predicting the $G_{sr(m)}$ The MSE refers to that which is achievable utilizing a modelling technique for unknown smooth function of continuous variables (Evans & Jones, 2002). When assessing the performance of the models using the statistical indicators, we ordinarily expect to have low RMSE though; few errors in the sum can result in the increased values of the indicators.

5. Results and Discussion

5.1. Gamma Test Analyses

GT analysis on *warri-solar.csv* with 2201 unique data points of the 5-inputs/ 1-output pair and $p_{\max} = 10$ resulted in a gamma statistic, gradient, standard error, expected absolute error and $V - ratio$ values that show that the data has a high capacity to predict the output. An evaluation of the noise variance when all the input parameters are included resulted in Γ = 0.0005. The noise variances for the alternate mask, 11110, 11101, 11011, 10111 and 01111 had values of 0.08890, 0.0006, 0.0009, 0.0161 and 0.0227 respectively. The increasing embedding search analyses gives varying gamma values from 0.0005 for 11111 to 0.0542 for mask 00001.



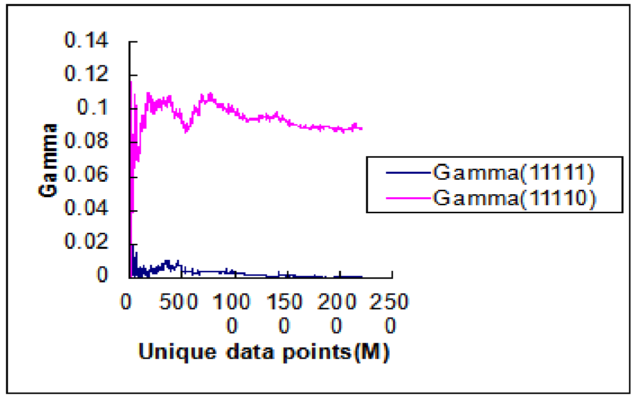


Fig. 1 Histogram of delta/gamma plot frequencies
Fig. 2 Graphical analysis of M-test for warri-solar.csv for Mask 11111 & 11110

This noise variance within the data is the result of the different delta values for the different gamma points which is illustrated using the frequency histogram shown in Fig.1. The 3D histogram illustrates the 'wedge shaped' area resulting from the analysis of *warri-solar.csv*. Though the plots are infrequent and variant, the results shows partial indications of an 'empty wedge' or strong outliers revealed by the actual frequency distribution of points in the scatter plot leading to the deduction that the data is somewhat noisy but the modelling is reasonably feasible. If similar inputs were to produce similar outputs, there should be a cluster of points around the origin which will provide a useful means of visualizing patterns within the numerical input/output data. A 2D Gamma test graphical analysis used in this study as well, provided for the different input combinations identified two indicators (the vertical intercept(Γ) which gives an estimate of the best MSE and the gradient(A) revealing the models function complexity). Results show a low standard error (SE) value of 0.0007 with a relationship between input and output being represented by a moderately complex function with gradient of 0.1005. The complexity of the defining function was found to increase for other possible combination represented by the following mask (01111, 10111, 11011, 11101, and 11110). For the different estimated MSE, the M-test analysis showed different values of the minimum data point required to produce a smooth data model (See Fig.2). The values ranged between 1200 data points for mask 11111 to 1800 for 11110. The result of the GT for different or alternate mask combination is summarized in table 1.

Input Parameter Combination	*Gamma* (Γ)	*Gradient* (A)	*Standard Error* (SE)	*V-ratio*	*Data Length* (M)	*Mask*
T_{mx}, T_{mn} R_f, Rh_{09}, G_o	0.0005	0.1005	0.007	0.0021	1200	11111
Tmx, Tmn, R_f, Rh_{o9}	0.0889	0.2053	0.0034	0.3556	1800	11110
T_{mx}, T_{mn}, R_f, G_o	0.0006	0.1314	0.0005	0.0023	1000	11101
T_{mx}, T_{mn}, R_f, G_o	0.0009	0.1174	0.0004	0.0036	1600	11011
T_{mx}, T_{mn}, Rh_{o9}, G_o	0.0161	0.0914	0.0008	0.0644	1400	10111
T_{mx}, R_f, Rh_{o9}, G_o	0.0227	0.1170	0.0126	0.0907	1400	01111

Table 1: Results of gamma analyses for warri-solar.csv (pmax=10)

5.2. Results of Increasing Embedding Search Analysis

The critical contribution of the extraterrestrial radiation and the relevance of other inputs as revealed in table 2 are rather instructive. The results were generated automatically using increasing embedding search heuristic of the Gamma test analyses. The models with T_{mx}, R_f, Rh_{o9}, G_o and Tmx, Tmn, R_f, Rh_{o9} inputs corresponding to mask 01111 and 11110 respectively are identified to be the least reliable for the model building activity.

Input Parameter Combination	Gamma Γ	Standard Error (SE)	R^2	Mean Absolute Error	Mask
G_o	0.054236	0.0033133	0.78306	0.23289	00001
Rh_{09}, G_o	0.034998	0.001557	0.86001	0.18708	00011
R_f, Rh_{09}, G_o	0.032774	0.0011642	0.86890	0.18104	00111
T_{mn}, R_f, Rh_{09}, G_o	0.022677	0.0012633	0.90929	0.15059	01111
T_{mx}, T_{mn}, R_f, Rh_{09}, G_o	0.0005357	0.0007056	0.99786	0.02145	11111

Table 2: Results gamma analyses with warri-solar.csv for increasing embedding

The decreasing values of the Standard error (0.0033133-0.0007056) and Mean Absolute Error (0.23289-0.02145) and the corresponding increasing R^2 (0.7831-0.9979) from mask 00001 – 11111 indicates that, better results are guaranteed with the inclusion of all relevant inputs parameters. For the purpose of this paper, the sull embedding heuristics of the *Model Identification* analysis produced 31 possible models out of which we have chosen six(6) and another five(5) sets of input combinations shown in table 2 resulting from the increasing embedding search for purposes of analysis.

5.3. ANN model development using warri-solar.csv

The input parameters as shown in table 1 using $p_{\max} = 10$ and the minimum unique data points (M) represented as data length for the different combinations generated during the Model identification process were used to produce the ANN models results shown in table 3. The degree of complexity of the defining function which is originally unknown is indicated by the slope (A). The results are produced by a 5-7-7-1 neural network architecture using multilayer feed-forward back propagation algorithms. Learning rate of 0.25, momentum of 0.1 and a regularization constant of 1E-7 was used. The average MSE reached using the training data were between 0.0013 for 1200 (M) of 11111 to 0.0885 for M=1800 of 11110 for the Back propagation 2-layer algorithm and the other two learning algorithms (CG and BFGS). The number of cycle or epochs for ranged from 130 to 14,000 cycles/second.

Input Parameter Combination	Training Data			Validation Data			Mask
	R^2	A	RMSE	R^2	A	RMSE	
T_{mx}, T_{mn} R_f, Rh_{09}, G_o	0.9988	0.9919	0.2585	0.9982	0.9934	0.2590	11111
T_{mx}, T_{mn}, R_f, Rh_{09}	0.9994	1.0109	0.2561	0.9991	1.0105	0.2557	11110
T_{mx}, T_{mn}, R_f, G_o	0.9975	1.0031	0.2961	0.9978	1.0057	0.2726	11101
T_{mx}, T_{mn}, Rh_{09}, G_o	0.9321	0.9608	1.7019	0.9336	0.9822	1.4880	11011
T_{mx}, R_f, Rh_{09}, G_o	0.8560	0.9979	1.8678	0.8377	1.0056	1.8900	10111
T_{mn}, R_f, Rh_{09}, G_o	0.3545	0.8992	3.8296	0.2685	0.9336	3.9415	01111

Table 3.Statistical performance indicators of ANN data model for predicting
G_{sr} Using the Backpropagation 2-layer neural network algorithm

Input Parameter Combination	Training Data			Validation Data			Mask
	R^2	A	RMSE	R^2	A	RMSE	
T_{mx}, T_{mn} R_f, Rh_{09}, G_o	0.9964	0.9996	0.3180	0.9959	0.9985	0.3391	11111
T_{mx}, T_{mn}, R_f, Rh_{09}	0.9984	1.0021	0.2373	0.9980	1.1005	0.7541	11110
T_{mx}, T_{mn}, R_f, G_o	0.9974	0.9988	0.2972	0.9969	1.0026	0.3053	11101
T_{mx}, T_{mn}, Rh_{09}, G_o	0.9258	0.9821	1.5505	0.9335	1.0055	1.3718	11011
T_{mx}, R_f, Rh_{09}, G_o	0.8996	0.9953	1.7315	0.8768	0.9987	1.8363	10111
T_{mn}, R_f, Rh_{09}, G_o	0.4589	0.9493	12.184	0.2957	0.9830	13.157	01111

Table 4: Statistical performance indicators of ANN data model for predicting G_{sr} Using the Conjugate Gradient (CG) neural network algorithm

Input Parameter Combination	Training Data			Validation Data			Mask
	R^2	A	RMSE	R^2	A	RMSE	
T_{mx}, T_{mn} R_f, Rh_{09}, G_o	0.9977	1.0016	0.2801	0.9970	1.0029	0.3002	11111
Tmx, Tmn, R_f, Rh_{09}	0.9983	1.0009	0.2433	0.9984	0.9984	0.2136	11110
T_{mx}, T_{mn}, R_f, G_o	0.9958	0.9965	0.3787	0.9954	1.0009	0.3621	11101
T_{mx}, T_{mn}, Rh_{09}, G_o	0.9261	0.9840	1.5411	0.9336	1.0076	1.3814	11011
T_{mx}, R_f, Rh_{09}, G_o	0.9027	0.9999	1.7270	0.8714	1.0009	1.8717	10111
T_{mn}, R_f, Rh_{09}, G_o	0.4267	0.9501	3.5426	0.2834	0.9803	3.5320	01111

Table 5: Statistical performance indicators of ANN data model for predicting G_{sr} Using the Bryoden-Fletcher-Goldfarb- Shanno (BFGS) neural network algorithm

Table 3 shows that there is a close agreement between measured and predicted global solar radiation with R^2 for both training and validation set raging from 0.2685 (11110) to 0.9988 (11111) for the Backpropagation 2-Layer algorithm, Table 4 reveals R^2 performance measures of 0.2957 (11110) to 0.9964 (11111) for the Conjugate Gradient and Table 5 also gives R^2 of 0.2834 (11110) to 0.9977 (11111) for the Bryoden – Fletcher-Goldfarb-Shanno. The low correlation value of R^2 for 11110 shows that the exclusion of G_0 as reflected in mask 11110 affects the predictive potential of the smooth data model significantly. This emphasizes the fact that extraterrestrial radiation is a critical input factor in building models that predict G_{sr} for such and similar locations in the country.

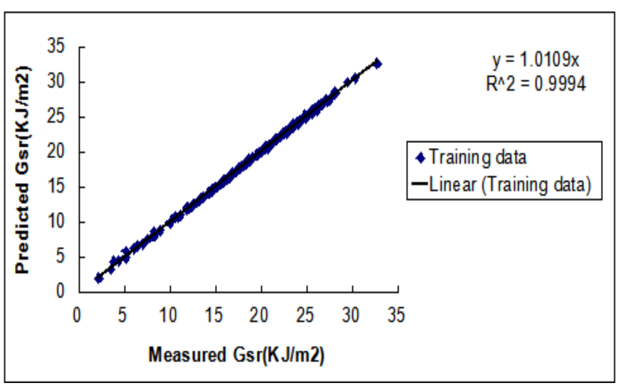


Fig. 3: Predicted and measured G_{sr} for the BP 2L NN algorithm using the training data

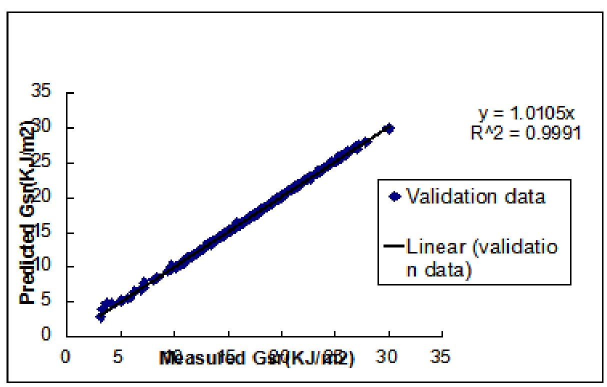


Fig. 4: Predicted and measured G_{sr} for the BP 2L NN algorithm using the validation data set.

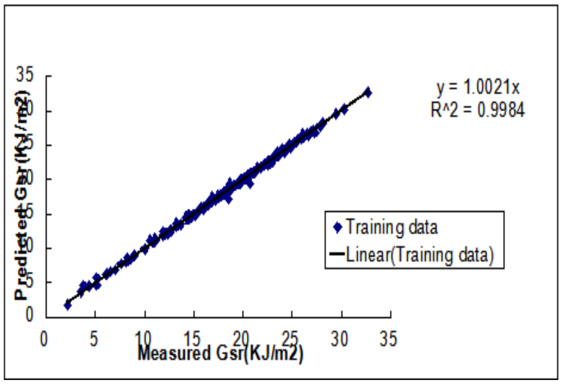


Fig. 5: Predicted and measured G_{sr} for the CG NN algorithm using the training data

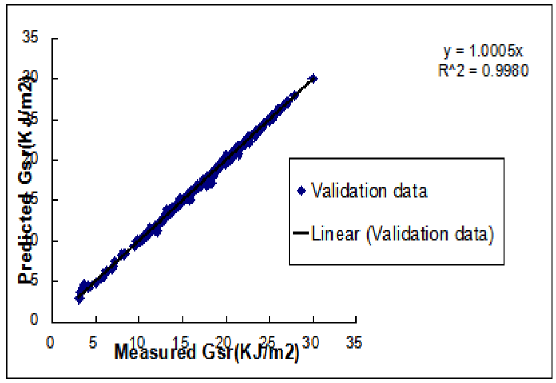


Fig. 6: Predicted and measured G_{sr} for the CG NN algorithm using the validation data set.

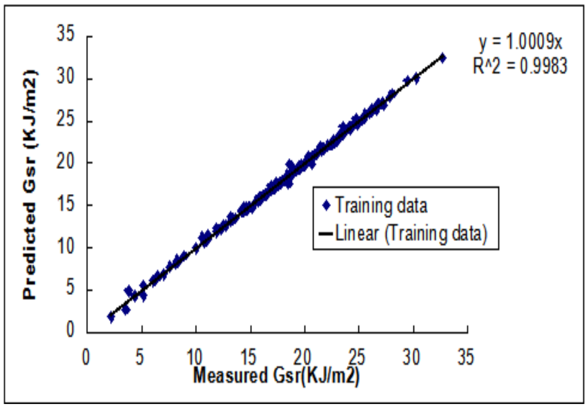


Fig. 7: Predicted and measured G_{sr} for the BFGS NN Algorithm using the training data

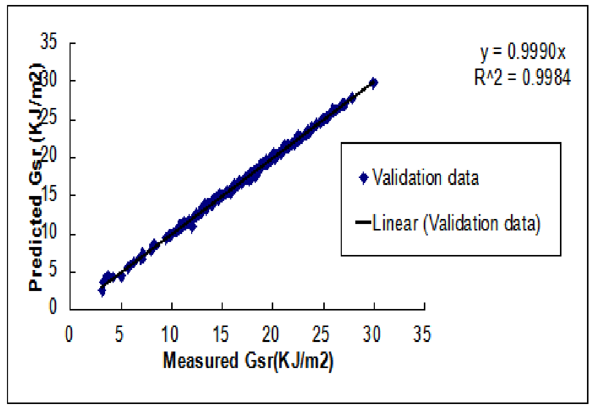


Fig. 8: Predicted and measured G_{sr} for the BFGS NN algorithm using the validation data set.

Figure 3 to figure 8 represent sampled graphical pictures of the relationships between the measured and predicted G_{sr} generated by the different algorithms as specified in the labels. The measured and predicted values of G_{sr} are observed to correlate significantly.

6. Conclusion

Artificial neural network technique for modelling global solar radiation (G_{sr}) using back- propagation 2 layer, Conjugate Gradient and Bryoden-Fletcher-Goldfarb-Shanno algorithms has been presented. The initial analysis of the data using Gamma test algorithm which seeks to determine the best MSE, the minimum amount of data and the possible relevant combinations of inputs for good model construction work has also been done. The performance parameters such as the Gamma Statistics, the gradient, standard error, V-ratio, root mean square error, (RMSE) and the co-efficient of determination have been represented both for the selected mask of *warri-solar.csv* and the models constructed using the ANN algorithms. The result shows that aside mask which excluded the extraterrestrial radiation, the indicators for all other mask shows that the GT-ANN methods adopted here are suitable and relatively more precise than the direct implementation of ANN in modelling G_{sr} . The estimated values are in good agreement with measured values of G_{sr} . The arbitrary partitioning of data for training and testing is approached more scientifically with the application of M-test which gives different m values for the selected mask representing different percentages of the data sets, which does not necessarily reflect the traditional use of 70% for training and 30% for testing. In this study with a total of 2201 data points of the daily values, 1000, 1200, 1400, 1600 and 1800 represents 45.4%, 54.5%, 63.6% 72.7% and 81.8% respectively while the validation data is extracted from the remaining percentage for the various mask shown in table 1. From the performance indices, the model that excludes the relative humidity measured at 9:00hours with mask defined by 11101 is the best model for the location under investigation with R^2 values of 0.9994 for the Back propagation Two Layer algorithm, 0.9984 for the CG and 0.9983 for the BFGS. Hence, Rh_{09} is the least significant input parameter among the inputs, used for this study. The results indicate a good promise for simulation and evaluation of global solar radiation using Gamma test, even for many other locations in Nigeria where there is dearth of G_{sr} records. A comparative study of the performance of other nonlinear modelling methods that use data directly without an initial subjection to Gamma analysis is being contemplated.

7. References

1. Abdelhay A. M. (2002); Application of Artificial Neural Networks to predict the carbon contact and the grain size for carbon steels. Egypt J. Soc. Vol, (25) no (2) pp 224 – 242
2. Abdulazeez M. A. (2011) Artificial Neural Network estimation of global solar radiation using meteorological parameters in Gusau, Nigeria. Scholars Res. Library, 3(2): 586 – 595
3. Angstrom, A. (1924) Solar and terrestrial radiation QJR meteoric SOC (UK), 50 PP 121-25.
4. Barbero, F. J., G. Lopez and F. J. Batlles (2006). Determination of daily solar ultraviolet radiation using statistical models and artificial neural networks. Ann. Geophys, 24, 2105 -2114.
5. Conley, J.R. and T. B. McKee (1983). An analysis of solar radiation data for Fort Collins, Colorado. Climatology report No. 83 – 2 PP 22 – 37.
6. Dogniaux, R. Proceedings of the symposium on solar radiation, Smithsonian Institution, Washington DC 1973, 50-60
7. Evans D, Jones AJ. (2002) A proof of the gamma test. Proceedings of the Royal Society. Series A
8. 458(2027): 2759-2799
9. Ewona, I.O (2011). Modelling of Basic Climatic Parameters in Nigeria using Correlation Analysis and Artificial Neural Networks. PhD thesis, Department of Physics, University of Calabar, Calabar-Nigeria.
10. Falayi, EO., Adepitan JO and Rabiu, AB (2008) ' Empirical models for the correlation of global solar radiation with meteorological data for Iseyin, Nigeria.' Pacific Journal of Science and Technology. Vol.9 No. 2 : 583-590.
11. Gopinathan KK, (1988) 'A general formula for computing the coefficients of the correlations connecting global solar radiation to sunshine duration'. Solar Energy 41: 499-502
12. Ibeh GF., Agbo GA., and Ibeh LM.(2012) Application of Artificial Neural Network for estimation of annual Global solar radiation of Warri with relative humidity" Pacific Journal of Science and Technology,. Vol.13. No. 1: 193-196.
13. Jonathan J. Corcoran, I. D Wilson, and J. A ware (2003): predicting the geo-temporal variations of crime and disorder international journal of forecasting 19:623-634.
14. Jones A. J. 2004. New tools in non-linear modeling and prediction. Computational Management Science 1: 109 – 149. DOI: 10. 1007/S 0287 – 0030006 – 1
15. Jones A.J, Margetts, S and Durrant, P (2001) The WinGamma™ User Guide. University of Wales, Cardiff 1998-2001. pp76
16. Krishnaiah, T., S. S Ruo, K. Madhumurthy and K. S. Reddy 2007 Neural Network approach for modeling global solar radiation. Journal of Applied Sciences Research. 3(10): 1105 – 1111.
17. Leckner, B. The Spectral Distribution of solar radiation at the earth's surface - Elements of a Model, Solar Energy 20, 143 - 150, 1978.
18. Lopez, G, F. J. Batlles, J. Tovar – Pescador (2004): selection of input parameters to model direct solar irradiance by using artificial neural networks. ENERGY (30): 1675 – 1684.
19. Minsky M. and Papert S. 1969. Perceptions MIT Press: Cambridge
20. Ninomiya H. (1994) Study on application of AMe DAS metrological data to the simulation of building heat environment Ph.D thesis, University of Tokyo, Tokyo 55 pages.
21. Prescott J. A. (1940) Evaporation from water surface in relation to solar radiation. Trans. Royal. Society. Austr. 64, 114
22. Renji Remesan, Muhammad Ali Shamin and Dawei Han (2008): Non-linear Modeling of daily solar radiation using the gamma test, Btts 10th National Hydrology Symposium Exeter.
23. Remesan R, M.A. Shamin and D. Han (2008) Model data selection using gamma test for daily solar radiation estimation. Hydrological processes 22 Pp 4301 – 4309.
24. Yang K, G. W. Huang and N. Tamai; A hybrid model for estimating global solar radiation Pergamon Pp 13 – 22, 2001.
25. Yeboah-Amankwah D. and Agyeman K. (1990) Differential Angstrom model for prediction insolation from hours of sunshine. Solar Energy 45, 371 – 377