Estimation of daily global solar radiation in tropical region in south India using linear regression and Artificial Neural Network

J. Vishnupriyan[1], P.S. Manoharan[2], V. Annapeachi[3]

[1]Research scholar, [2]Associate Professor, [3]PG student

Department of Electrical and Electronics Engineering, Thiagarajar College of Engineering, Madurai, Tamil Nadu, India

vishnupriyanj@gmail.com, psmeee@tce.edu, joe251294@gmail.com

Abstract - *Estimation of daily global solar radiation (GSR) is important feature for non availability of observatory station in the study area. In this study, daily GSR models are developed for radiation for Madurai location in Tamil Nadu, India. The three years hourly solar radiation data collected from Tamil Nadu Agriculture University using pyranometer measurement data. The regression coefficient of selected radiation models were performed with curve fitting tool in MATLAB. In addition, the developed models are validated with Artificial Neural Network to estimation the prediction accuracy. The accuracy of the models was evaluated by statistical indexes. This estimation results further validated through pyranometer measurement data.*

Keywords: Approximate solar radiation, Artificial Neural network, Levenberg-Marquardt, Curve fitting.

1. Introduction

The solar energy is the one of the best decentralized power generation for urban, rural and remote location. The modeling and designing of solar based renewable energy system requires estimation of daily GSR for a specific location. Many of previous works are focused on monthly average daily GSR model development. The authors in [1] have developed a model which is based on the correlation between the inputs parameters has been evaluated. The different network configuration of direct normal irradiance and global radiation is evaluated in the literature [2]. Estimation of the daily GSR by day of the year is proposed in six cities of Yucatan peninsula, Mexico [3]. In this case study, the interpolation the missing and erroneous data problem could be evaluated.

The Liu and Jordan approach is utilized to estimate mean hourly global radiation value based on clearness index and sunshine duration [4]. The correlation between two input variable models is presented for monthly mean value for the Indian location based on extraterrestrial radiation [5]. Additionally, the daily temperature, wind speed, relative humidity and particular mater are the input key parameter for estimating daily GSR [6]. The short term, medium and long term solar irradiance predictions by trend and level factor in time series analysis using Pro-Energy algorithm is obtained [7]. The over fitting and outlier observation problem are overcome by using the (FRF-SVM) approach [8, 9].

The statistical indicator such as root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), mean absolute percentage error (MAPE), absolute fraction of variance (R^2) are performed and the accuracy of the developed model is evaluated by using Artificial Intelligence[10-12]. The comparative study of forecasting global horizontal radiation using SVM, FRF and QKSVM-FR for number of selected variables is performed [13]. The design of hybrid photovoltaic–wind system for twelve location the accuracy and model has developed with ANN model [14]. The different methodology and importance of selection of different solar radiation are obtained and analyzed in many literatures. The estimation of daily solar radiation based on diffuse, extraterrestrial, clearness index is detailed studied in the literatures [15-17].

The previous studies indicates that there is no comprehensive works on estimating daily GSR models, particularly in south India, where a great amount of solar potential exist. In this study, a new strategy is applied for daily GSR model estimation. The irradiance values above than 120W/m^2 are considered as approximate bright sunshine for that particular day.

2. Methodology adopted

2.1. Resource assessment

Resource data’s are collected from Tamil Nadu Agricultural University (TNAU), Madurai and Ministry of New and Renewable Energy (MNRE). The collection of resource data are temperature and solar radiation for Madurai location. The geographical location of the study region is given in Fig.1. This study area comes under tropical climatic region. This area have average temperature in the ranges between 18°C and 40°C.

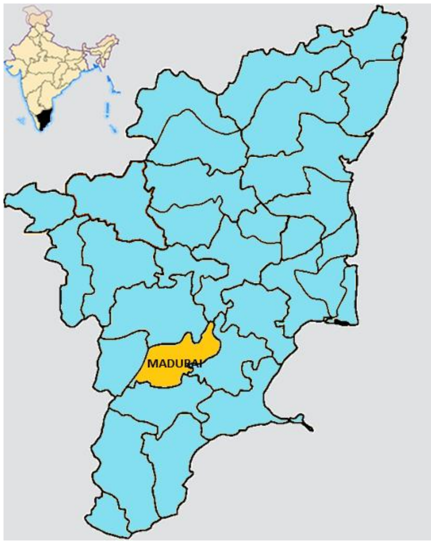

Fig. 1. Location of study region

2.2. Mathematical model development

The hourly solar irradiation data of two years for Madurai location is taken as input variable. In the present study, the models were developed by correlating the data of $\left(\frac{\bar{s}}{\bar{s}_0}\right)$. The approximate sunshine duration on the horizontal surface is considered about 120 MJ/ (m2 day) of measured solar radiation and the maximum possible radiation $(\bar{s}_0)$ based on measured data from the considered location.

$$\bar{H}_0 = \frac{24}{\pi} Hsc \left(1 + 0.033\cos\left(\frac{360}{365}\times n\right)\right) X \left(\cos\emptyset \cos\delta \sin\omega s + \frac{\pi}{180}\omega s \sin\emptyset \sin\delta\right) \quad (1)$$

Where, H_{sc} is the solar constant with the value of 1367 W/m^2. The tilt angle (δ) of solar panel can be mathematically calculated as;

$$\delta = 23.45° \sin\left[\frac{360\,(284+n)}{365}\right] \quad (2)$$

$$s_0 = -\tan\emptyset \tan\delta \quad (3)$$

In this study, the regression analysis is utilized for model development based on approximate solar radiation. The total measured solar radiation dataset was separated in to two parts namely training dataset and testing dataset. The training dataset consist of 2/3rd of the data used to develop the mathematical model while validation dataset consist of the remaining 1/3rd of the data to test models.

2.3. Evaluation criteria

Mean absolute percentage error (MAPE): It expresses the accuracy of the data. When its value is zero, the predicted value is closer to observed value. It may be expressed as;

$$MAPE = \frac{1}{p}\sum_{i=0}^{p}\left|\frac{(H_{i,e}-H_{i,m})}{H_{i,m}}\right| \times 100 \quad (4)$$

Where $H_{i,e}$ is measured value, $H_{i,m}$ is predicted value, and n is the total number of observations.

Mean error (ME): It indicates the deviation from the measured data. It provides long term performance. Mean error is given by;

$$ME = \frac{1}{p}\sum_{i=0}^{p}\left(H_{i,e} - H_{i,m}\right) \quad (5)$$

Root Mean Square Error (RMSE): It is the standard deviation between predicted value and actual value. This gives the short term performance. RMSE is given by;

$$RMSE = \frac{\sqrt{\left(\frac{1}{p}\sum_{i=1}^{p}(H_{i,e}-H_{i,m})\right)^2}}{\frac{1}{p}\sum_{i=1}^{n}(H_{i,m})} \times 100 \quad (6)$$

Mean square error (MSE): If the MSE is zero then estimated and measured are approximately equal.

$$MSE = \frac{1}{p}\sum_{i=0}^{p}\left(H_{i,e} - H_{i,m}\right)^2 \quad (7)$$

Goodness fit (R^2): If the value of R^2 is one, then the relative correlation between the estimated and measured value is approximately equal.

$$R^2 = \sum_{i=1}^{n}\frac{(H_{i,e}-H_{i,m})^2}{(H_{i,e}-H_{i,m})^2} \quad (8)$$

3. Model development

3.1. Curve fitting method

Six models developed to estimate the daily GSR on horizontal surface. The curve fit application of the three existing models is described in the literature [15]. The curve fitting, regression coefficient and statistical analysis of the selected models were performed with curve fitting tool in MATLAB.

Linear first order model: The authors in the reference [15] implemented the simple empirical model which is used for interpolation or extrapolation.

$$\frac{\bar{H}}{H_0} = a\left(\frac{\bar{s}}{so}\right) + b \quad (9)$$

Quadratic order: Polynomial of degree two is used to characterize data using global fit.

$$\frac{\bar{H}}{H_0} = a\left(\frac{\bar{s}}{so}\right)^2 + b\left(\frac{\bar{s}}{so}\right) + c \quad (10)$$

Exponential: If the coefficient associated with exponential is negative then, its result in decreased value. If the coefficient associated with exponential is positive then, its result in increased value.

$$\frac{\overline{H}}{H_0} = \mathrm{a}\ exp^{\mathrm{b}\left(\frac{\bar{s}}{\bar{s}_0}\right)} \tag{11}$$

Fourier first order: The Fourier series is a sum of sine and cosine function that is used to describe the periodic data. It is represented in the trigonometric form.

$$\frac{\overline{H}}{H_0} = a + bcos\left(\left(\frac{\bar{s}}{\bar{s}_0}\right) \times c\right) + d\sin\left(\left(\frac{\bar{s}}{\bar{s}_0}\right) \times c\right) \tag{12}$$

Where *a*, *b* and *c* are the coefficient parameter which fits the curve between the variables by least squares estimation of nonlinear models.

3.2. Artificial Neural Network

The ANN is an interconnected group of artificial neurons that uses a mathematical model for information processing. The back propagating network is utilized, it guaranteed to converge only towards some local minima.

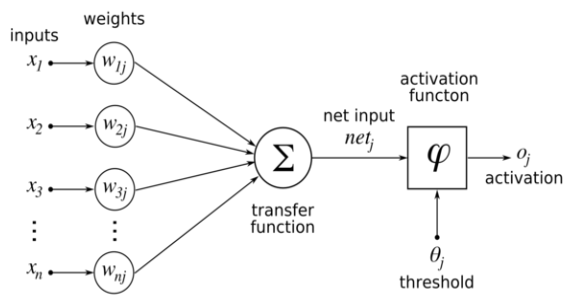

Fig. 2. Schematic view of Neural Network

In Fig. 2, *n* is the number of input parameters and x_i is the input parameter of i^{th} of the network. It is the feed forward network which the first layer is the input layer and second layer is the hidden layer which is interconnected with the output layer. Where W_{ij} is the neuron weight and O_j is the trained output of the network.

4. Result and discussion

4.1. Curve fit performance

The performance of the different models for the present location is presented in Table 1. The Fourier first order model had the best performance for present location according to overall performance (R^2=0.701, RMSE=0.122 MJ/(m^2day) and MAPE=3.85%). Quadratic order model had the second best of overall performance (R^2=0.698, RMSE=0.123 MJ/(m^2day) and MAPE=3.87%). Models 1 and 3 had the worst performance when analyzed by the statistical performance. For the model 1 (R^2=0.694, RMSE=0.124 MJ/(m^2day) and MAPE=3.94%) and for model 3 ((R^2=0.690, RMSE=0.126 MJ/(m^2day) and MAPE=3.96%).

Table 1. Regression coefficients and statical indicators

Models	Regression coefficients				Statical indicators		
	a	b	c	d	MAPE (%)	RMSE (MJ/(m^2day))	R^2
Linear first order model	0.891	0.090	-	-	3.94	0.124	0.694
Quadratic order	0.310	0.542	0.149		3.87	0.123	0.698
Exponential	0.233	1.477	-	-	3.96	0.126	0.690
Fourier first order	1.018	-0.861	0.354	1.16	3.85	0.122	0.701

4.2. ANN configuration

In this preset study, the given real time dataset is also analyzed by changing the number of hidden layer and statistical analysis is performed for each neuron selected in hidden layer. The training dataset consist of 2/3[rd] of the data used to develop the models while validation dataset consist of the remaining 1/3[rd] of the data to test models. The regression analyzed between the measured and predicted value of daily dataset. The result of mean square error corresponding to the neurons in the hidden layer is presented in Fig. 3.

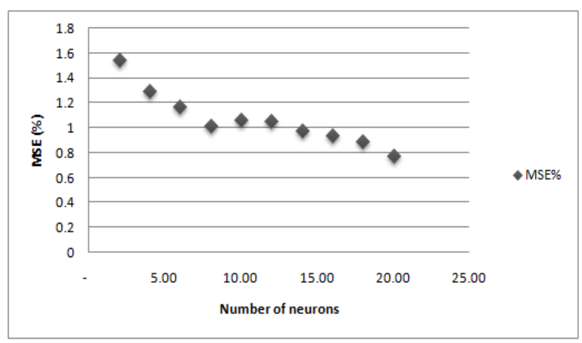

Fig. 3. Performance of neuron numbers

For the modeling and predicting the amount of daily global solar radiation by using Multilayer perceptron Artificial Neural Network (MLP-ANN), one layer and 12 hidden layers have been used. According to the results elaborated in Table 2, increasing the number of neurons the decreasing in the mean square error is obtained. Though the mean square error is decreased, it can be seen that performance of the neuron is decreased with increasing the neurons. The 12 neuron is the first best neuron in the hidden layer. It had analyzed that (MSE=0.06 $MJ^2/(m^4day)$, RMSE=0 MJ/(m^2day) and R=0.88) for the training dataset. The second best is neurons are (2, 4 and 6), since RMSE value is zero and the performance (MSE=0.155 $MJ^2/(m^4 day)$, RMSE=0 $MJ/(m^2 day)$, R=0.854), (MSE=1.301 $MJ^2/(m^4 day)$, RMSE=0 $MJ/(m^{2.} day)$, R=0.790) and (MSE=1.17 $MJ^2/(m^4 day)$, RMSE=0 $MJ/(m^2 day)$, R=0.87).

Similarly while training for the neurons (8, 10, 14, 16, 18 and 20), it had observed that mean square error is low. But the RMSE value of these neurons is higher and this neuron shows the worst performance (for n=8, MSE=1.02 $MJ^2/(m^4 day)$, RMSE=0.130 $MJ/(m^2 day)$, R=0.88), (for n=10, MSE=1.07 $MJ^2/(m^4 day)$, RMSE=0.277 $MJ/(m^2 day)$, R=0.88), (for n=14, MSE=0.098 $MJ^2/(m^4 day)$, RMSE=0.344 $MJ/(m^2 day)$, R=0.89), (for n=16, MSE=0.094 $MJ^2/(m^4 day)$, RMSE=0.04 $MJ/(m^2 day)$, R=0.88), (for n=18, MSE=0.08 $MJ^2/(m^4 day)$, RMSE=0.02 $MJ/(m^2 day)$, R=0.88) and (for n=20, MSE=0.07 $MJ^2/(m^{4.} day)$, RMSE=0.052 $MJ/(m^{2.} day)$, R=0.88). Though the regression correlation of this neuron is increased the error in the training data is also increased. Hence the 12 neuron in the hidden layer shows the best performance, since the correlation of this neuron is higher and the error is zero.

Table 2 Performances of the training phase varying the neuron number

Number of neuron	MSE ($MJ^2/(m^4 day)$)	RMSE ($MJ/(m^2 day)$)	R^2
2	0.155	0	0.854
4	1.301	0	0.790
6	1.170	0	0.870
8	1.020	0.130	0.880
10	1.070	0.277	0.880
12	0.060	0	0.880
14	0.098	0.344	0.890
16	0.094	0.041	0.880
18	0.089	0.02	0.880
20	0.078	0.052	0.880

As it is shown in the Fig. 4, there is the better harmony with the ANN-1-12 model for the estimating the daily GSR. Therefore, approximate global solar radiation parameter as network input parameter is important in predicting the solar radiation.

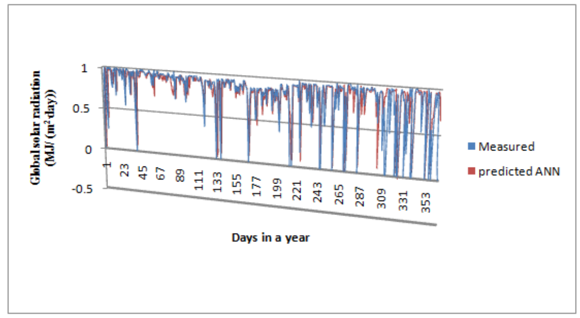

Fig.4. Validation of approximate solar radiation parameter with measured data

The correlation between the measured and predicted value for the testing dataset is approximately 0.7558 is given in Fig. 5. Though for the training dataset the correlation between the measured and predicted dataset is approximately is 0.88, the correlation obtained for the testing dataset is 0.7588. Due to the fact that the correlation is roughly close to 1, it can be inferred that the predicted value is highly accurate. According to the Fig. 5, it can be observed that there is a better correlation between the empirical measurements and prediction value for the testing dataset.

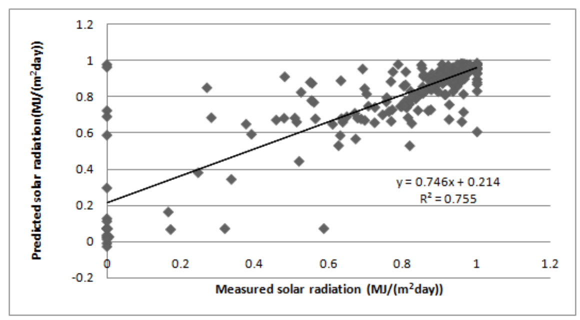

Fig.5. Correlation between the measured and the predicted value for the testing dataset

In the validation result of the given dataset, the accuracy of the model had decreased while compared to the training dataset. In the Table 3 shows the accuracy between the models developed for validation dataset. ANN-1-12 model had the best performance for present location (R^2=0.756, RMSE=0.276 $MJ/(m^2 day)$, MAPE=2.53%) and the Fourier model had the second better performance for the present location (R^2=0.691, RMSE=0.341 $MJ/(m^2 day)$, MAPE=4.02%) and other three models had the worst performance for the present location (R^2=0.688, RMSE=0.377 $MJ/(m^2 day)$, MAPE=4.97%), (R^2=0.675, RMSE=0.398 $MJ/(m^2 day)$, MAPE=5.21%) and (R^2=0.597, RMSE=0.421 $MJ/(m^{2.} day)$, MAPE=5.54%).

Table 3.Comparison of accuracy between the models developed for validation dataset.

Models	MAPE (%)	RMSE (MJ/ (m^2day))	R^2	Performance
Exponential	5.54	0.421	0.597	Poor
Linear first order model	5.21	0.398	0.675	Worst
Quadratic order	4.97	0.377	0.688	Worst
Fourier first order	4.02	0.341	0.691	Better
ANN-1-12	2.53	0.276	0.756	Good

5. Conclusion

The estimation of daily global solar radiation (GSR) in southern part of India, Tamil Nadu was carried out by considering approximate sunshine hours. The estimation results reveals that Fourier first order model had improved accuracy compared to other models in Curve fit tool in MATLAB. The Artificial Neural Network (ANN) contributing the best approach to estimate the daily GSR for the considered study area. For the Madurai location ANN-1-12 model shows the highest accuracy of R^2=0.88 for the training dataset and R^2=0.756 for the testing dataset. In addition, the ANN predicted data is validated with measured data. It shown the improved prediction accuracy.

Acknowledgement

The financial support for the first author, J. Vishnupriyan by the Technical Education Quality Improvement Programme (TEQIP) (Ref: TCE/Dean-R&D/TEQIP/T&RAssistantship/2016-17) and the second author, P.S. Manoharan, University Grant Commission (UGC), Government of India, under Research Award Scheme (RA-2016-18-OB-TAM-7210).

Reference

1. E.F. Alsina, M. Bortolini, M. Gamberi, A. Regattieri, “Artificial neural network optimization for monthly average daily global solar radiation prediction”, Energy Conversion and Management, Vol.20, pp.320-329, 2016.
2. C. Renno, F. Petito, A. Gatto, “ANN model for predicting the direct normal irradiance and the global radiation for a solar application to a residential building”, Journal of Cleaner Production, Vol.135, pp.1298-1316, 2016.
3. V.H. Quej, J. Almorox, M. Ibrakhimov, L. Saito, “Estimating daily global solar radiation by day of the year in six cities located in the Yucatán Peninsula, Mexico”, Journal of Cleaner Production,Vol.141, pp.75–82, 2017.
4. S.A. Kalogirou, S. Pashiardis, A. Pashiardi, “Statistical analysis and inter-comparison of the global solar radiation at two sites in Cyprus”, Renewable Energy, Vol.101, pp.1102–1123, 2017.
5. M. Vakili, S.R. Sabbagh-Yazdi, S. Khosrojerdi, K. Kalhor, “Evaluating the effect of particulate matter pollution on estimation of daily global solar radiation using artificial neural network modeling based on meteorological data”, Journal of Cleaner Production, Vol.141, pp.1275–1285, 2017.
6. D.I. Jeong , A. St-Hilaire , Y. Gratton , C. Belanger , C. Saad, “A guideline to select an estimation model of daily global solar radiation between geostatic interpolation and stochastic simulation approaches”, Renewable Energy , Vol.103, pp.70–80, 2017.
7. Kadir Bakirci, “Prediction of global solar radiation and comparison with satellite data”, Journal of Atmospheric and Solar–Terrestrial Physics, Vol.152-153,PP.41-49,2017
8. F. Baser, H. Demirhan, “A fuzzy regression with support vector machine approach to the estimation of horizontal global solar radiation”, Energy, Vol.23, pp.229–240, 2017.
9. A. Sharma, A. Kakkar, “Development of modified Pro-Energy algorithm for future solar irradiance estimation using level and trend factors in time series analysis”, Journal of Renewable and Sustainable Energy, Vol.9, pp.1-16, 2017.
10. Mauricio Bruno Prado da Silva, Joao Francisco Escobedo, Cicero Manoel dos Santos, Taiza Juliana Rossi, Silvia Helena ModeneseGorla da Silva, “Performance of the Angstrom-Prescott Model (A-P)and SVM and ANN Techniques to estimate the daily Global Solar Irradiation in Botucatu/SP/Brazil”, Journal of Atmospheric and Solar–Terrestrial Physics, Vol.160, pp.11-23,2017.
11. Victor H. Quej, Javier Almorox, Javier A. Arnaldo, Laurel Saito, “ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment”, Journal of Atmospheric and Solar–Terrestrial Physics, Vol.155, pp.62-70,2017.
12. Sayed Saber Sharifi, VahidRezaverdinejad, VahidNourani, “Estimation of Daily Global Solar Radiation using Wavelet Regression, ANN, GEP and Empirical Models: A

Comparative Study of Selected Temperature-Based Approaches", Journal of Atmospheric and Solar–Terrestrial Physics, Vol.149, pp.131-145, 2016.

13. B. Jamil, N. Akhtar, "Estimation of diffuse solar radiation in humid-subtropical climatic region of India: Comparison of diffuse fraction and diffusion coefficient models", Energy, Vol.131, pp.149–164, 2017.

14. H. Jiang, Y. Dong, "Global horizontal radiation forecast using forward regression on a quadratic kernel support vector machine: Case study of the Tibet Autonomous Region in China", Energy, Vol.133, pp.270–283, 2017.

15. Basharat Jamil, Abid T. Siddiqui, "Generalized models for estimation of diffuse solar radiation based on clearness index and sunshine duration in India: Applicability under different climatic zones", Journal of Atmospheric and Solar–Terrestrial Physics, Vol.157-158, pp.16-34, 2017.

16. K.N. Shukla, SarojRangnekar, K. Sudhakar, "Comparative study of isotropic and anisotropic sky models to estimate solar radiation incident on tilted surface: A case study for Bhopal, India", Journal of Atmospheric and Solar–Terrestrial Physics, Vol.1, pp.96-103, 2015.

17. Amrita Das, Jin-ki Park, Jong-hwa Park, "Estimation of available global solar radiation using sunshine duration over south Korea", Journal of Atmospheric and Solar–Terrestrial Physics, Vol.134, pp.22-29, 2015.