Linear Regression Model for Estimating Sustainable Generation: A Case Study in Tamil Nadu

A. Geetha[1], S. Usha[1], J. Santhakumar[2,*], Amrit Kalash[3], Harshit Saini[3], Shashwat Sinha[3]

[1]Department of Electrical and Electronics Engineering, SRM Institute of Science and Technology, Chennai, India
[2]Department of Mechanical Engineering, SRM Institute of Science and Technology, Chennai, India.
[3]Department of Electronics and Communication Engineering, SRM Institute of Science and Technology, Chennai, India.

Abstract

This article aims at developing a statistical model for the prediction of DC and AC generated power from the installed PV plant. A proper understanding of the PV plant characteristics is highly in need of predicting the yield based on the solar and atmospheric parameters. This study focusses on investigating the relationship among the factors such as beam and diffused solar radiations, atmospheric temperature and wind speed for predicting the hourly generated powers. The location involved in the investigation is Chennai city, Tamil Nadu state, India. The meteorological data for the selected location is obtained from NREL and using a simple linear regression model prediction equations for DC and AC solar output power was built using Minitab 16.2.1 version. The methodology used has a capability of better correlation coefficient than the other techniques. The developed regression models show R^2 value of 99.24% and 99% for DC and AC power and the predicted R^2 (R_{pred}) values obtained are 86.54% and 83.22% for DC and AC power respectively.

*Corresponding author. Email: Santhakj@srmist.edu.in

1. Introduction

Recently, the demand for energy increases rapidly as the usage of electrical appliances in day today life increases. Hence the burden on dependency of fossil fuel also increases. The depletion in fossil fuel has a greater negative impact on environment. However the drastic technology developments on sustainable sources such as solar, wind etc. [1, 2]. Out of those solar based power plants are attracting the researchers whose energy is eco-friendly, clean, unlimited and free when compared to other conventional fuels [3].

India being a developing country is much investing on solar energy systems for meeting the energy demands [4]. It is noted that, in India, a particular state called Tamil Nadu alone possess a total PV plant capacity of 4000 MW by the year 2020. It is also to know that the installed plant had a solar generated power for the period of 2019-2020 was 3842 million number of units [5]. The above data reveals the solar

availability in Tamil Nadu and its usefulness for the yield of electrical energy for the current energy demand situation.

Hence the prediction of those yield in a specific location can help the investors for a proper designing and installation at accurate capacity of solar plants. The two different prediction techniques in literature is historical data based or model based methods [6, 7]. Among those regression analysis, artificial neural network, support vector machines and other types of machine learning techniques are most attracted and accurate prediction models used for the prediction of solar panel generated powers in the literature [8 -10].

Based on the land cover area of India, the total potential of the country is found to be 5000 trillion kWh/year. It is estimated that a single square metre of solar panel area can generate maximum of 4-6 kWh/day. National Institute of Solar Energy (NISE) suggested minimum 3% of total waste land in India can able to produce a maximum power of 750 GW [11].

Beam and diffused solar radiation are the two main elements of solar radiation. The addition of beam and

diffused radiation is global solar radiation [12]. This article uses the hourly meteorological data obtained for the selected location. The hottest city in Tamil Nadu is Chennai which is selected as a location for prediction. The latitude and longitude of the selected city are 13.08° N, 80.27° E respectively. Satellite data, Situ measurements and statistical methods are the techniques utilized for collecting information about solar radiations. Buying and installing measuring instruments such as pyrometer, sunshine recorder etc., under situ measurement technique in the location under study for all cases are very expensive and challenging. It is also tedious for proper maintenance and servicing of those instruments. On the other hand, the information obtained from satellite is accurate and reliable however it highly depends on the efficient communication from the satellite stations.

Statistical method called regression modelling overcomes the above difficulties where the available historical data is utilized for generating a set of prediction equations by fitting those empirical data. Linear Regression is a relatively simple technique that may be carried out quickly and reliably. Additionally, these models may be trained quickly and effectively on systems with less computing capacity than other sophisticated methods. When compared to some of the other machine learning methods, linear regression has a much reduced time complexity. Linear regression's mathematical formulae are also very simple to comprehend and interpret. The prediction equations may be of yearly, monthly, daily or hourly as per the available data [13 – 15]. However similar studies on the prediction of generated power from the have been found limited in India. Hence, the presented article focusses on the estimation of DC and AC generated power in a specific location in India which experiences non-seasonal weather conditions over the years.

2. Methodology

Hourly data of the selected locations were collected from National Renewable Energy Laboratory (NREL). Averagely Chennai is subjected to 9 hours of solar radiation per day. The correct location as well the detailed parameters for obtaining the historical data is given in Figure 1 and Table 1 respectively.

Table 1. Parameters to NREL

Parameters	Details
Location	Chennai
Lat (deg N):	12.85
Long (deg E):	80.05
Plant Size (kW):	4
Array	Fixed
Array Tilt (deg):	20
Array Azimuth	180

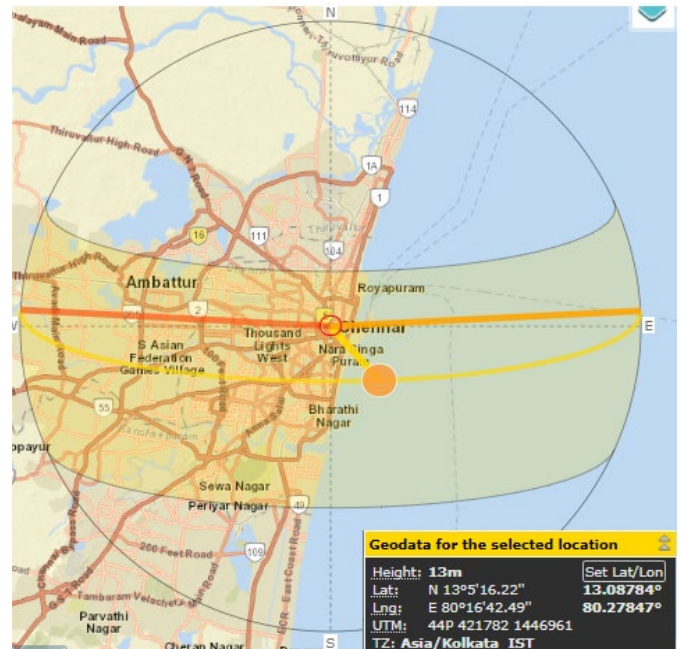|	
(deg):	
System Losses:	14.08
Invert Efficiency:	96
DC to AC Size Ratio:	1.2

Figure 1. Sun path from SunCalc.

The DC and AC hourly generated powers along with direct beam and diffused radiation, ambient temperature and wind velocity of the selected location is gathered.

The prediction of hourly generated power developed for Chennai is made using regression analysis (RA). It compares the characteristics of the model developed to that of the observed information. This model utilizes a single variable 'u' which is independent which possess a linear relationship with an output variable 'v'. Equation (1) denotes the linear regression model.

$$v = \lambda_o + \sum_{i=1}^{4} \lambda_i + \varepsilon \qquad (1)$$

Here, intercept λ_o and slope λ_i represents the constants (unknown) which are found using meteorological data and the random error is denotes as ε. The best fit in align with the straight line is termed as sum of squared residuals. The sample of data received from NREL for the 1st January is given in Table 2.

3. Regression Model for DC and AC Power

Based on the average hourly data for the site Chennai, Tamil Nadu, India, the regression model for generation of DC and AC power outputs are obtained.

Table 2. Sample of Meteorological data from NREL

Hour	Beam Irradiance (W/m^2)	Diffuse Irradiance (W/m^2)	Ambient Temperature (C)	Wind Speed (m/s)	DC Array Output (W)	AC System Output (W)
0	0	0	25.776	2	0	0
1	0	0	25.834	2	0	0
2	0	0	25.911	1	0	0
3	0	0	26.013	1	0	0
4	0	0	26.049	1	0	0
5	0	0	24.492	1	0	0
6	0	0	24.679	0	0	0
7	195	92	24.711	0	622.992	589.576
8	522	140	27.168	0	1550.004	1490.554
9	649	173	28.665	0	2144.566	2064.213
10	719	192	29.759	0	2507.496	2412.769
11	753	202	30.458	0	2671.626	2569.996
12	759	203	30.731	0	2647.127	2546.544
13	739	197	30.51	0	2446.45	2354.227
14	630	210	29.863	1	2087.052	2008.865
15	590	157	28.716	1	1538.536	1479.457
16	318	140	27.124	2	844.081	805.181
17	0	23	25.505	2	80.267	58.386
18	0	0	24.702	3	0	0
19	0	0	24.183	3	0	0
20	0	0	23.846	3	0	0
21	0	0	23.633	3	0	0
22	0	0	23.527	3	0	0
23	0	0	23.587	2	0	0

Two regression equations have been obtained from the prediction model. DC and AC power from PV panel are represented as equations (2) and equation (3) which have a linear relation with beam radiation, diffused radiation, temperature and wind speed.

The regression model is a statistical way of finding out the relationship among the control factors. Before generating a regression equation, it is necessary to check for residual plots to verify for linear regression.

$$PDC = x_0 + x_1 B + x_2 D - x_3 T + x_4 S \qquad (2)$$

$$PAC = y_0 + y_1 B + y_2 D - y_3 T + y_4 S \qquad (3)$$

Where, B is Beam Irradiance (W/m^2), D is Diffuse Irradiance (W/m^2), T is Ambient Temperature (C) and S is Wind Speed (m/s). PDC and PAC are the DC and AC power generated in Watt respectively. The co-efficient

of beam, diffuse irradiance, temperature and wind speed are provided in Table 3.

A graphical way of analysing residuals are most important to check for a 'good fit' regression model. A residual value is a measure of how much a regression line vertically misses a data point. Regression lines are the best fit of a set of data. The residual plots provide the information about the correlation among the trial run experiments and the predicted values. The residual plots of DC and AC are depicted in Figures 2 and 3 respectively. It is observed that the points are spread linearly straight as shown in Figures 2 (a) and 3 (a). It denotes a better correlation among the predicted and trial run (experiment) values. It is termed as a normal probability plot.

Table 3. Regression Co-efficient Terms

Co-efficient	DC Equation	Co-efficient	AC Equation
x0	352.3	y0	339.0
x1	2.4028	y1	2.3177
x2	3.3958	y2	3.2596
x3	15.194	y3	14.730
x4	7.44	y4	7.33

Residual values and predicted values are compared in Figures 2 (b) and 3 (b). Both seem to be closest to each other hence a very minimal difference is observed between them.

Figures 2 (c) and 3 (c) represent the histogram plots of DC and AC power. The clear statistics about the residuals are displayed in histogram graphs. The residuals versus order of trial runs are shown in Figures 2 (d) and 3 (d). The residual values are present in both positive and negative which indicates the presence of certain correlations. As a result of the complete analysis of DC and AC residual plots, the models are promising for adequacy.
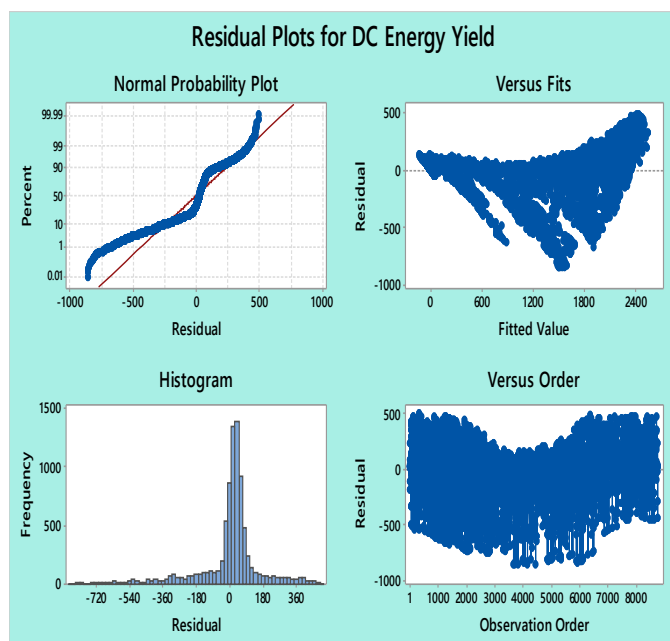

Figure 2. Residual Plots DC Power (a) Normal Probability Plot (b) Residual Vs Fitted Value (c) Histogram Plot (d) Residuals versus the Order of the Trial Runs
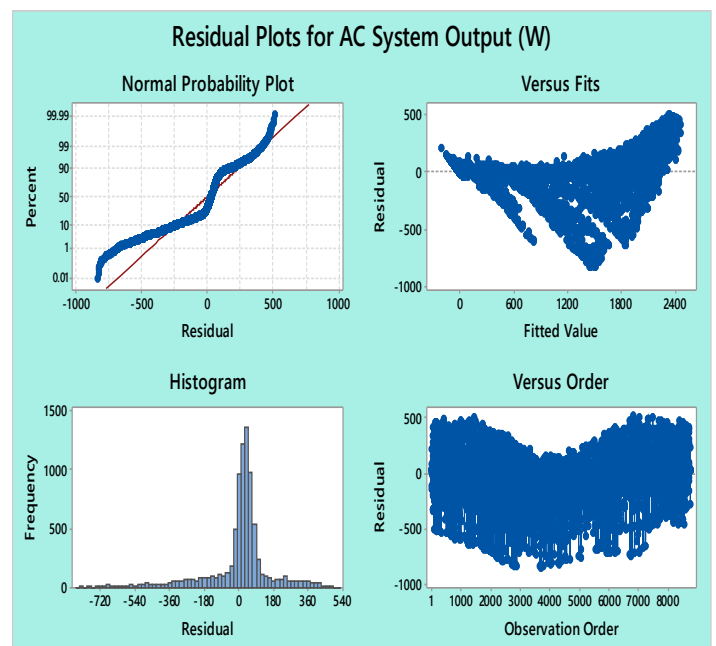

Figure 3. Residual Plots AC Power (a) Normal Probability Plot (b) Residual Vs Fitted Value (c) Histogram Plot (d) Residuals versus the Order of the Trial Runs

The created regression model possesses a better R^2 (Co-efficient Determination) value which denotes sufficient accuracy. The R^2 values obtained are 99.24% and 99% for DC and AC power respectively. The adjusted R^2 (Radj) values are 96.95% and 95.98% for DC and AC power respectively which implies that the developed regression model has high importance. Also, the predicted R^2 (Rpred) values obtained are 86.54% and 83.22% for DC and AC power respectively.

The effect of irradiance and temperature on the DC generated power is shown in Figure 4. In figure 4 (a), maximum array output is noticed at maximum beam irradiance and temperature median.

Further increase in temperature may lead to decrease in output. Figure 4 (b) reveals that high array output at median of diffused radiation and temperature. It is observed that for more output yield from the installed plant, the temperature must be in median rather than low or high.

The effect of irradiance and wind speed on the DC generated power is shown in Figure 5. The plot in figure 5 (a) reveals the interaction between the beam irradiance and wind speed effect on DC output. Here, for the case of high DC output, high beam irradiance and medium wind speed is required. Similarly, in the case of figure 5 (b), the maximum output is obtained at the median of the plot. Hence it is concluded that for the output yield to be high, maximum beam irradiance, medium diffused radiation, temperature and wind speed are found to be the optimum case.
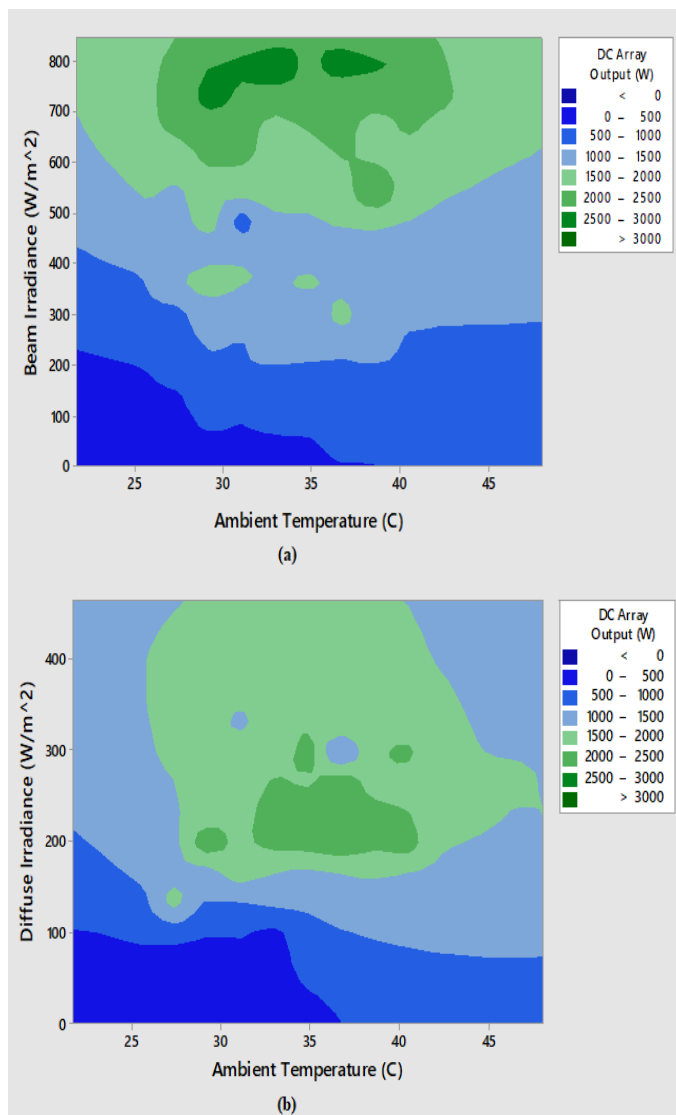
Figure 4. Effects on DC power output (a) influence of Beam Irradiance and Temperature, (b) influence of Diffuse Irradiance and Temperature.
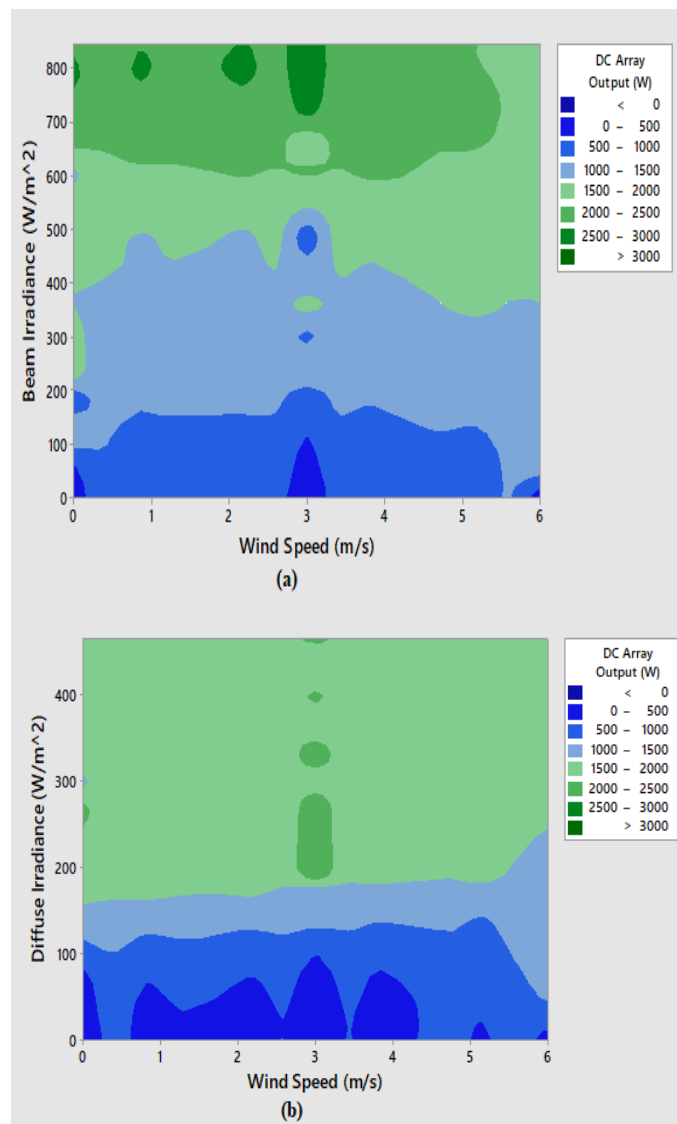

Figure 5. Effects on DC power output (a) influence of Beam Irradiance and Wind Speed, (b) influence of Diffuse Irradiance and Wind Speed.

The presented work is compared with the previous work done in the similar area and the comparative table is listed in the Table 4.

Table 4. Comparative Results

Model	R^2	RMSE
Proposed Work	99	2.293
Kahlil et al. [12]	95	6.34

Conclusion

The scarcity in fossil fuel and rise in environmental pollution, leads to the development of renewable based energy sources. A proper prediction of solar power from the PV panels are highly in need of today's scenario for a particular location. Hence this article can provide an insight idea to the researchers to predict the DC and AC power generated from the installed power plant in Chennai, Tamil Nadu, India. The prediction of DC and AC output power based on the beam radiation, diffused radiation, atmospheric temperature and wind speed has been carried out based on the meteorological data from NREL. Linear regression model based prediction equations are obtained using a software tool called Minitab 16.2.1. The developed

regression model owns a better R2 and Rpred value which denotes the model is best fit with high accuracy.

References

[1] Daut, I.; Irwanto, M.; Irwan, Y. M.; Gomesh, N.; Ahmad, N. S. Combination of Hargreaves method and linear regression as a new method to estimate solar radiation in Perlis, Northern Malaysia. Journal of Solar Energy. 2011; 85: 2871-2880.

[2] Aynur, U.; Figen, B. Assessment of wind power potential for turbine installation in coastal areas of Turkey. Renewable and Sustainable Energy Reviews. 2010; 14: 1901-1912.

[3] C. Furlan, A. P. de Oliveira, J. Soares, G. Codato, and J. F. Escobedo. The role of clouds in improving the regression model for hourly values of diffuse solar radiation. Applied Energy. 2012; 92: 240–254.

[4] Kiesecker, J.; Baruch-Mordo, S.; Heiner, M.; Negandhi, D.; Oakleaf, J.R.; Kennedy, C.M.; Chauhan, P. Renewable Energy and Land Use in India: A Vision to Facilitate Sustainable Development. Sustainability. 2020; 12(1): 281 – 289.

[5] Tamil Nadu state of Environment and related issues http://tnenvis.nic.in/ (Accessed on October 2020).

[6] Brancucci, C.; Martinez-Anido, B.; Botor, A. R.; Florita, C.; Draxl, S.; Lu, H. F. Hamann; and B. M. Hodge. The value of day-ahead solar power forecasting improvement. Solar Energy. 2016; 129: 192 – 203.

[7] Al-Dahidi, S.; Ayadi, O.; Adeeb, J.; Alrbai, M.; and Qawasmeh, R. B. Extreme Learning Machines for Solar Photovoltaic Power Predictions. Energies. 2018; 11(10): 2725 - 2736.

[8] Das, U. K.; Tey, K. S.; Seyedmahmoudian, M.; Mekhilef, S.; Idris, M. Y. I.; Van Deventer, W., Horan, B., and Stojcevski, A., Forecasting of photovoltaic power generation and model optimization: A review. Renew. Sustain. Energy Rev. 2018; 81: 912–928.

[9] Izgi, E.; Oztopal, A.; Yerli, B.; Kaymak, M. K.; and Şahin, A. D. Short-mid-term solar power prediction by using artificial neural networks. Solar Energy. 2012; 86(2): 725–733.

[10] Al-Dahidi, S.; Ayadi, O.; Alrbai M.; and Adeeb, J. Ensemble Approach of Optimized Artificial Neural Networks for Solar Photovoltaic Power Prediction. IEEE Access. 2019; 7(81): 741-758.

[11] Ministry of New and Renewable Energy, https://mnre.gov.in/solar/current-status/ (Accessed on December 2020)

[12] Khalil, S. A.; and Shaffie, A. M. A comparative study of total, direct and diffuse solar irradiance by using different models on horizontal and inclined surfaces for Cairo, Egypt. Renewable and Sustainable Energy Reviews. 2013; 27: 853–863.

[13] Sabzpooshani M. and Mohammadi, K. Establishing new empirical models for predicting monthly mean horizontal diffuse solar radiation in city of Isfahan, Iran. Energy. 2014; 69: 571–577.

[14] Zhao, N.; Zeng, X.; and Han, S. Solar radiation estimation using sunshine hour and air pollution index in China. Energy Conversion and Management. 2013; 76: 846–851.

[15] Rajendran, P.; and Smith, H. Implications of longitude and latitude on the size of solar-powered UAV. Energy Conversion and Management. 2015; 98: 107–114.