International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No. (03), January 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

PV MODULE PERFORMANCE USING REGRESSION ANALYSIS

E. Fernandez[1],Sandhya Prajapati[2]

[1,2]*Electrical Engineering Department, IIT Roorkee, Roorkee, Uttarakhand, (India)*

ABSTRACT

PV module performance is analyzed in this paper using two types of regression models, namely, linear regression model and non linear regression model with logarithmic linearization. Data used for the models is obtained from manufacturer's specifications of different PV modules available commercially. The results show that the log model fares better in terms of accuracy

Keywords-PV performance, Regression, Comparative analysis, PV module, Linear and Non linear models

I.INTRODUCTION

The photovoltaic (PV) module is a reliable renewable energy based technology which directly converts solar radiation into electrical power. Since 2009, the PV installed capacity is increasing globally with more than 40% growth increase all over the world [1]. Large scale PV power systems have been commercialized in numerous countries due to their substantial long term benefits and promotion of sustainable "*green energy*". Financial initiatives are provided by most governments to promote this sustainable form of *green* energy. However, due to the high investment cost on PV modules, it becomes necessary to size systems satisfactorily by making use of optimal utilization of the available solar energy.

Different methods have been adopted to classify and estimate PV power generation. Some may directly provide energy values, while others compute the power generated by the system, thereafter integrating it to the given time to obtain the energy generated during the period considered. Some other methods use parameters such as current curves and PV generating voltage, while yet others use data directly from the electrical power generated [2,3] Certain other methods employ electrical and climatic data on linear regression models or artificial neural networks (ANN), data from the literature data and/or manufacturers together with climatologically related databases. However, these data may have significant variations in mean values, affecting the quality of the results [4, 5].In general, the use of PV performance evaluation methods is complex because they depend on multiple variables like intrinsic characteristics of the generators and weather intermittency and the climatic peculiarities of each region.

One of the convenient ways to model PV performance is by the use of Linear Regression models. This is a statistical approach that obtains the best coefficients required to model the system with a minimum cumulative least square error. The accuracy obtainable for the prediction will depend on the value of the parameter R^2. This value lies between 0 and 1, and the nearer is this value to 1, the better is the predicted accuracy of the regression model designed for the PV module. In the literature, some researchers have tried to apply linear regression to model PV systems. Some of these studies are reported in [6-10].

However, the linear regression model is best suited to model phenomenon that is essentially linear in functional form. In the case of the PV system, the governing performance equations are non linear. Hence, the linear regression models needs to be converted to a linearized form using logarithmic transformation. This will result in a non linearized model that is mathematically represented in a linearized form for performance evaluation.

The aim of this paper is to examine and compare the accuracy of two model- a linear regression model and a non-linear regression model for the PV module performance prediction. The non linear model is mathematically represented by a logarithmic linearized equivalent model.

II. THE REGRESSION MODEL

Linear Regression is an approach for modeling the relationship between a scalar dependent variable *y* and one or more explanatory variables (or independent variables) denoted *X* in a linear frame. If one explanatory variable is involved the regression is termed as *simple linear regression*. For more than one explanatory variable, the process is called *multiple linear regression*. A multiple regression model is given by the general expression:

$$y_i = \beta_0 1 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^\top \beta + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where: y_i denotes the dependent variable and x_{in} th independent variables. ε_i denotes the error term involved [11] . Often these equations are grouped and collectively represented in vector form by the expression:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

Where:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$X = \begin{pmatrix} \mathbf{x}_1^\top \\ \mathbf{x}_2^\top \\ \vdots \\ \mathbf{x}_n^\top \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Ordinary least squares (OLS) is the most commonly used estimator being conceptually simple and computationally straightforward. The OLS method minimizes the sum of squared residuals, and leads to a closed-form expression for the estimated value of the unknown parameter β:

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \left(\sum \mathbf{x}_i \mathbf{x}_i^\top\right)^{-1} \left(\sum \mathbf{x}_i y_i\right).$$

The estimator is unbiased and consistent if the errors have finite variance and are uncorrelated with the regressors, i.e.

$$\mathrm{E}[\,\mathbf{x}_i \varepsilon_i\,] = 0.$$

The degree of "fit" of the model is evaluated in terms of the R^2 statistic value that lies between 0-1. A value closer to unity indicates a good fit to the input data.

When the variable in the regression equation are related in a non-linear fashion (as is the case for the PV module), the linearity condition of the regression can be maintained if logarithmic transformations of the variables are involved. This gives rise to use of the Log-Log models in which both the dependent variable as well as the independent variables are in the logarithmic form[12]. Such a model is expressed as:

$$Ln\,(y) = B_0 + B_1 * Ln(x_1) + B_2 * Ln(x_2) + B_3 * Ln(x_3) + \ldots\ldots$$

III. THE DEVELOPED REGRESSION MODELS

Four control variable were identified namely, Open Circuit voltage of the module (Voc), The Short Circuit current of the module (Isc), The level of Solar radiation (Insolation level)and the Temperature of the module (Temperature)

The linear regression model was structured both as a linear form as well as a Log-Log form. These forms are given below as:

Linear Form

$$Pmax = \beta 0 + \beta 1 * (Voc) + \beta 2 * (Isc) + \beta 3 * (insolation\ level) + \beta 4 * (Temperature)$$

Logarithmic Form

$$Ln(P\,max) = Ln(\beta 0) + \beta 1 * Ln(Voc) + \beta 2 * Ln(Isc) + \beta 3 * Ln(Insolation\ level) + \beta 4 * Ln(Temperature)$$

The data for developing the models was obtained from a created database involving different PV modules specifications as available from various manufacturers. Additionally some research papers providing output results of experimental investigations have also been consulted. The coefficients $\beta 0$, $\beta 1$, $\beta 2$, etc will be different

for the two models The values so obtained for the two models using MS EXCEL are shown in Tables 1,2.

Table 1 Results of the Regression Model (Linear Form)

S.No	Feature	Value		
1.	R^2	0.92448		
2.	**Standard Error**	18.9037		
3.	**F Statistic**	91.8066		
4.	***Coefficients of regression terms***	***Coefficient Value***	***Standard error***	***t- statistic***
	βo	-194.114	31.0356	-6.2545
	$\beta 1$	5.027044	0.549786	9.143629
	$\beta 2$	32.08994	2.529024	12.68867
	$\beta 3$	-0.01815	0.024088	-0.7536
	$\beta 4$	-0.17474	0.263908	-0.66214

Table 2 Results of the Regression Model (Logarithmic Form)

S.No	Feature	Value		
1.	R^2	0.9804		
2.	**Standard Error**	0.0907		
3.	**F Statistic**	185.9828		
4.	***Coefficients of regression terms***	***Coefficient Value***	***Standard error***	***t- statistic***
	βo	0.8636	0.6603	1.3079
	$\beta 1$	0.9357	0.1051	8.8976
	$\beta 2$	0.9996	0.0817	12.2299
	$\beta 3$	-0.1000	0.0971	-1.0299
	$\beta 4$	-0.0634	0.0476	-1.3318

IV.RESULTS

Taking a random sample of 20 sets of readings from the developed database, the mean percentage error in estimation of the two regression models (linear form, and Logarithmic form) were analysed.

Table 3 shows the results of this analysis.

IJARSE
ISSN: 2319-8354

Table 3: Results of relative comparison of mean errors of the two models

S.No	Pmax (actual)	Linear model Value	% absolute error	Log model value	% absolute error
1	280.858	274.377653	2.3073393	281.11325	0.090885
2	310.709	296.0060456	4.7320658	304.46968	2.008089
3	205.117	206.1315465	0.4946184	207.22621	1.028297
4	216.062	213.4605103	1.2040477	214.06405	0.92471
5	178.565	178.3969781	0.0940956	181.46423	1.62363
6	214.361	226.8354342	5.8193580	190.50492	2.357244
7	209.25	222.1610374	6.1701492	195.01476	2.412597
8	234.885	239.0602171	1.7774280	133.27703	1.071086
9	196.016	186.3228436	4.9451327	207.92451	4.419616
10	305.321	294.014944	3.7031325	220.67426	2.945156
11	158.064	160.0641087	1.2653790	165.02848	4.40612
12	230.124	236.2882222	2.6786524	231.51774	0.60565
13	239.695	247.2267076	3.1419895	247.43528	3.229009
14	220.129	230.554304	4.7359975	230.74124	4.82092
15	240.787	234.5082396	2.6075994	230.07169	4.450117
16	235.64	231.1711141	1.8964886	227.21692	3.574552
17	250.17	254.997841	1.9298241	254.38712	1.685704
18	172.627	189.8969002	10.004170	191.58172	10.98016
19	205.2	214.9084819	4.7312289	212.16318	3.393364
20	121.5	114.454691	5.7986082	132.37063	8.947021

	Mean %error Linear form	3.501865	Mean %error Log form	3.248697

V. DISCUSSION OF RESULTS

The results appear to indicate that there is not much variation in the accuracy of the linear and the logarithmic forms of linear regression as far as analysis for PV module performance is concerned. The probable reason for this observation is that the characteristics (I-V & P-V) for the PV module are basically linear in form except in the region near the knee of the curve. Thus, the non-linearity is extended to a relatively small region of the entire performance curve. Thus both models will show almost equal levels of accuracy. In this work, it appears that the logarithmic model is slightly better due to lesser error.

VI. CONCLUSIONS

An attempt was made in this paper to model the performance of a photovoltaic (PV) module using regression analysis. A database was prepared from a compilation of various manufacturers' data in relation to the PV modules manufactured by them. The input data involved four variables namely, open circuit voltage (Voc), short circuit current (Isc), solar insolation and temperature of operation of the module .The dependent output was the maximum power of the module (P max). Two types of regression models were tried i.e. linear form and logarithmic form.

REFERENCES

[1] Amit Kumar Yadav,, S.S. Chandel , "Identification of relevant input variables for prediction of 1-minute timestep photovoltaic module power using Artificial Neural Network and Multiple Linear Regression Models", *Renewable and Sustainable Energy Reviews* 77 (2017) 955–969

[2] N. Aste, C. Del Pero, F. Leonforte, M. Manfren, "A simplified model for the estimation of energy production of PV systems", *Energy* 59 (2013) pp.503-512.

[3] C.Rus-Casas, J.D. Aguilar, P. Rodrigo, F. Almonacid, P.J. P_erez-Higueras, "Classification of methods for annual energy harvesting calculations of photovoltaic generators", *Energy Conversion and. Management* 78 (2014) pp.527-536.

[4] S. Labed, E. Lorenzo, "The impact of solar radiation variability and data discrepancies on the design of PV systems*", Renewable. Energy* 29 (7) (2004), pp.1007-1022.

[5] M. Ernst, A. Thomson, I. Haedrich, A. Blakers, "Comparison of ground-based and satellite-based irradiance data for photovoltaic yield estimation*", Energy Procedia* , 92 (2016) pp.546-553.

[6] S. H. Oudjana, A. Hellal, and I. Hadj Mahammed, "Power Forecasting of Photovoltaic Generation" ,International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering Vol:7, No:6, 2013

[7] Salim Moslehi, T. Agami Reddy, Srinivas Katipamula "Evaluation of data-driven models for predicting solar photovoltaics power output", Energy Vol.142, pp 1057-1065, 2018(Available online 10 September 2017)

[8] A. M. Muzathik, " Photovoltaic Modules Operating Temperature Estimation Using a Simple Correlation", International Journal of Energy Engineering,Vol. 4 (4,) pp. 151-158, Aug. 2014.

[9] T. Bhattacharya ,A. K. Chakraborty , and K. Pal, "Statistical Analysis of the Performance of Solar Photovoltaic Module with the Influence of Different Meteorological Parameters in Tripura, India", International Journal of Engineering Research Volume No.4(3), pp : 137 – 140, March 2015.

[10] Abhineet Samadhiya and Ruchi Pandey, "Analysis of PV Panels under Various Weather Conditions", International Journal of Emerging Research in Management &Technology, Volume 5(2), February 2016.

[11] https://en.wikipedia.org/wiki/Linear_regression#Simple_and_multiple_linear_regression

[12] Damoder N Gujarati,"Basic Econometrics", Fourth Edition ,Mc Graw-Hill Companies , 2004