Photovoltaic forecasting model in Thailand: case study solar farm at Nakhon Ratchasima province

Withaya Yongchareon[1,*], Sumlee Thongthew[2], and Pensiri Juncharoen[3]

[1]Faculty of Engineering. Chulalongkorn University

[2]Faculty of Education. Chulalongkorn University

[3]Graduate School, Chulalongkorn University

* Corresponding Author: Withaya Yongchareon: yongchareon@gmail.com

Abstract

The photovoltaic module performance was normally tested at the standard conditions. The use of the actual field performance has not been investigated. One purpose of this study was to find the predicted equation for forecasting the energy produced from photovoltaic modules. Data for this study were collected using the data from the solar farm in Nakhon Ratchasima province: solar irradiation, ambient temperature, module temperature and wind velocity. The daily composite data were 1,461 from 1st January 2015 – 31st December 2018. The five predicted equations, including both the main effect and the interaction effect, were statistically analysed by regression method using the Minitab software program for multiple linear regression. The requiring dependent variable (y) was the energy produced from the photovoltaic module. The four independent variables were solar irradiation (x_1), ambient temperature (x_2), module temperature (x_3) and wind velocity (x_4). For prediction preciscion, all the predicted equations were validated with the new gathered data from 1st January 2019 – 30th June 2019. The analysis results showed, the forecasting model together with the appropriate predicted equation with the highest coefficient of determination (R^2 of 0.9873) and the standard deviation of prediction (S = 2.67%). This predicted equation, $\hat{y}_2 = 5258 + 5310.0x_1 - 100.31x_2 + 66.2x_4$, consisted of the three external independent variables; solar irradiation, ambient temperature, and wind velocity. The independent variable sensitivities were also determined. The solar irradiation was the most sensitive to the predicted equation. Moreover, this predicted equation would be suggested to utilize as the estimator for the energy produced from the new photovoltaic power plant to be installed at the north-eastern part of Thailand.

Keywords: Photovoltaic, Forecasting Model, Regression Method, Solar Farm

1. Introduction

There is evidence that a solar panel play a crucial in generating clean, emission free electricity. One of the greatest challenges is it produces only direct current electricity (DC) and is limited by appliances use. Solar photovoltaic systems (solar PV systems) are often made of solar PV panels (modules) and inverter (changing DC to AC). Solar PV panels are of interest because they are mainly made of solar photovoltaic cells. As much module performance is generally rated under standard test conditions (STC): irradiance of 1,000 W/m^2, the solar spectrum of AM 1.5 and module temperature at 25°C.

So far, however, there has been little discussion about, the actual field performance. Evidence suggests that the actual voltage and current output of the module changes according to lighting, temperature, and load conditions. Therefore, there is never one specific voltage, current, or power at which the module operates. Performance varies depending on the time of the day, amount of solar insolation, direction and tilt of modules, cloud cover, shading, state of charge, temperature, geographic location, and day of the year. Concerning the photovoltaic power plant, the electrical power output, according to the production license, must be guaranteed. Since the lifetime of the solar panels is generally last at approximately 25 years, the electrical power output produced must be forecasted before actual use to make sure it meets the guaranteed power output. Preliminary work on multiple regression analysis technique was under taken by Pichet Vongkiem [1]. In an analysis of multiple regression analysis, he showed that it can forecast the electrical energy needs for the PEA (Provincial Electricity Authority). This paper purposes a new methodology of the predicted equation for forecasting the energy produced from photovoltaic modules using the past actual data from the solar farm at Nakhon Ratchasima province. As can be seen from Table 1 below, electrical solar cell module performance at standard conditions include maximum power (P_{MAX}), open circuit voltage (V_{OC}), short circuit current (I_{SC}), maximum power voltage (V_{MPP}), maximum power current (I_{MPP}), and module efficiency (%).

Table 1. Solar Cell Module Performance

Electrical Performance	Unit	Value
Irradiation	W/m2	1000
Module Temperature	^{o}C	25
Maximum Power	W	210
Maximum Power Voltage	V	26.6
Maximum Power Current	A	7.9
Open Circuit Voltage	V	33.2
Short Circuit	A	8.58
Module Efficiency	%	12.1

Table 1. also shows that the module temperature increased when electrical power and efficiency decreased. It was found that the module temperature depended on the balance of heat gain and heat loss from the module, in the environment.

2. Regression method [4]

The regression method has been proposed to analyse the forecasting model, and the predicted equation containing various parameters. The regression method, through the statistic test can eliminate the unnecessary parameter in the predicted equation and it has less contribution to the dependent variable, which is the solar power output in this study.

To demonstrate regression method, the simple (1st order) linear regression was selected as an example. A set of observed data consisted of independent variable y and independent variable x. As shown in the following procedures.

1) The assumed model was expressed as in the equation (1)

$$y_i = a + bx_i + \varepsilon_i \tag{1}$$

where y_i and x_i were the observed variables
a and b were constant parameter
ε_i was the random error variable of y and normally distributed with the mean of 0 and variance of 1 and also independent on the level of x

Then, the predicted equation could be expressed as in the equation (2)

$$\hat{y}_i = a + bx_i \qquad (2)$$

2) The least square of error was applied as in the equation (3)

$$\sum \varepsilon_i^2 = \sum (y_i - (a + bx_i))^2 = \text{mininum} \qquad (3)$$

To satisfy equation (3) the partial derivative of parameters should be set to zero as in the equation (4)

$$\frac{\partial \sum \varepsilon_i^2}{\partial a} = 0 \text{ and } \frac{\partial \sum \varepsilon_i^2}{\partial b} = 0 \qquad (4)$$

From Equation (4) the parameter a and b could be mathematically solved as in equation (5)

$$a = \bar{y} - b\bar{x} \quad \text{and} \quad b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \qquad (5)$$

where $\bar{y}$ was the mean value of y_i and $\bar{x}$ was the mean value of x_i

3) Finding the goodness of fit by using the coefficient of determination R^2 as in the equation (6)

$$R^2 = \frac{SS_{REG}}{SS_T} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \qquad (6)$$

where SS_{REG} was sum square due to regression and SS_T was the total sum square

$R^2=1$ was the perfect fit, all data points of y were lined on the regression line. However, in the actual situation there were errors due to measurement uncertainties and the assumed model was not the true model. Thus $R^2<1$

4) Testing the hypothesis by using statisticT for individual parameter and statistics F for all parameters. If the individual parameter was significant, then it was in the predicted equation.

3. Solar farm data [2]

The recorded field data used in this study were solar irradiation, ambient temperature, module temperature wind velocity and energy production. The daily composite data were 1,461 from 1 January 2015 - 31 December 2018 for construction of models. The data were 183 from 1 January 2019 - 30 June 2019 for validation of the model. The example of recorded data was given in Table 2. The statistic data of 1461 was shown in Table 3.

Table 2. The example of recorded data on the solar farm

Date	Irr (kWh/m2)	Amb temp. (◦C)	Module temp. (◦C)	Wind velocity (m/sec)	Energy Production (kWh)
1/1/2015	6.397	25.130	33.283	2.092	38,400
2/1/2015	6.308	25.240	33.797	1.833	37,600
3/1/2015	6.081	26.281	35.484	1.097	35,920
4/1/2015	6.405	30.114	41.766	0.541	36,880
5/1/2015	6.104	32.507	43.484	0.463	34,800
6/1/2015	5.885	33.141	43.285	0.636	33,440
7/1/2015	4.454	32.072	39.764	0.918	25,360
8/1/2015	1.192	25.660	26.341	1.282	7,280
9/1/2015	4.675	28.787	36.584	1.412	27,360
10/1/2015	5.396	29.456	38.765	1.326	31,520

Table 3. The statistic data of solar farm

List	Irr (kWh/m2)	Amb temp. (◦C)	Module temp. (◦C)	Wind velocity (m/sec)	Energy Production (kWh)
Mean	5.484	32.9	42.3	1.1	31156
Minimum	0.620	2.4	16.9	0.0	3680
Maximum	7.517	41.9	54.3	3.1	42800
Range	6.896	39.5	37.4	3.0	39120
Standard Deviation	1.200	3.7	5.4	0.5	6464

4. Selection of Models

The linear polynomial model method is particulary useful in studying since most of the engineering data could be fitted with a popular polynomial equation, and the required model was a simple approximation model with acceptable precision**.** However, the higher-order term such as the radiation loss term which is the function of (temperature)^4 was not included in the model because the value of temperature variables was less than 100^{o}C.

All recorded data of solar farm were used. Energy Production was selected as the dependent variable (y), while solar irradiation (x_1), ambient temperature (x_2), module temperature (x_3), and wind velocity (x_4) were independent variable. The forecasting approximation model could be many models that represent the field data set. First, the used full model consisted of all field variables and only the main effect of x_1, x_2, x_3, and x_4 (see Equation (7)). The predicted equation was also given in Equation (8). The model included an error to represent the data points.

$$y_i = a + bx_{1i} + cx_{2i} + dx_{3i} + ex_{4i} + \varepsilon_i \quad (7)$$

$$\hat{y}_i = a + bx_{1i} + cx_{2i} + dx_{3i} + ex_{4i} \quad (8)$$

Where ε_i was the random error in y, it is normally distributed and independ to the level of x. It can be seen from the data in Table 4., the regression analysis results obtained from the Minitab software program.

Table 4. Analysis of Variance (ANOVA)

Source	Degree of freedom	Sum Square	Mean Square	F-Value	P_Value
Regression	4	60255940916	15063985229	29006.63	0.000
Solar Irradiation	1	13607620571	13607620571	26202.31	0.000
Amb.Temp.	1	1541163	1541163	2.97	0.085
Module Temp.	1	20156982	20156982	38.81	0.000
Wind velocity	1	9096653	9096653	17.52	0.000
Error	1455	755623775	519329		
Total (corrected)	1459	61011564690			

The predicted equation was also provided in Equation (9)

$$\hat{y}_1 = 4594 + 5494.6x_1 + 40.2x_2 - 118.0x_3 + 90.8x_4 \quad (9)$$

with coefficient of determination R^2 = 0.9876 and standard deviation S = 720.6 kWh. In this current study the predicted equation represented the field data with the precision of 98.76%. Table 4 illustrates the ambient temperature (x_2). From this data, it was not significant because the probability of F (P_ value) in the last column, was greater than the significant at the p = 0.05 level . It can thus be suggested that the ambient temperature would not existed in the predicted equation. Secondly, there were many approximation models and predicted equations which consisted of the linear combinations of x-variables including only the main effect, for example $\hat{y}_2$ and $\hat{y}_3$, or including both the main effect and the interaction effect, for example $\hat{y}_4$ and $\hat{y}_5$. The predicted equations fitted to the observed field data (see Table 5.). Table 5. presents the results of the regression analysis of some predicted equations. It reviewed that these four selected predicted equations was a high value of R^2 showing more than 0.97. Further analysis showed that these approximation predicted equations could represent the observed solar farm data set. It also could explain the variation of data, showing more than 97 %. However, the equation $\hat{y}_2$ had the high value of R^2 of 0.9873 and contained only the three external independent variables; solar irradiation, ambient temperature, and wind velocity. While x_3, module temperature, depends on solar irradiation, ambient temperature, and wind velocity. Therefore, this predicted equation $\hat{y}_2$ was appropriately predicted equation representing the forecasting equation.

Table 5. Analysis Results of Regression

Predicted Equation	R^2
$\hat{y}_2 = 5258 + 5310.0x_1 - 100.31x_2 + 66.2x_4$	0.9873
$\hat{y}_3 = 4849 + 5448.1x_1 - 86.51x_3 + 80.7x_4$	0.9876
$\hat{y}_4 = 2471.7 + 5165.7x_1 + 0.2382x_2x_3x_4$	0.9847
$\hat{y}_5 = 2241.5 + 5175.4x_1 + 15.07x_2x_4$	0.9786

The result of the residual ($y_i - \hat{y}_i$) composed of the lack of fit and the random noise ε_i . The state of lack of fit could be observed from the residual plots. If the distribution of the residual was nearly the normal distribution then it shows a small state of lack of fit. The residual plot of the predicted equation $\hat{y}_2$ was nearly normal distribution because of its high value of R^2 (see Figure 1.)

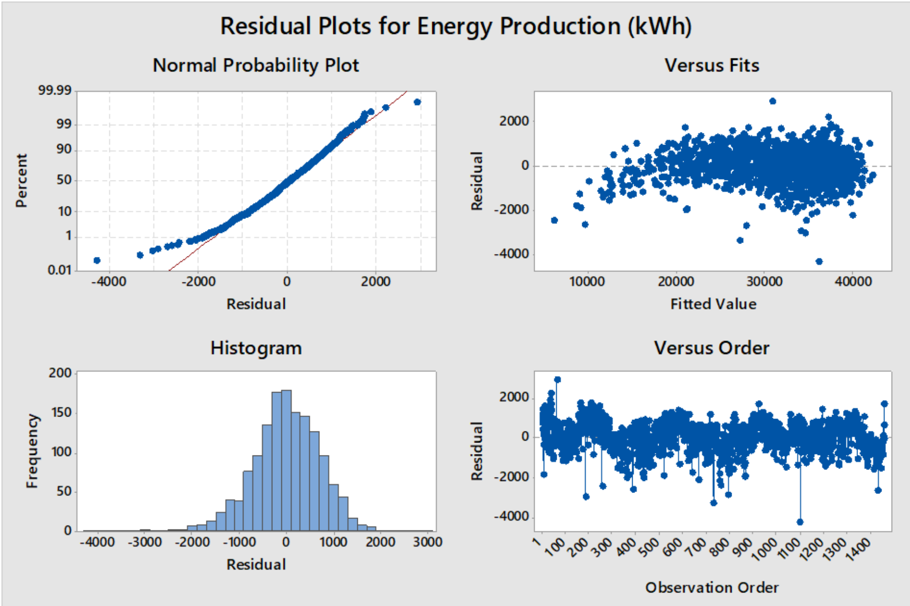

Figure 1. The residual plots of the predicted equation $\hat{y}_2$

In general, therefore, it seems that the predicted equation could be used to represent the observed solar farm data set. It is possible to use this predicted equation as the estimator for the energy produced from the exiting photovoltaic power plant or the new one to be installed in the future. This predicted equation need to be validated with the new data set from 1 January 2019 - 30 June 2019. The predicted values from the predicted equation were used to compare with the observed data set. As shown in Figure 2., the straight line passed to the origin. The equation was y = 1.0063 x with the R^2 of 0.9749. These findings suggest that the predicted equation $\hat{y}_2$ could be used as the etimator for energy production.

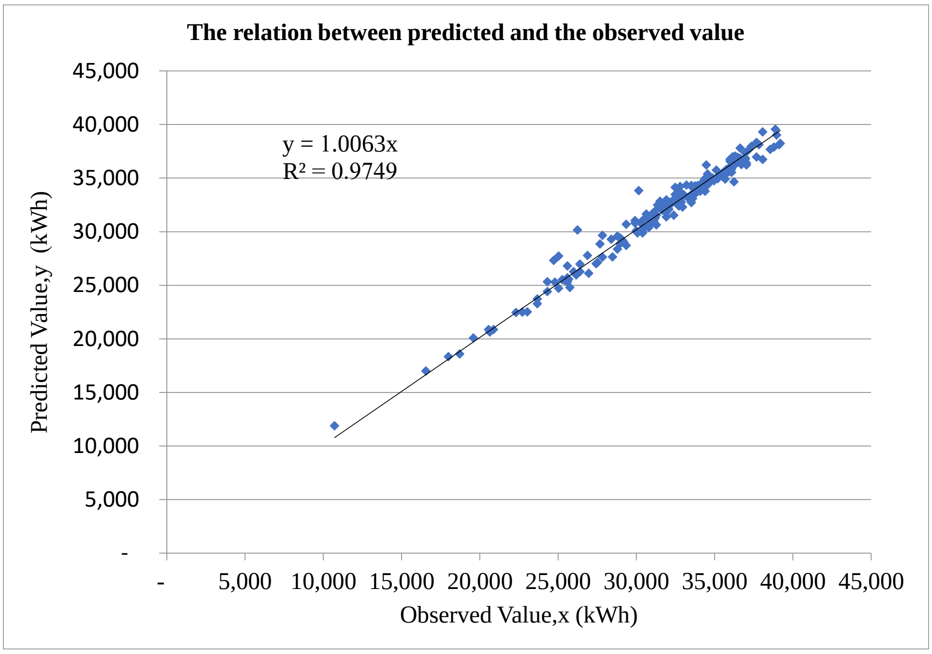

Figure 2. The relation between the predicted and the actual value of solar farm

5. Sensitivity Analysis

The energy production $\hat{y}_2$ depended on the solar irradiation x_1, ambient temperature x_2, and wind velocity x_4. To find how much each variable had an influence on the energy production, the variable sensitivity had been analyzed by changing each variable by 10 % while keeping other variables constant and obtained the change of the energy production. The sensitivity results were given in Table 6.

Table 5 provides the solar irradiation and it was the most sensitive variable. Whilst the wind velocity was the least sensitive variable. Also, the ambient temperature was a negative sensitivity. When the ambient temperature increased the energy production decreased because of the increased in the module temperature. This finding is consistent with that of Keattisak Phunjumpa [3]

Table 6. Analysis results of Sensitivity of the predicted Equation

	Daily Actual Average				
List	**Irradiation kWh/m2**	**Amb temp. ◦C**	**Wind velocity m/s**	**Energy Production kWh**	**% Diff. $(y_{+10\%} - \hat{y})/\hat{y}$**
	x_1	x_2	x_4	$\hat{y}_2$	
Actual	5.484	32.9	1.1	31151	
+10% by variable	6.032	32.9	1.1	34063	9.35%
	5.484	36.2	1.1	30821	-1.06%
	5.484	32.9	1.2	31158	0.02%
Note:	$\hat{y}_2 = 5258 + 5310x_1 - 100.31x_2 + 66.2x_4$				

6. Conclusion

This study has identify the data of solar farm that comprised of solar irradiation, ambient temperature, module temperature and wind velocity. This study had found that generally daily composite data were 1,461 from 1st January 2015 – 31st December 2018 for construction of the model. The data were 183 from 1 January 2019 - 30 June 2019 for validation of the model.The method used in this current study was a regression and it is chosen to forecast the model along with the appropriate predicted equation with the highest coefficient of determination (R^2 = 0.9873) and the standard deviation of prediction (S = 2.67%). This predicted equation contained only the three external independent variables; solar irradiation, ambient temperature, and wind velocity. The independent variable sensitivities were also determined. The solar irradiation was the most sensitive to the predicted equation. The finding reported here shed new light on the estimator for the energy produced from the new photovoltaic power plant that will be installed in the north-east of Thailand.

Acknowledgments

The researcher was grateful to obtain the data of the solar farm from the solar power company.

References

[1] Vongkiem P and Rerkpreedapong D 2015 *J. Kasetsart Engineering* **91** 31-40
[2] Juncharoen P and Yongchareon W 2019 *Proc. Conf. on Solar Energy Forecasting Model in Thailand (Thai Version)* (Phitsanulok: Thailand) 1-6
[3] Phunjumpa K Kongnok K and Plangklang B 2017 *Proc. Conf. on Study on the Effect of Ambient Temperature to Electrical Energy Production from Solar Cells (Thai Version)* (Chanthaburi: Thailand) p 217
[4] Draper N.R., Smitch.H, 1966 *Applied Regression Analysis* (London: John Wiley & Sons) Chapter1 pp 1-33