Applied Mathematics and Nonlinear Sciences

<https://www.sciendo.com>

Visualized calculation of regional power grid power data based on multiple linear regression equation

Kun Chen^{1†}, Xinhua Wang¹, Daniyal M. Alghazzawi², Yanfeng Wang¹

¹Grid Planning & Research Center, Guangdong Power Grid Corporation, CSG, Guangzhou, Guangdong, 510080, China ²Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Submission Info

Communicated by Juan Luis García Guirao

Received June 16th 2021

Accepted September 24th 2021

Available online December 15th 2021

Abstract

The power generation of distributed power generation systems in regional power grids is mainly affected by external environmental factors, and there are volatility and intermittent power generation issues. This paper establishes a relatively simple multiple linear regression prediction model to predict grid-connected photovoltaic power generation systems. The article uses a hybrid modelling method to filter more important weather forecast input variables through a step-by-step linear regression method to realize the visual calculation of power data. The research results show that the multiple linear regression equation methods proposed in the article have pervasive application value in the actual application of solar photovoltaic power consumption data.

Keywords: regional power grid, multiple linear regression, power data visualization, photovoltaic power generation data.

AMS 2010 codes: 62J05

1 Introduction

The rapid development of society and the economy has led to a sharp increase in energy consumption, increasingly depleted fossil energy sources, and increased environmental pollution. This poses a severe threat to the ecological balance of the earth and the survival of humankind. Large-scale photovoltaic power plants are an effective way to meet human demand for energy. Because photovoltaic power generation is affected by solar radiation intensity, battery component temperature, weather, cloud cover, and some random factors, the system operation process is an unbalanced random process. Its power generation and power output are highly random, fluctuating, and uncontrollable. This is particularly prominent when the weather changes suddenly.

[†]Corresponding author.

Email address: 474906215@qq.com

This kind of power generation inevitably brings a series of problems to the safety and management of the power grid after it is connected to the power grid [1]. Therefore, it is imperative to predict the output of the photovoltaic system accurately. At present, there are several types of prediction models for the power generation of photovoltaic power generation systems, such as neural network models, radial basis function models, and multi-layer perception models. In addition, the photovoltaic system is a typical grey system with partly clear and partly unknown information [2]. Grey theory forecasting models have a wide range of applications in power planning and load forecasting. Because these models are relatively complex, this paper establishes a relatively simple multiple linear regression prediction model to predict the power generation of grid-connected photovoltaic power generation systems.

2 Theoretical calculation of power generation of solar cell modules

Photovoltaic power generation uses the photovoltaic effect of the P–N junction to convert light energy into electrical energy. The equivalent circuit of its components and system under ideal conditions is shown in Figure 1. The maximum output power is represented as follows:

$$P_m = U_m I_m = U_{oc} I_{sc} FF \quad (1)$$

Here, U_{oc} , I_{sc} , FF are the open-circuit voltage, short-circuit current, and fill factor of the battery. Silicon solar cells are $n = 1$, $R_s \rightarrow 0$, $R_{sh} \rightarrow \infty$ under ideal conditions [3]. Therefore, the open-circuit voltage, short-circuit current, and fill factor can be expressed as follows:

$$\begin{cases} U_{oc} = \frac{n_s n k T}{q} \ln \left(\frac{I_L}{I_D} + 1 \right) \\ I = I_L \approx I_{sc} \\ FF = \frac{u_{oc} - \ln(u_{oc} + 0.72)}{u_{oc} + 1} \left(1 - \frac{R_s I_{sc}}{U_{oc}} \right) \end{cases} \quad (2)$$

Here, n_s , n are the number of batteries in series in the module and the ideality factor of the diode, respectively; B is the normalized open-circuit voltage; $u_{oc} = qU_{oc}/nkT$; are the load current, photogenerated current, and diode response saturation current.

Fig. 1 Equivalent circuit diagram of an ideal DC model of a photovoltaic module. DC, direct current.

The performance of monolithic silicon solar cell modules is mainly affected by radiation intensity and temperature. Battery's output power and power generation also strongly depend on both these factors. The short-circuit current is proportional to the radiation intensity within several orders of magnitude and is proportional to the light-receiving area of the battery [4]. The open-circuit voltage tends to saturate quickly with radiation, regardless of the battery area. The fill factor does not have a simple, functional relationship with the radiation amount. The open-circuit voltage, fill factor, and output power all decrease with an increase of component temperature. The temperature coefficients are $-2.3 \text{ mV}/^\circ\text{C}$, $-0.0015/^\circ\text{C}$ and $-(0.004-0.005) \text{ W}/^\circ\text{C}$, respectively. As

the temperature of the components increases, the short-circuit current also increases slightly. The temperature coefficient is $+0.107 \text{ mA/}^\circ\text{C}$. Therefore, the output power of the photovoltaic system can be expressed as a function of the radiation intensity and temperature: $P_m = P_m(G, T)$. The relationship between the short-circuit current on the one hand and the photogenerated current, the incident light intensity and the temperature is shown as follows:

$$I_{sc} \approx I_L = I_{scR} \left(\frac{G}{G_R} \right) + K_T(t - t_R) \quad (3)$$

Here, I_{scR} is the component's short-circuit current obtained under standard testing conditions; is the radiation intensity under the standard test (1000 W/m^2); K_T is the temperature coefficient of the short-circuit current (amperes per degree Celsius [$\text{A/}^\circ\text{C}$]); t, t_R are the temperatures of the battery under arbitrary conditions and standard conditions (25°C), respectively. Because the temperature has less influence on the short-circuit current, Eq. (3) can be rewritten as follows:

$$I_{sc} \approx I_L = I_{scR} \left(\frac{G}{G_R} \right) = \gamma G \quad (4)$$

In the equation, γ is a constant, which is related to the battery area. We substitute Eq. (4) into Eq. (2) to obtain the open-circuit voltage:

$$U_{oc} = \frac{n_s n k T}{q} \ln(\gamma G / I_D + 1) \approx \frac{n_s n k T}{q} \ln(\gamma G / I_D) = \frac{n_s n k T}{q} [\ln(\gamma / I_D) + \ln G] = a + b \ln G \quad (5)$$

Among them, $a = \frac{n_s n k T}{q} \ln \left(\frac{\gamma}{I_D} \right)$, $b = \frac{n_s n k T}{q}$. We substitute Eqs (4) and (5) into Eq. (2) to get the fill factor, as follows:

$$\begin{aligned} FF &\approx \frac{u_{oc} - \ln(u_{oc} + 0.72)}{u_{oc} + 1} \left(1 - \frac{R_s \gamma G}{a + b \ln G} \right) = \frac{U_{oc} n_s / b - \ln[(U_{oc} n_s / b) + 0.72]}{(U_{oc} n_s / b) + 1} \\ &\times \left[1 - \frac{R_s \gamma G}{a(1 + (\ln \gamma - \ln I_D)^{-1} \times \ln G)} \right] \approx \frac{U_{oc} - b \ln(U_{oc} / b)}{U_{oc}} \times \left(1 - \frac{R_s \gamma G}{a} \right) \end{aligned} \quad (6)$$

Therefore, the output power can be expressed as

$$P_m = U_{oc} \times I_{sc} \times FF \times K_{T,P} \quad (7)$$

Here, $K_{T,P} = 1 + \varepsilon(t - t_R)$ is the weight of the relationship between output power and temperature; ε is the temperature coefficient of output power [5]. The output power on the silicon solar cell monolithic module can be obtained by actually measuring the radiation intensity on the solar cell surface and the temperature of the module. The hourly power generation capacity of the solar cell can be calculated by adding up within hours.

3 Multiple linear regression model

The general form of the multiple linear regression model is $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_j x_{ji}$, where j is the number of explanatory variables, β_j is the regression coefficient, and x_{ji} is a variable. In the photovoltaic power generation system, the radiation amount and temperature greatly influence the battery output [6]. Therefore, the article builds a multiple linear regression model of a photovoltaic power generation system around these two factors. We substitute Eqs (4)–(6) into Eq. (7) to obtain the following expressions:

$$P_m = \gamma G(a + b \ln G) K_{T,P} \{ (1 - cG) \frac{a + b \ln G - b \ln[(a + b \ln G)/b]}{a + b \ln G} \} \quad (8)$$

$c = R_s \gamma / a$. We can simplify the above formula to get

$$P_m = (\gamma G - \gamma c G^2) [a - b \ln(a/b) + b \ln G - b \ln \left(1 + \frac{b}{a} \ln G\right)] K_{T,P} \quad (9)$$

We perform Taylor series expansion on the term $\ln \left(1 + \frac{b}{a} \ln G\right)$ in Eq. (9) and take the first three terms to obtain

$$\ln \left(1 + \frac{b}{a} \ln G\right) \approx c^0 + c^1 \ln G + c^2 (\ln G)^2 \quad (10)$$

In this expression, c_0, c_1, c_2 are the expanded constant terms respectively. Eq. (10) is inserted into Eq. (9), and after finishing, we get

$$\begin{aligned} P_m &= \{ [\gamma a - \gamma b \ln(a/b) - \gamma b c^0] \\ &\quad G + (\gamma b - \gamma b c^1) G \ln G + (-\gamma b c^2) [G (\ln G)^2] \\ &\quad \begin{cases} N_1 = \gamma a - \gamma b \ln(a/b) - \gamma b c_0 & N_2 = \gamma b - \gamma b c^1 \\ N_4 = \gamma b c \ln(a/b) - \gamma a c + \gamma b c c_0 & N_3 = \gamma b c^2 \\ N_5 = \gamma b c c_1 - \gamma b c & N_6 = \gamma b c c^2 \end{cases} \\ &\quad \begin{cases} x_1 = K_{T,P} G & x_4 = K_{T,P} G^2 \\ x_2 = K_{T,P} G \ln G & x_5 = K_{T,P} G^2 \ln G \\ x_3 = K_{T,P} G (\ln G)^2 & x_6 = K_{T,P} G^2 (\ln G)^2 \end{cases} \end{aligned} \quad (11)$$

Then, Eq. (11) can be changed to

$$P_m = N_1 x_1 + N_2 x_2 + N_3 x_3 + N_4 x_4 + N_5 x_5 + N_6 x_6 \quad (12)$$

Here, N_1-N_6 are constants, x_1-x_6 are six unknowns about temperature and radiation intensity. In the grid-connected photovoltaic system, the photovoltaic array is composed of many photovoltaic modules connected in series (N_s) and parallel (N_p). The conversion efficiency of each part of the entire system is not 100%. The overall efficiency of the system η is $\eta = \eta_1 \eta_2 \eta_3$, where η_1, η_2, η_3 are the photovoltaic array efficiency, inverter efficiency, and alternating current (AC) grid-connected efficiency [7]. Then, the output power of the photovoltaic array is

$$P_A = P_m N_s N_p \eta \quad (13)$$

We substitute Eq. (12) into Eq. (13) to get

$$P_A = N'_1 x_1 + N'_2 x_2 + N'_3 x_3 + N'_4 x_4 + N'_5 x_5 + N'_6 x_6 \quad (14)$$

Here, $-N'_1-N'_6$ are constants. It can be seen that there is a multivariate linear relationship between the output power and the unknown variables. They can collect a lot of radiation intensity and temperature data. We use regression methods to obtain the six constants in Eq. (14). The predictive regression equation can be obtained. In a short period, we consider the test volume to be approximately constant [8]. At the same time, we can predict the power and output of the photovoltaic system by using the radiation and temperature measured at the last moment.

4 Data testing and regression analysis

4.1 Theoretical calculation of components

The photovoltaic system consists of 18 Kyocera KC130GH-2P polysilicon solar cell modules connected in series. The parameters of the single-piece components are shown in Table 1.

Table 1 Kyocera KC130GH-2P polysilicon module parameters

Component	P_{max} , W	U_m , V	I_m , A	U_{oc} , V	I_{sc} , A	Area, m ²
Polysilicon	130	17.6	7.39	21.9	8.02	0.81

Fig. 2 The relationship between the output power of the module and the amount of radiation.PV, photovoltaic system.

We calculate the relationship between the open-circuit voltage U_{oc} and short-circuit current I_{sc} with the radiation amount G under the standard condition of a single module. Figure 2 shows the relationship between output power and radiation and the component temperature.

When the component's temperature does not change, there is a logarithmic relationship between the open-circuit voltage and the amount of radiation. When the radiation becomes $>500 \text{ W/m}^2$, the open-circuit voltage tends to a particular value. There is a linear relationship between the short-circuit current and the amount of radiation, and there is also a linear relationship between output power and radiation [9]. When the temperature of the component gradually increases, the output power decreases significantly. Therefore, photovoltaic modules should be kept in as low-temperature working environments as possible.

4.2 Test system design and method

We tested the 2 kW grid-connected photovoltaic system for several months. The test system is shown in Figure 3. We collected data after installing the test equipment according to the requirements of the instrument. The test parameters are as follows: radiation intensity (G), ambient temperature (T_a), component back panel temperature (T_b), wind speed (v), power generation (P_A), and hourly power generation (Q). Among them, G, T_a, T_b, v and P_A are collected every 3 minutes and Q is recorded every hour.

$$\begin{cases} P_{A,I} = -1.522x_1 + 0.549x_2 - 0.083x_3 + 0.0135x_4 - 0.0028x_5 + 1.49 \times 10^{-4}x_6 \\ P_{A,II} = -0.672x_1 + 0.339x_2 - 0.056x_3 + 0.011x_4 - 0.0023x_5 + 1.5 \times 10^{-4}x_6 \\ P_{A,III} = -0.802x_1 + 0.404x_2 - 0.067x_3 + 0.013x_4 - 0.003x_5 + 1.3 \times 10^{-5}x_6 \\ P_{A,IV} = -0.47x_1 + 0.237x_2 - 0.039x_3 + 0.0076x_4 - 0.0016x_5 + 8.6 \times 10^{-5}x_6 \end{cases} \quad (15)$$

4.3 Regression analysis

We process classified and unclassified (IV) test data. The weather is classified into three categories: sunny (I), cloudy (II) and sunny to cloudy (or overcast to cloudy) (III). Then, we perform linear regression on the classified and unclassified test data. The regression results are shown in Tables 2–4.

From Tables 2–4, the following conclusions can be drawn:

- (1) The linear regression equation for the three weather types and unclassified conditions is shown in Eq. (15).

Fig. 3 Test system.AC, alternating current; DC, direct current; PV, photovoltaic system; V_{dc}

That is the photovoltaic system generating the power prediction equation.

(2) The complex correlation coefficient between the six independent variables and the dependent variable in the linear regression model is $R > 0.996$ or more [10]. The better the weather, the greater is the R -value, and the corresponding significance values are all zero. This means that the result is significant at the 0.01 level. After adjustment, $R^2 > 0.993$, which shows that the established model fits better.

Table 2 Regression statistics

Weather type	Sunny (I)	Cloudy (II)	Partly cloudy/overcast to cloudy (III)	Uncategorized (IV)
R	0.997415	0.99715	0.99652	0.99667
R^2	0.994838	0.99431	0.99314	0.99336
R -adjusted	0.98404	0.98545	0.98373	0.99026
Standard error, %	10.61	9.11	9.99	10.34
F -value	3051.26	3379.02	2,652.47	8,301.23
Significance value	0	0	0	0
Observation point	101	122	117	340

Note: R is the precision of regression coefficient; R^2 is the precision of model fitting coefficient; R -adjusted is the precision of the fitting coefficient after correction.

(3) The constant term in the regression Eq. (15) is zero. The regression coefficients are within the range of 95% confidence. The corresponding significance values are all zero. This shows that the results are significant.

(4) In sunny and cloudy weather, the classified regression equation $P_{A,I}$, can be used to predict the power generation of the photovoltaic system. Its prediction accuracy is higher than that obtained by directly using the unclassified regression equation. For sunny to cloudy (or overcast to cloudy), we use the unclassified regression equation $P_{A,IV}$ to predict the accuracy.

Table 3 Analysis of variance

Weather type		Regression analysis	Residual	Sum
Sunny (I)	<i>Df</i>	6	95	101
	<i>SS</i>	205.93	1.068	206.998
	<i>MS</i>	34.32	0.0112	–
Cloudy (II)	<i>Df</i>	6	116	122
	<i>SS</i>	166.97	0.955	167.9272
	<i>MS</i>	27.828	0.0082	–
Partly cloudy/overcast to cloudy (III)	<i>Df</i>	6	110	116
	<i>SS</i>	160.17	1.107	161.2801
	<i>MS</i>	26.69	0.0101	–
Unclassified (IV)	<i>Df</i>	6	333	339
	<i>SS</i>	532.81	3.562	536.38
	<i>MS</i>	88.8	0.0106	–

Table 4 The output results of the coefficients of each equation in the regression calculation

Weather type					N'_1	N'_2	N'_3
$N_4N_5N_6$ unny (I)	<i>X</i>	-1.15	0.55	-0.08	0.01	0.00	1.49×10^{-4}
	Standard deviation, %	24.14	113.67	16.85	2.57	0.52	0.03
	<i>T</i>	-0.48	0.48	-0.49	0.52	-0.53	0.53
Cloudy (II)	<i>X</i>	-0.80	0.40	-0.07	0.01	0.00	1.5×10^{-4}
	Standard deviation, %	63.56	31.45	5.05	0.94	0.20	0.01
	<i>T</i>	-1.26	1.29	-1.33	1.42	-1.43	1.44
Partly cloudy/overcast to cloudy (III)	<i>X</i>	-0.68	0.34	-0.06	0.01	0.00	1.3×10^{-5}
	Standard deviation, %	53.43	26.59	4.31	0.82	0.17	0.01
	<i>T</i>	-1.28	1.29	-1.32	1.38	-1.38	1.39
Unclassified (IV)	<i>X</i>	-0.47	0.24	-0.04	0.01	0.00	8.6×10^{-5}
	Standard deviation, %	34.72	17.13	2.73	0.50	0.10	5.6×10^{-3}
	<i>T</i>	-1.35	1.38	-1.43	1.51	-1.52	1.53

(5) It can also be seen from the data (Table x) that the regression coefficient and T -statistic of the unknown quantity X_1 – X_6 in the regression equation are also significant. This shows that the six established unknowns have a significant linear effect on the photovoltaic power generation system [11]. At the same time, the F -detection method is used to detect the significance of the regression equation. This also shows that the established model has an excellent fitting effect.

5 Analysis and discussion of forecasted results

We apply the prediction model to the 2 kW grid-connected photovoltaic system. Among them, the weather of sunny (I), cloudy (II), sunny to cloudy (or overcast to cloudy) (III) adopts the $P_{A,I}$, $P_{A,II}$, $P_{A,IV}$ forecast models, respectively [12]. We predict the average power generation (in kilowatts [kW]) and power generation (in kilowatt hour [kWh]) per hour for 2 consecutive days under different weather types. The forecasted values for sunny days (I) are higher than the measured values, the forecasted values for sunny to cloudy (or overcast to cloudy) (III) days are lower than the measured values, and cloudy days (II) have a better forecasting effect. The average residuals under sunny (I), cloudy (II), and sunny to cloudy (or overcast to cloudy) (III) weather are about 0.095, 0.066, and 0.126, respectively. The comparison result of the prediction residuals is shown in Figure 4.

Fig. 4 Comparison of three types of weather prediction residuals.

6 Conclusion

We have carried out theoretical calculations on the monolithic components of silicon crystal cells. On this basis, we establish a multiple linear regression model of photovoltaic system power generation and power generation with radiation and module working temperature as variables. Multiple linear regression analyses were conducted on the data under weather classification (I, II, and III) and unclassified conditions. The model has a good fit, and its accuracy is high.

Acknowledgement: Technology project of China Southern Power Grid Corporation (GDKJXM20198046): Research and application of key technologies in engineering design and review based on digital development thought – Topic 2: Research and application of key technologies in 3D design review.

References

- [1] Akhter, M. N., Mekhilef, S., Mokhlis, H., & Shah, N. M. Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. *IET Renewable Power Generation*, 2019. 13(7): 1009-1023
- [2] Zhen, Z., Liu, J., Zhang, Z., Wang, F., Chai, H., Yu, Y.,... & Lin, Y. Deep learning based surface irradiance mapping model for solar PV power forecasting using sky image. *IEEE Transactions on Industry Applications*, 2020. 56(4): 3385-3396
- [3] Jewell, J., Vinichenko, V., Nacke, L., & Cherp, A. Prospects for powering past coal. *Nature Climate Change*, 2019. 9(8): 592-597
- [4] Touchent, K., Hammouch, Z. & Mekkaoui, T. A modified invariant subspace method for solving partial differential equations with non-singular kernel fractional derivatives. *Applied Mathematics and Nonlinear Sciences*, 2020. 5(2): 35-48
- [5] Aghili, A. Complete Solution For The Time Fractional Diffusion Problem With Mixed Boundary Conditions by Operational Method. *Applied Mathematics and Nonlinear Sciences*, 2020. 6(1): 9-20
- [6] Giorgi, F., Raffaele, F., & Coppola, E. The response of precipitation characteristics to global warming from climate projections. *Earth System Dynamics*, 2019. 10(1): 73-89
- [7] Abdel-Nasser, M., & Mahmoud, K. Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Computing and Applications*, 2019. 31(7): 2727-2740
- [8] Mehr, A. D., Nourani, V., Khosrowshahi, V. K., & Ghorbani, M. A. A hybrid support vector regression–firefly model for monthly rainfall forecasting. *International Journal of Environmental Science and Technology*, 2019. 16(1): 335-346
- [9] Mehr, A. D., Nourani, V., Khosrowshahi, V. K., & Ghorbani, M. A. A hybrid support vector regression–firefly model for monthly rainfall forecasting. *International Journal of Environmental Science and Technology*, 2019. 16(1): 335-346
- [10] Zhang, Z., & Hong, W. C. Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm. *Nonlinear Dynamics*, 2019. 98(2): 1107-1136
- [11] Sanikhani, H., Kisi, O., Maroufpoor, E., & Yaseen, Z. M. Temperature-based modeling of reference evapotranspiration using several artificial intelligence models: application of different modeling scenarios. *Theoretical and Applied Climatology*, 2019. 135(1): 449-462
- [12] Ramon, J., Lledo, L., Torralba, V., Soret, A., & Doblas-Reyes, F. J. What global reanalysis best represents near-surface winds?. *Quarterly Journal of the Royal Meteorological Society*, 2019. 145(724): 3236-3251

This page is intentionally left blank