PV Power Production Estimation by Using radiometric and Meteorological Data

Trigo-González M. [1], Marzo A.[1-4], Alonso-Montesinos J.[2-3], Cortes M.[1-4], Portillo C.[1-4], Batlles F.J.[2]

[1] Centro de Desarrollo Energético de Antofagasta (CDEA). Universidad de Antofagasta, 02800, Antofagasta, Chile)

[2]Departamento de Química y Física, Universidad de Almería, España.

[3]CIESOL, Joint Centre of the Universidad de Almería-CIEMAT, España.

[4]Solar Energy Research Center (SERC-Chile), Chile.

Abstract

World total installed photovoltaic (PV) capacity reached 320 GWp by the end of 2017. New market shave shown a rapid development in the implementation of large PV plants connected to the grid, but it is needed to face new challenges to overcome barriers for the massive deployment. The incorporation of renewable energy sources, such as solar, can affect the stability of the electrical grid. Predicting the energy production of PV plants is crucial for the massive integration of PV technologies in the grid. The aim of this research is to estimate the energy production within an infrahour period using minimal knowledge about weather conditions. In order to achieve this goal, a mathematical model is proposed which directly computes the energy as a function of module temperature and the available solar resource. The results of the model validation test show a root mean square error value of 6.42% and an mean bias error value of 1.51% and a good performance regardless the sky condition.

Keywords: estimation, production, photovoltaic, multiple linear regression, solar energy.

1. Introduction

Total installed capacity in photovoltaic (PV) plants worldwide reached 320 GW at the end of 2016 (Ise, 2018). The main contributors in this capacity are China with 26%, Europe with 33%, North America with 15%, Japan with 13%, and the rest of the countries of the world with 13%. New markets such as Asians, Africans and South Americans have shown a rapid development in the implementation of large photovoltaic plants connected to the grid, however, they face great challenges to overcome the different barriers that a massive production of photovoltaic electricity implies (Haas et al., 2018). Among these challenges, it is worth highlighting the variability of the solar resource. At present, there is a great business interest in photovoltaic technology, both in the issue of energy self-consumption, and in large photovoltaic plants, connected to the electricity grid. This fact is mainly due to the decrease in the prices of photovoltaic panels. At the end of 2016, the average sale price of the photovoltaic panel, worldwide, was of the order of 0.37 USD / Wp (Metz et al., 2017). The variability of solar radiation and climatic conditions have a direct impact on the production of a photovoltaic plant. To massively integrate photovoltaic technology, it is essential to accurately estimate the electrical production of the plant.

The methods used to predict the production of PV plants can be classified as indirect and direct. The indirect ones require first, to predict the solar irradiance (Moretón et al., 2017), and later determine the production of power through a parametrized model of the plant. On the other hand, direct methods directly determine the power produced by statistical techniques and artificial intelligence. Additionally, there are hybrid methods, which are a mixture of both methods. The largest effort in research has been made in direct methods, which correspond to 72% of published works, while 11% and 17% correspond to indirect and hybrid methods (Antonanzas et al., 2016). Statistical models do not need information on the internal model of the plant. These are fed with past data from which relationships are extracted that allow predicting the behavior in real time and future of the plant. The quality of historical data is essential for a good quality estimate. This technique is superior to indirect methods; however, it requires a large amount of historical data (meteorological, and power measurements, among others) recorded with the plant operating. This method has the advantage that it allows to correct systematic errors that originate in the measurements of the input variables.

The main statistical techniques used are regressive, autoregressive and artificial intelligence models, among which the Artificial Neural Network (RNA) stands out. The regressive techniques determine the correlation between a

doi:10.18086/eurosun2018.02.18 Available at http://proceedings.ises.org

dependent variable, in our case the electrical energy produced by the plant and the independent variables, such as radiometric, meteorological variables, as well as those of the photovoltaic panels. As an example, in (Zamo et al., 2014) different regression methods were implemented to study the relationship between energy produced and meteorological variables. The results obtained show that the RMSE for individual plants is between 9% and 10%, while the prediction of the total production of the plants presents an error between 10% and 12%.

Different regressive methods have been proposed by the international literature. Notable among these is Osterwald's (Fuentes et al., 2007) which expresses the maximum power of the cell according to the maximum power under standardized test conditions (STC), and the temperature coefficient of the maximum power of the cell. Another model is the one proposed by Araujo-Green (Araujo et al., 1982) which expresses the power produced as a function of the irradiance, the electrical variables and the parameters of the modules. Investigations carried out (Almonacid et al., 2011) show that the relative error of the Osterwald model is in the range of 15% - 21%, and 14% - 18% for the Araujo-Green method. More recent models consider nonlinear relationships between the power produced and variables such as the irradiance of the inclined plane, ambient temperature, and wind speed (Dias et al., 2017, Myers, 2009).

From the reviews made in the previous paragraphs it is concluded that the main effort has been oriented to the development of short-term to long-term prediction models of solar radiation and energy production of photovoltaic plants. However, few works are reported in relation to estimation of real-time power production of photovoltaic plants. Considering this aspect, the objective of this work is to develop a model that allows to estimate the production of a photovoltaic plant from minimum radiometric and meteorological data. To do this, a linear regression model is designed to estimate PV production with the minimum environmental parameters.

2. Materials and methods

2.1 Climatological description of the place under study

The Antofagasta Region is one of the four Chilean Regions located in the Atacama Desert in the north of the country. This Region is characterized by having a well-defined morphology and four climatological zones as defined by the Chilean Meteorological Directorate in (Cruz Silva and Calderón Suenzen, 2008), see Fig. 1

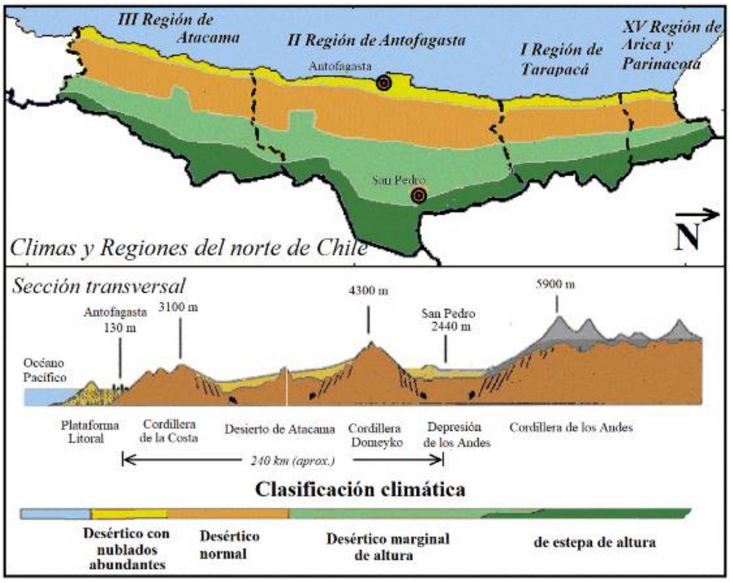

Fig. 1 - Above: climatic zones of the Northern Regions of Chile: XV-Arica and Parinacota, I-Tarapacá, II-Antofagasta and III-Atacama. Below: cross section of the Antofagasta Region, its different geological formations and elevations, and climatic classification.

The study area is located in the town of San Pedro de Atacama, is located in the interior of the desert, at 2440 meters above sea level, 22.55 degrees south and 68.12 degrees west. A high thermal amplitude of 20 ° C is registered throughout the year and its annual precipitation is 68.2 mm, concentrating most during the summer season, with 19.6 mm. The maximum registered temperature is 28.6 ° C during the summer and 19.1 ° C in winter, reaching minimum values of 0 ° C at this time of year. The location of this town, where the desert climate predominates, makes it possible to observe a high number of clear days throughout the year.

2.2. Description of the PV plant used in the study

The Energy Development Center of Antofagasta (CDEA) of the University of Antofagasta, since 2010 has a pilot photovoltaic plant consisting of cadmium telluride modules (CdTe) located in the basic school E-26 of San Pedro de Atacama. This plant consists of 40 CdTe modules with an inclination angle of 10 ° facing north. Its nominal power is 3.00 kWp with an area of 28.8 m2. The modules are divided into two groups, each consisting of 8 strings in parallel that are connected to an SMA SB3000TL-20 inverter.

2.3. Instrumentation and database

The parameters measured in the plant previously mentioned for the realization of the present study were: the environmental temperature (Tamb), the temperature of the module (Tmod), the irradiation in the generation plane (G), the wind speed (WS) and the energy injected into the network (Eac).

All these variables were recorded in a Weblog PRO (ADQ) datalogger. Table 2 shows the technical data of the monitoring system of the photovoltaic plant.

Tab. 1: Instrumentation used to record data in the PV plant

Nomenclature	Instrument	Model	Measurement range
G	Calibrated photocell	Si-12Tc	0 a 1500 W m^{-2}
Tamb	Thermocouple	PT100	-20, +100 °C
Tmod	Thermocouple	PT100	-20, +100 °C
WS	Anemometer	Cup	0.8, 40 m/s
ADQ	Data acquisition system	Meteocontrol	16 bits

From the San Pedro plant, 3767 registered samples were used with a sampling frequency of 5 minutes during the years 2010 and 2014.

3. Description of the methodology

The objective of this work is to estimate the production of PV plants in Wh terms by means of the multiple linear regression method (MLR). Figure 2 describes the steps of the proposed methodology.

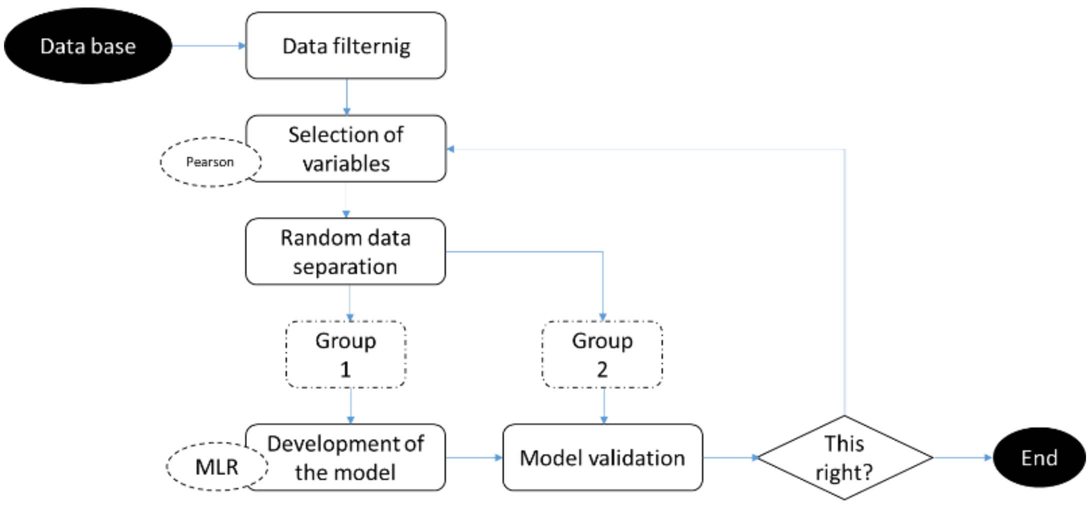

Fig. 2 - Scheme of the proposed methodology for the development of the production estimation model in PV plants.

- From the available database, the elimination of atypical data that interfere in the adjustment of the model is done. And at the same time the hour values of the variables were calculated, with the aim of developing a model that estimates the hourly photovoltaic production.
- Subsequently, a statistical study was carried out to measure the degree of correlation between the different variables by calculating the Pearson coefficient and, in this way, selecting the parameters to be taken into account.
- The database was randomly divided into two groups: the first group of 2/3 of the database was used for the development of the model, the second group consisted of one third of the data and served for the validation of the generated model.
- The model was adjusted using the statistical technique of multiple linear regressions (MLR).
- The models were validated using the group of validation data, different from the one used to adjust the model, to corroborate the effectiveness of the extrapolation to other geographical locations.

4. Results and discussion

4.1 Selection of variables

For the selection of the variables to be used in the model, the Pearson coefficient between them was calculated. This allowed to identify the degree of direct correlation of the variables with the production of the PV plants. This coefficient varies in the interval [-1, 1]: when the absolute value of the coefficient approaches one, it indicates a linear correlation; while moving away from this value, approaching zero, shows that the correlation between variables is not linear (Hall, 2000). This criterion allowed discarding variables that did not contribute to the adjustment of the model.

Figure 3 shows a set of graphs that represent the correlation between the variables measured in both PV plants, and the result of the Pearson coefficient of each variable is also shown.

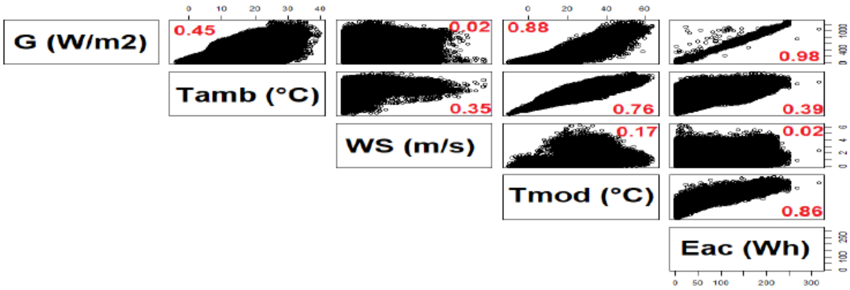

Fig. 3 - Pearson coefficients of the set of variables measured in the PV plants of Cd-Te technology.

If we look at the graphs in the last column, these represent the relation of the Photovoltaic Energy (EAC) vs the meteorological variables that are G, Tamb, WS and Tmod, the results of the Pearson coefficient were 0.98, 0.39, 0.02, 0.86, respectively. Which leads to selecting the variables G and Tmod for the adjustment of the model, because it has a very high linear correlation with photovoltaic energy.

4.2 Development of the estimation model of PV plant production with RLM

As mentioned above, the present work focused on the calculation of a model for estimating the production of PV plants based on CdTe technology. For this purpose, two groups of data were randomly separated: one was used for the development of the models, while the remaining quantity was reserved for the validation of the same models.

The expression proposed in equation 1, relates the statistical adjustment of the energy production of the plant (PPV) in Wh with the hour average of the module temperature under insolation conditions ($\Delta Tmod$) in ° C, the hourly solar irradiation (G) in Wh m^{-2}.

$$P_{pv} = 4.08 + 0.21 \cdot G - 0.03 \cdot T_{mod} \qquad \text{(eq. 1)}$$

A method to quantify the effectiveness of the estimation is by means of two statistical indicators, namely: the root means square error (RMSE) method, given by the equation. 2, and the average bias error (MBE), given by Eq. 3.

$$RMSE(\%) = \frac{100}{Eac_{media}}\left[\sqrt{\frac{1}{N}\sum_{j=1}^{N}(Eac_{est.} - Eac_{med.})^2}\right] \quad \text{(eq. 2)}$$

$$MBE(\%) = \frac{100}{Eac_{media}}\left[\frac{1}{N}\sum_{j=1}^{N}(Eac_{est.} - Eac_{med.})\right] \quad \text{(eq. 3)}$$

The result of the validation was an RMSE of 6.42% with an MBE of 1.51%.

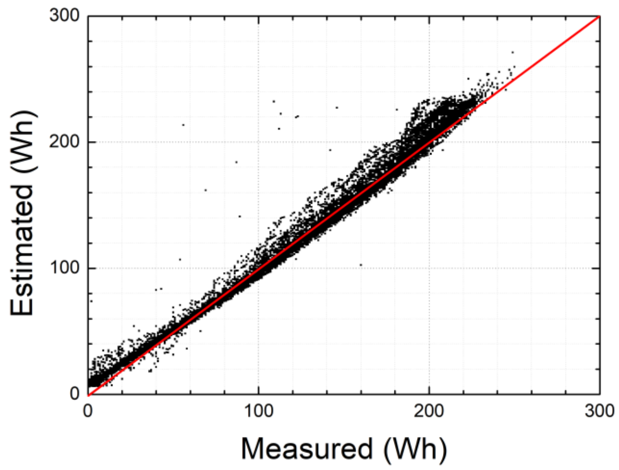

Fig. 4 – Estimation of the PV production of a CdTe plant at San Pedro de Atacama, Chile, by using the model base don multiple linear regression method.

4.3 Test

A test was carried out with days cleared, partially cloudy and totally cloudy, using the data registered in the town of San Pedro de Atacama. The objective is to study the effectiveness of the model in different sky type conditions, with the possibilities of extrapolating the results of the estimates in other places. The results are shown in the following table.

Tab. 2: Results of the errors of the statistical indicators for different types of sky.

Type of sky	MBE%	RMSE%	R^2
Clear sky	2.12	5.40	0.92
Party cloudy sky	-1.75	5.35	0.95
Cloudy sky	10.29	17.24	0.85

The results of the table indicate that, for a clear and partially cloudy type of sky, the results are approximately equal, with an RMSE of 5.4% and 5.35%, respectively. It is observed that for the partially cloudy day the MBE has to underestimate, with respect to the others.

In the case of the cloudy sky type, the error is higher, and this is mainly due to the fact that the adjustment of the model was made under conditions of a climate that most of the year is with clear days.

Figure 5 shows the actual measurement of the energy generated by the photovoltaic plant (black color) and the model estimates (red color), for the three conditions of the sky.

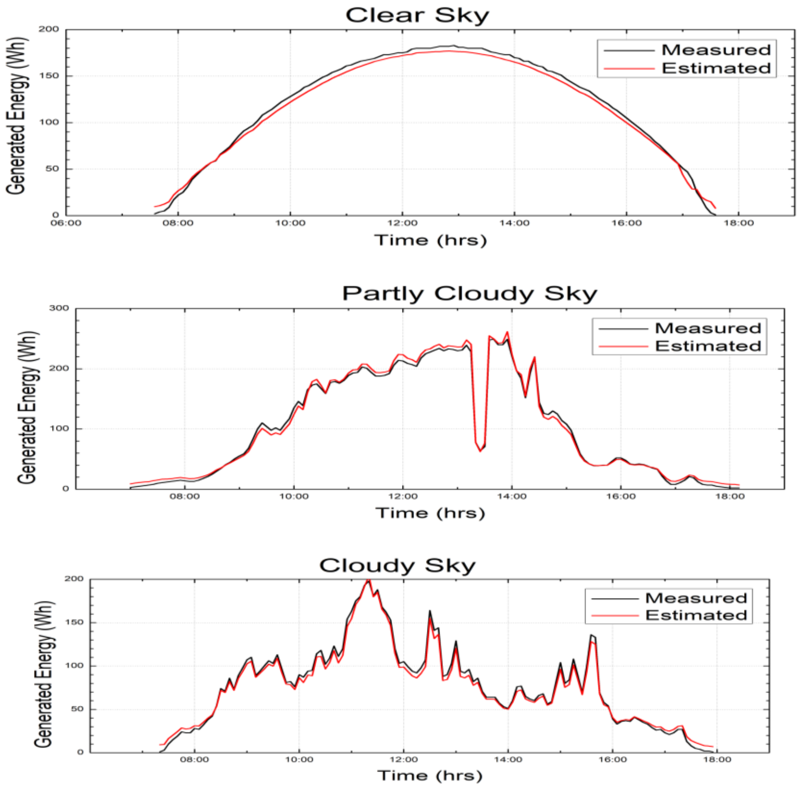

Fig. 5 - Results of the test for different sky conditions, with a model adjusted by the MRL technique

The results highlight the good fit of the model for any condition of the sky, observing in the previous graphs that the model has a good response to follow the trend of the radiation curve for the 3 types of sky conditions.

5. Conclusions

- In this work we propose a model to estimate the electrical production of a CdTe technology photovoltaic plant by using multiple linear regression method.
- The Pearson coefficient was used to select the variables of the model. Irradiance and module temperature showed the highest linear correlation with photovoltaic energy production.
- The results of the model validation show a root mean square error value of 6.42% and an mean bias error value of 1.51%
- The model continues to show good performance according to the generation of energy regardless the sky condition.

References

Almonacid, F., Rus, C., Pérez-Higueras, P., Hontoria, L., 2011. Calculation of the energy provided by a PV generator. Comparative study: Conventional methods vs. artificial neural networks. Energy 36, 375–384. doi:10.1016/J.ENERGY.2010.10.028

Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F.J., Antonanzas-Torres, F., 2016. Review of photovoltaic power forecasting. Sol. Energy 136, 78–111. doi:10.1016/J.SOLENER.2016.06.069

Araujo, G.L., Sánchez, E., Martí, M., 1982. Determination of the two-exponential solar cell equation parameters from empirical data. Sol. Cells 5, 199–204. doi:10.1016/0379-6787(82)90027-8

Cruz Silva, C., Calderón Suenzen, J., 2008. Guía Climática Práctica. Dirección Meteorológica de Chile, Santiago de Chile, Chile.

Dias, C.L. de A., Castelo Branco, D.A., Arouca, M.C., Loureiro Legey, L.F., 2017. Performance estimation of photovoltaic technologies in Brazil. Renew. Energy 114, 367–375. doi:10.1016/J.RENENE.2017.07.033

Fuentes, M., Nofuentes, G., Aguilera, J., Talavera, D.L., Castro, M., 2007. Application and validation of algebraic methods to predict the behaviour of crystalline silicon PV modules in Mediterranean climates. Sol. Energy 81, 1396–1408. doi:10.1016/j.solener.2006.12.008

Haas, J., Palma-Behnke, R., Valencia, F., Araya, P., Díaz-Ferrán, G., Telsnig, T., Eltrop, L., Díaz, M., Püschel, S., Grandel, M., Román, R., Jiménez-Estévez, G., 2018. Sunset or sunrise? Understanding the barriers and options for the massive deployment of solar technologies in Chile. Energy Policy 112, 399–414. doi:10.1016/J.ENPOL.2017.10.001

Hall, M.A., 2000. Correlation-based feature selection of discrete and numeric class machine learning, Computer Science Working Papers, Computer Science Working Papers. University of Waikato, Department of Computer Science.

Ise, F., 2018. Photovoltaics Report.

Korachagaon, I., Mudgal, D.N., Kottur, R.M., Patil, S.K., Bapat, V.N., 2015. Global solar radiation estimation model with two parameters and its ANN validation, in: 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO). IEEE, pp. 1–3. doi:10.1109/ISCO.2015.7282286

Marzo, A., Trigo-Gonzalez, M., Alonso-Montesinos, J., Martínez-Durbán, M., López, G., Ferrada, P., Fuentealba, E., Cortés, M., Batlles, F.J., 2017. Daily global solar radiation estimation in desert areas using daily extreme temperatures and extraterrestrial radiation. Renew. Energy 113, 303–311. doi:10.1016/j.renene.2017.01.061

Mellit, A., Sağlam, S., Kalogirou, S.A., 2013. Artificial neural network-based model for estimating the produced power of a photovoltaic module. Renew. Energy 60, 71–78. doi:10.1016/J.RENENE.2013.04.011

Metz, A., Fischer, M., Trube, J., 2017. International Technology Roadmap for Photovoltaics (ITRPV) 8 th edition: Crystalline Silicon Technology–Current Status and Outlook.

Monteiro, R.V.A., Guimarães, G.C., Moura, F.A.M., Albertini, M.R.M.C., Albertini, M.K., 2017. Estimating photovoltaic power generation: Performance analysis of artificial neural networks, Support Vector Machine and Kalman filter. Electr. Power Syst. Res. 143, 643–656. doi:10.1016/j.epsr.2016.10.050

Moretón, R., Lorenzo, E., Pinto, A., Muñoz, J., Narvarte, L., 2017. From broadband horizontal to effective in-plane irradiation: A review of modelling and derived uncertainty for PV yield prediction. Renew. Sustain. Energy Rev. 78, 886–903. doi:10.1016/j.rser.2017.05.020

Myers, D., 2009. Evaluation of the Performance of the PVUSA Rating Methodology Applied to Dual Junction PV Technology: Preprint (Revised).

Quesada-Ruiz, S., Linares-Rodríguez, A., Ruiz-Arias, J.A., Pozo-Vázquez, D., Tovar-Pescador, J., 2015. An advanced ANN-based method to estimate hourly solar radiation from multi-spectral MSG imagery. Sol. Energy 115, 494–504. doi:10.1016/J.SOLENER.2015.03.014

Zamo, M., Mestre, O., Arbogast, P., Pannekoucke, O., 2014. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production, part I: Deterministic forecast of hourly production. Sol. Energy 105, 792–803. doi:10.1016/J.SOLENER.2013.12.006