Photovoltaic Power Generation Estimation Using Statistical Features and Artificial Neural Networks

D A Elvira-Ortiz, D Morinigo-Sotelo, R J Romero-Troncoso, and R A Osornio-Rios

1HSPdigital – CA Mecatronica, Facultad de Ingenieria Campus San Juan del Rio, Universidad Autonoma de Queretaro, Rio Moctezuma 249, 76807 San Juan del Rio, Queretaro, Mexico
2HSPdigital – Research Group ADIRE, University of Valladolid, UVa., Paseo del Cauce, 59, 47011 Valladolid, Spain

Received 22 June 2018; revised 04 December 2018; accepted 17 February 2019

Photovoltaic generation completely depends on environmental factors like sun irradiance and cell temperature. Therefore, it is necessary to develop methodologies that allow to predict the power generated by photovoltaic systems under different weather conditions. This work presents a methodology for estimating the power delivered by a photovoltaic inverter using statistical features coming from weather signals and an artificial neural network for predicting the power level delivered by the photovoltaic system.

Keywords: Artificial neural network; Photovoltaic systems; Statistical features; Power generation; Environmental factors

Introduction
The recent environmental and economic problems have led to a growth in the inclusion of renewable energies, being photovoltaic (PV) generation the most promising technology. However, PV generation is variable and introduce some problems because it depends on environmental factors. In this sense, several methodologies using signal processing techniques have been developed for the forecasting of PV system output. Artificial neural networks (ANN) are a powerful tool for predictions tasks, making them one of the most common solution for this and other tasks regarding PV generation. Despite the advantages offered by ANN they may turn complex when the inputs are not well selected. In this sense, it is possible that the use of statistical features instead of using the raw signals, simplifies the resultant neural network. This paper presents a methodology that fuses the statistical feature extraction with ANN for the estimation of the power delivered by a 100 kW PV inverter.

Statistical features
Statistical features provide information regarding the behavior of the environmental parameters. Additionally, these features provide more information than the raw data, because the statistical features change between two days with similar conditions when these days occur in different seasons three of the year. For this particular work it is proposed the use of features: mean, variance and skewness. The mean represents a measure of the central tendency from a data set and it is mathematically expressed by Eq. (1)

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \ldots \ (1)$$

where: $\bar{x}$ is the mean of the data set, $n$ is the number of samples, and $x_i$ is the $i$-th sample from the data set.

The second feature is the variance that measures how far is a set of numbers from its mean or expected value. The expression used for the variance computation is presented in Eq. (2).

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \ldots \ (2)$$

Where: $\sigma^2$ is the variance, $\bar{x}$ is the mean of the data set, $n$ is the number of samples, and $x_i$ is the $i$-th sample from the data set.

Finally, the third order statistic called skewness is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean, and it is defined in Eq. (3).

$$S_k = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^3 \quad \ldots \ (3)$$

Where: $S_k$ is the skewness of the set of values, $\bar{x}$, $n$ and $x_i$ are the same variables defined in Eq. (1) and Eq. (2).
Active power

The active power \( P \) (also called real power), is the average value of the instantaneous power \( p(t) = v(t) \cdot i(t) \) during a specific time interval \( T \), and its mathematical expression is presented in Eq. (4).

\[
P = \frac{1}{T} \sum_{k=1}^{n} v_k \cdot i_k \quad \ldots (4)
\]

Where \( v_k \) and \( i_k \) are the \( k \)-th sample of the voltage and current signals respectively and \( n \) is total number of samples comprised in the time interval \( T \).

Power estimation in PV generation

Figure 1 shows the methodology followed for the developing of this work. This methodology requires to measure and store two variables associated with environmental conditions: sun irradiance and PV cell temperature. The global irradiance that reaches the PV panels is measured every minute using a reference cell, whereas the cell temperature is measured using a thermocouple also on a minute basis. Additionally, it is necessary to acquire the voltage and current signals delivered by the PV inverter for their later use in the active power computation. Voltage and current signals are acquired using a proprietary data acquisition system (DAS). The DAS acquires and stores data from 7 seven simultaneous channels: 3 for voltage signals and four for current signals. The data are acquired at sampling rate of 8000 samples per second with a 16-bit resolution. Additionally, the DAS incorporates micro-SD memories for storing the totality of the acquired data. For this particular application, a DAS with lower sampling frequency and resolution may be used without compromising the results. However, the large storage capacity provided for the incorporated micro-SD memory is necessary in order to be able of collecting the data from several days. The PV inverter under test is a three-phase Ingecon sun 100 that delivers 230V at 50Hz and a rated power of 100kW. Then, the environmental parameters are passed through three blocks for obtaining some statistical features from these signals. The first block is the mean computation, the second is the variance computation, and the third one is the skewness computation. Each one of these blocks delivers two outputs that represent the mean, variance, and skewness from the two environmental variables. There are some other features that could be considered. Nevertheless, to identify if additional statistical features are whether relevant or redundant for the analysis, it is necessary to perform a feature selection and optimization task using techniques such as the principal component analysis (PCA) or the linear discriminant analysis (LDA)\(^{5}\). However, this type of analysis is left as a prospective for future work. Every feature is computed on a five-minute basis, and the six statistical features that have been extracted are used as inputs for the next block that is the application of the ANN. The ANN proposed in this work is a simple perceptron with six neurons in the input layer. The six inputs correspond to the mean, variance and skewness of the sun irradiance, and the mean, variance, and skewness of the PV cell temperature. Also, the ANN counts with ten neurons in the hidden layer that will be in charge of processing the inputs and finding the best fitting with the targets of the ANN. For the output layer, the ANN has only one neuron that corresponds to the estimation of the power delivered by the PV inverter. A backpropagation algorithm has been used for the training stage, and a maximum iteration number of 300 is set as stop criterion. The proposed methodology also performs the active power
computation. This block uses Eq. (4) and it is observed that the output of this block enters to the ANN block. This is because the active power is used as target for training the proposed ANN. By means of consistency, the active power is calculated using the same time interval defined for the statistical features. This work uses the data from nine different days comprised between November 7\textsuperscript{th} and November 15\textsuperscript{th}, 2016. Seven days are dedicated for the training process of the ANN and the two remaining days allows observing the output of the ANN and validating the methodology. The last step of the proposed methodology consists on comparing the active power delivered by the PV inverter with the power estimated with the proposed methodology to determine the reached level of fitting. It is important to mention that this work does not aim to perform a cloud presence forecasting but a PV system output estimation under different atmospheric conditions.

**Results and Discussion**

Figure 2 shows the profiles for sun irradiance and PV cell temperature for the two days of analysis. Day 1 (see Figure 2(a)), presents a pattern without any abrupt variation in irradiance and temperature during the day. The irradiance profile of the second day (see Figure 2(b)), presents a lot of abrupt and unexpected variations associated with a severe cloud presence. These variations also affect the cell temperature. Figure 2 also presents the comparison between the active power calculated using Eq. (4) and the power estimated by the proposed ANN. It is observed that the performed estimation is good. In day one (Figure 2(c)), it can be seen that the blue line (power estimated with the ANN) and the red line (active power) are very close to each other most of the day. For day two (Figure 2(d)), the estimation presents a higher deviation from the real values than the one presented for the first day of analysis. Nevertheless, the approximation remains in a reasonable range. From Figure 2 it can be inferred that the unexpected variations associated with clouds introduce a bigger error in the estimation performed by the ANN. Though, it is observed that the error is always in reasonable values making this methodology useful for estimating the short-term power delivered by a PV inverter under different weather conditions. It can be inferred from the results that both parameters, irradiance and temperature, are relevant for the power forecasting; however, the sun irradiance presents a more significant contribution because its correlation with the PV inverter output is higher than the one presented by the temperature. To quantify the accuracy of the estimation, it is carried out a goodness of fit test. This test delivers the percentage of fitting that exist between a reference variable and an

![Fig. 2](image-url) — Sun irradiance and cell temperature for (a) the first day of analysis, and (b) the second day of analysis; and comparison between the real power and the estimated with the proposed methodology for: (c) first day of analysis, and (d) second day of analysis.
estimated value. The mathematical expression that describes this test is presented in Eq. (5).

\[
fit = 1 - \frac{\|X_{\text{ref}} - \bar{X}\|^2}{\|X_{\text{ref}} - \bar{X}_{\text{ref}}\|^2} \cdot 100
\]  

(5)

Where: \(fit\) is the percentage of fitting between the estimated and the reference variables, \(X_{\text{ref}}\) is the vector containing the reference variable, \(\bar{X}\) is the vector containing the estimated variable, \(\bar{X}_{\text{ref}}\) is the mean of the reference variable, and the symbol \(\|\|\) represents the norm of the vector. The value of the goodness of fit reaches a value of 99.3363 in an open sky day and 92.9360 in a cloudy day. This result corroborates the fact that the variations associated with cloud presence difficult the estimation task.

Conclusions

PV generation is a promising technology for dealing with recent economic and environmental problems regarding power generation. For this reason, it is important to count with methodologies for developing models that allow estimating the power that will be delivered by a PV system under different environmental conditions. The use of statistical features from environmental parameters along with ANN proved to be an effective tool for estimating the power delivered by a PV inverter. Using statistical features allows to simplify the ANN required for the estimation task. Results prove that a simple perceptron with six inputs and 10 neurons on the hidden layer, is enough for reaching fitting values higher than 90% regardless the weather conditions of the day under test. This is an advantage because it reduces the computational effort required in comparison with other methodologies that uses more complex topologies. This methodology aims to be a useful tool for preventing losses on industries that work with PV generation. Moreover, this methodology could be extended for estimating not only the power generation but also the quality of the power delivered by the system.

Acknowledgements

This work was partially supported by CONACyT scholarship 415315 and by project FOFl –UAQ 2018 FIN201812.

References