Expert Systems with Applications 40 (2013) 7141–7150

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Data mining and statistical techniques for characterizing the performance of thin-film photovoltaic modules

Rafael Moreno Sáez [a,1], Mariano Sidrach-de-Cardona [a,1], Llanos Mora-López [b,2,*]

[a] Departamento de Física Aplicada II, E.U. Politécnica, Universidad de Málaga, Campus de Teatinos, 29071 Malaga, Spain
[b] Departamento de Lenguajes y Ciencias de la Computación, ETSI Informática, Universidad de Málaga, Campus de Teatinos, 29071 Malaga, Spain

ARTICLE INFO

Keywords:
Solar spectral distribution
Photovoltaic modules
Performance ratio
Data mining
Statistical models

ABSTRACT

A method for characterizing the performance ratio of thin-film photovoltaic modules based on the use of data mining and statistical techniques is developed. In general, this parameter changes when modules are working in outdoor conditions depending on irradiance, temperature, air mass and solar spectral irradiance distribution. The problem is that it is usually difficult to know how to include solar spectral irradiance information when estimating the performance of photovoltaic modules. We propose five different solar spectral irradiance distributions that summarize all the different distributions observed in Malaga. Using the probability distribution functions of these curves and a statistical test, we first checked when two spectral distributions measured can be considered to have the same contribution of energy per wavelength. Hence, using this test and the *k*-means data mining technique, all the measured spectra, more than two hundred and fifty thousand, are clustered in only five different groups. All the spectra in each cluster can be considered as equal and the *k*-means technique estimates one centroid for each cluster that corresponds to the cumulative probability distribution function that is the most similar to the rest of the samples in the cluster. The results obtained proves that 99.98% of the functions can be considered equal to the centroid of its cluster. With these five types of functions, we have explained the changes in the performance ratio measured for thin-film photovoltaic modules of different technologies.

1. Introduction

The increase in the performance of solar photovoltaic modules is a pressing issue in order to improve solar photovoltaic market share and make this industry more competitive. There are many elements where this efficiency can be achieved. The elements to be studied to better establish module performance are the key parameters given by the standard test conditions (STC) used to characterize and correct modules as defined by some international standards such as IEC 60904-3 (IEC, 1989) or IEC 61215 (IEC, 1993). The typical standard variables are temperature (25°), irradiance (1000 W/m^2) and AM 1.5 spectrum. This paper is focused on the study and characterization of the spectrum (solar irradiance spectral distribution) and its influence on thin-film photovoltaic devices. The performance of these devices has a greater dependence on the spectrum than photovoltaic modules built using conventional technologies such as mono and policristalin-Si.

Two standards for the AM 1.5 spectrum were defined for terrestrial use by the photovoltaic industry and the ASTM (American Society for Testing and Materials) (ASTM, 2012). The importance of this parameter lies in the fact that photovoltaic modules have a spectral response, (Martín and Ruiz, 1999), which will only allow a part of the whole solar spectral irradiance received on their surface 'to be seen'. Spectral response is a characteristic of every cell technology. Depending on this factor a module will be able to provide more or less energy for one quantity of irradiation received. To study how solar spectral irradiance affects different cell technologies and to explain the different module performances observed, it is first necessary to have an indepth knowledge of the solar spectrum in any outdoor conditions where photovoltaic modules will be operating.

Previous papers used different parameters to characterize solar spectrum. These include the Spectral Factor (SF), (Fabero and Chenlo, 1991), Mismatch Factor (M, MM or MMF), (Poissant et al., 2003), Usefull Fraction (UF), (Gottschalg et al., 2003), Photovoltaically Active Fractions (PAFs), (Berman et al., 1999), and the Average Photon Energy (APE), (Gottschalg et al., 2003). One of the reasons why solar spectrum is not among the typical variables is that measuring solar spectral irradiance involves using very expensive devices. We propose to analyze the APE as this parameter uses a single number to characterize a whole

* Corresponding author. Tel.: +34 951952299.
E-mail addresses: rmorenosaez@ctima.uma.es (R. Moreno Sáez), msidrach@ctimauma.es (M. Sidrach-de-Cardona), llanos@lcc.uma.es, llanos@uma.es (L. Mora-López)rmorenosaez@ctima.uma.es (R. Moreno Sáez), msidrach@ctima.uma.es (M. Sidrach-de-Cardona), llanos@lcc.uma.es, llanos@uma.es (L. Mora-López).
[1] Tel.: +34 951952299.
[2] Tel.: +34 952132802.

0957-4174/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2013.06.059

spectrum what makes easier for it to be included and treated as another meteorological variable, once calculated. The uniqueness of the relationship between the solar spectral irradiance distribution and the APE is proven, (Minemoto et al., 2009); however, these authors have proven this correspondence by dividing the values of spectral irradiance into several 50 nm bands and using these integrated spectral irradiance values. We propose a different approach, based on the use of a statistical test that enables us to use all the values of the spectrum. Moreover, the authors suggest that the results obtained should be verified with data from locations with different climatic conditions.

In this study we are going to characterize the spectra at the location of Malaga, Andalusia (Spain) with a Mediterranean climate whose typical features are high humidity and mild and warm temperatures. Solar spectral irradiances have therefore been recorded at the laboratory facilities of the Solar Photovoltaic Energy group at the University of Malaga during a period of over one year. By using more than one year of data a complete range of spectrum under real working conditions is obtained and a possible seasonal effect is isolated. Once the spectrum is defined a more precise prediction of the energy produced by modules can be performed. This will result in better tools for the photovoltaic industry to size up photovoltaic plants and their integration into the power grid.

One of the objectives of this paper is to determine the different types of solar spectral irradiance distribution curves and how these different curves are related to the different estimated APE values. Once these relationships are established, further goal of this paper is to analyze how the performance ratio (PR) for different photovoltaic modules is explained using the relationship found. PR values for these modules are estimated using data collected outdoors for over one year.

We propose to use both statistical and data mining techniques that allows us to handle a large quantity of data in an automatic way to analyse and characterize the different APE values in our laboratory. On the one hand, we have used the well known Kolmogorov–Smirnov two sample test to check whether or not two distributions are the same (homogeneity test). On the other hand, we have used the *k*-means data mining technique to cluster all the estimated APEs in several clusters where all the measurements included have the same cumulative probability distribution of solar spectral irradiance.

Many different areas have used clustering including text mining, statistical learning and pattern recognition, (Duda et al., 2001; Hastie et al., 2001; Jain et al., 1999). In previous papers, spectra were gathered using the APE value but data mining techniques will here be used to characterize the spectra in a few types, aimed at trying to put into the same group spectra with different values of APE but with the same relative distribution of the irradiance for each wavelength.

Once the different solar spectral irradiance measurements are characterized taking into account the similarities observed, this information has been used to analyze the performance ratio of different modules for many different outdoor conditions.

In the second section, a detailed description of the data set is presented. In the third section the values of performance ratio for thin-film modules of different technologies are depicted and analyzed depending on several parameters. The fourth section is dedicated to analyzing the relationship between solar spectral irradiance and APE value using statistical techniques and to characterizing the different types of spectra irradiance measured using the *k*-means data mining technique. The fifth section is dedicated to presenting the obtained results and discussing these results for their use in the characterization of performance ratio of thin-film photovoltaic modules. Finally, the last section summarizes the conclusions of the work.

2. Data set

The device used is a grating spectroradiometer that works on the spectral range of visual and near-infrared where most of the cell technologies have their spectral response. The model used for this study is a Grating Spectroradiometer EKO MS-710 prepared for continuous outdoor exposure. It has a silicon sensor that allows a spectral measurement ranges from 330 to 1050 nm, VIS and NIR. The collection time is reduced to between 10 ms and 5 s, so that a fine spectrum with no distortion is obtained on days with moving clouds. This device is placed on a fixed 21° slope frame where continuous spectra were captured during a period of over one year, in the Solar Photovoltaic Energy group facilities at the University of Málaga (Spain), latitude 36.7°N, longitude 4.5°W, height 50 m. The spectroradiometer has a glass dome. To avoid dust and other substances being deposited on the dome, daily cleaning and maintenance are carried out in order to measure solar spectral irradiance as more accurate as possible. There are some other photovoltaic devices close to the spectroradiometer such as piranometers, calibrated cells and modules of different technologies on the same tilted frame structure and placed on the same plane, with the same angle of inclination as recommended in the international standard IEC 60904-7 (IEC, 1995).The manufacturer provides its own software to use the spectroradiometer to measure solar spectra. A custom-built system for I–V curves developed by Piliougine et al. (2011), at the group of PV systems of the University of Malaga was used to collect the meteorological data and the electrical module parameters.

For over a full year, solar spectral irradiances were collected all day long from sunrise to sunset at the rate of one spectrum per minute between November 2010 and May 2012, both included. Solar spectral samples were measured under the full range of meteorological conditions expected in Malaga. This meant that snow was never recorded, but heavy rain, humidity, wind, cloud intervals and sun were experienced. Based on these premises, around four-hundred-thousand spectra were captured with a spectral resolution below 8 nm at wavelength interval of 0.75 nm.

The data recovered for the module were obtained with a time interval of five minutes where their electrical parameters I_{SC}, V_{OC}, T_{MOD}, G and I–V curves plus meteorological data are captured simultaneously.

Although these measurements are very accurate the exact time when sun rises and sets is very difficult to implement in the measurement system, and therefore some of the spectra were captured in the dark at night. Therefore, the data was filtered to avoid reflection produced by angle of incidence and other effects. Spectra taken with an elevation angle under fifteen degrees are removed, following (Nann and Riordan, 1991).

Eq. (1) was used (Reda and Andreas, 2004) to calculate the topocentric elevation angle without atmospheric refraction correction -α- for a surface oriented in any direction.

$$\alpha = \arcsin(\sin\varphi \cdot \sin\delta' + \cos\varphi \cdot \cos\delta' \cdot \cos H') \quad (1)$$

where:

φ is the geocentric latitude
δ' is the topocentric sun declination
H' is the local hour angle for the sun transit

The total amount of data remaining after filtering is up to 70% of the initial measurements, a total amount of two-hundred-eighty-thousands spectra.

Some thin-film modules and the above described spectroradiometer are installed nearby in the same frame. Three commercial thin-film modules are used for this paper: an a-Si/μc-Si, a single