Environmental Modelling & Software 53 (2014) 163–172

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Modelling the distribution of solar spectral irradiance using data mining techniques

Rafael Moreno-Sáez [a,1], Llanos Mora-López [b,*]

[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, E.T.S.I. Informática, Universidad de Málaga, Campus de Teatinos, 29071 Malaga, Spain

ARTICLE INFO

Article history:
Received 30 May 2013
Received in revised form
27 November 2013
Accepted 9 December 2013
Available online 27 December 2013

Keywords:
Solar spectral distribution
K-means clustering
Data mining techniques
Statistical techniques
Average photon energy

ABSTRACT

A procedure for modelling the distribution of solar spectral irradiance is proposed. It uses both statistical and data mining techniques. As a result, it is possible to simulate solar spectral irradiance distribution using some astronomical parameters and the meteorological parameters solar irradiance, temperature and humidity. With these parameters, the average photon energy and the normalization factor, which characterise the solar spectra, are estimated. First, the Kolmogorov–Smirnov two-sample test is used to analyse and compare all measured spectra. The k-means data mining technique is subsequently used to cluster all measurements. We found that three clusters are enough to characterise all observed spectra. Finally, an artificial neural network and a multivariate linear regression are estimated to simulate the solar spectral distribution matching certain meteorological parameters. The results obtained show that over 99.98% of cumulative probability distribution functions of measured spectra are the same as simulated ones.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Classical methods that are used to characterise the solar spectral irradiance distributions are based on physically modelling the atmospheric processes and they require the use of several environmental factors that account for atmospheric conditions. Basically, two different physical methods have been proposed: atmospheric transmittance methods and radiative transfer methods. The first methods are simpler: the atmosphere is modelled using a one-layer medium in which scattering and absorption processes attenuate the extraterrestrial solar radiation (Leckner, 1978; Bird et al., 1982; de La Casinier et al., 1997); examples of such methods are SPCTRAL2 (see Bird and Riordan, 1986) and SMARTS2 (see Gueymard, 2001); for the respective applications of these methods see Jacovides et al. (2004) and Kaskaoutis and Kambezidis (2008). The second methods are the radiative transfer methods, which are more rigorous: they use several scattering and absorbing layers to account for the vertical atmospheric in-homogeneity (see Liou, 1980; Stamnes et al., 1988); examples of such methods are MODTRAN models (or its predecessor LOWTRAN) (see Anderson et al., 1993). There are also easier models that include the effects of atmospheric components, such as gaseous pollutants and aerosols (see Jacovides et al., 2000) and there are methods that propose the characterisation of the solar spectral irradiance using global and diffuse irradiance measurements and account for the modification of the spectrum under different atmospheric conditions (see Kaskaoutis et al., 2006).

In general, the goal of these methods is to obtain the best representation of the atmosphere using local geographic coordinates, several types of atmospheric measurements and different aerosol models, which use the aerosol optical thickness (usually corresponding to 500 nm) and the Angstrom turbidity coefficient as inputs (see Utrillas et al., 1998). These last two parameters are estimated from the global, direct and diffuse integrated irradiance values. However, these methods are not appropriate for some engineering applications due to the detailed required inputs and software. Several fields in which information about the solar spectral irradiance distribution is useful, and which lack such detailed measurements, are illumination engineering or solar thermal and photovoltaic applications; for example, in the photovoltaics field, new materials that are used for solar photovoltaic modules have a performance that depends on the solar spectral distribution of solar radiation (see, for example, Martín and Ruíz, 1999; Minemoto et al., 2009; Gottschalg et al., 2003; Piliougine et al., 2011; Myers, 2012). For this reason, several studies have been conducted in this field for tackling the problem of knowing

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

1364-8152/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.envsoft.2013.12.002

the solar spectral irradiation distribution. Statistical and data mining methods that do not require a detailed description of the atmospheric composition but that only use the most typical atmospheric measurements that are available, such as global solar irradiance, temperature and humidity, have been used. Previously obtained results so far allow us to know some aspects of the statistical relationship among the spectral distribution of solar radiation and different meteorological variables. Most of the proposed statistical methods are based on the use of one parameter that characterises, and hence describes, the solar spectral distribution: Fabero and Chenlo (1991) use the Spectral Factor, Poissant et al. (2003) propose the Mismatch Factor and Williams et al. (2003) use the Average Photon Energy (APE).

All of these studies use only classical statistical methods to address the analysis and characterisation of the solar spectrum. Recently, other approaches that are based on the use of data mining models have been proposed, for example, some of these models are used in Moreno-Sáez et al. (2013) for characterising the solar spectral irradiance distribution by using a few meteorological parameters. Data mining techniques are widely applied in different areas, especially when it is necessary to address a large amount of data; these techniques have proved to be very useful. They allow us to process data and to identify models and patterns. A combination of different techniques to characterize environmental parameters has also been proposed in previous papers, such as in modelling ground-level ozone using principal component and multiple regression analysis (Abdul-Wahab et al., 2004) and using principal component regression and artificial neural networks (Al-Alawi et al., 2006).

This study seeks to obtain a model that allows us to establish the solar spectral irradiance distribution by using only meteorological parameters, which are usually available for different applications. For developing the model, first we characterised the measured spectra: we deeply analysed the relationships among the parameter average photon energy (APE) and the solar spectral irradiance distribution, and we also analysed how many different solar spectral irradiance distributions can be found in all of the measured spectral curves and how these different curves can be related by means of APE to some meteorological parameter. For performing these analyses, we have used both statistical and data mining techniques, as were proposed in Moreno-Sáez et al. (2013). Once we found a one-to-one or biunivocal relationship among APE and solar spectral irradiance distribution curves, we analysed how to simulate these curves using frequently available meteorological parameters. For this scope, we have determined the relationships between these meteorological parameters and the parameters that characterise the spectra, namely the APE value and the total energy received in the range of the wavelengths used.

The following section describes the materials and methods that we propose to use for characterising and simulating solar spectral irradiance values. The third section describes the proposed methodology. The fourth section describes the data that was used for this paper. The obtained results are presented in the fifth section. Finally, the conclusions summarise the most relevant results that were obtained in this study.

2. Materials and methods

In this section, we describe the parameters and methods that we propose to use for characterising the solar irradiance spectral curves and for simulating these curves once we have found the significant parameters that characterise the spectra and their relationships with the available meteorological parameters. We propose to use a hybrid approach that is based on the use of both statistical and data mining methods. We have already used, in a previous study, two of the techniques that we propose now: the Kolmogorov–Smirnov two-sample test and the k-means clustering method; the results that were obtained in this previous work are similar to some of the results that we present here, but those are focused on how the solar spectrum affects the performance of thin-film photovoltaic modules (Moreno-Sáez et al., 2013).

2.1. Average photon energy

We propose to use the average photon energy (APE) parameter for characterising the solar spectral distribution. The APE value is an index that was proposed by Gottschalg's group (Loughborough University) (Williams et al., 2003). APE is defined as the average energy per photon included in the spectrum (Williams et al., 2003); it is calculated by dividing the integrated irradiance by the integrated photon flux density, according to the following expression:

$$\mathrm{APE} = \frac{\int_a^b E(\lambda)\mathrm{d}\lambda}{q\int_a^b \Phi(\lambda)\mathrm{d}\lambda}(\mathrm{eV}) \tag{1}$$

where $E(\lambda)$ is the energy at wavelength λ, $\Phi(\lambda)$ is the photon flux density at wavelength λ, and a and b are the considered wavelength boundaries. The APE value for the standard spectrum AM 1.5 is 1.88 eV in the wavelength range that we are working in, which is 350–1050 nm.

Minemoto et al. (2009) proved that the APE value can be used to statistically characterise the spectral irradiance using the methodology adopted by the International Electrotechnical Commission (IEC, 2007). They compare only spectra that have similar APE values by using the mean and standard deviation and conclude that an APE value uniquely yields the shape of a solar spectrum, although they remark that the uniqueness of APE was verified only for spectra that were collected in the measurement location. This uniqueness means that, with the APE value, it is possible to know the shape of the solar spectral irradiance, in other words, the relative irradiance that corresponds to each wavelength is determined by the value of the APE. We propose to use this index as one of the variables that explains the solar spectral irradiance distribution. To analyse the relationship between this variable and the different distributions recorded in Malaga, we have used both statistical and data mining techniques, which are described in the following sections.

2.2. Statistical test for comparing solar spectral distributions

We propose using a statistical test that addresses all of the measured data in the spectral curves and not only the wavelength interval bands of 50 or 100 nm, as in previous studies (see, for example, Minemoto et al., 2009). We have implemented the classical statistical Kolmogorov–Smirnov two-sample test to analyse the different solar spectral distributions measured, as explained in Mora and Mora-López (2010). We propose to check if two solar spectral irradiance distributions are the same. We are given the following:

$$\left\{X_{\lambda_i}\right\}_{\lambda_1=350}^{\lambda_n=1050} \text{ and } \left\{Y_{\lambda_i}\right\}_{\lambda_1=350}^{\lambda_n=1050}$$

which are the solar spectral irradiance values for the different wavelengths λ_i of both measurements. Denote the probability distribution function (p.d.f.) of X as $f_X(\cdot)$, which is estimated by using:

$$\widehat{f}_X(\lambda_j) \equiv \frac{E_{\lambda_i}}{E_t} \tag{2}$$

where:

$$E_t = \sum_{\lambda_i=350}^{1050} E_{\lambda_i} \tag{3}$$

Denote the cumulative probability distribution function (c.p.d.f.) of X as $F_X(\cdot)$ and the c.p.d.f. of Y as $F_Y(\cdot)$. Both $F_X(\cdot)$ and $F_Y(\cdot)$ are assumed to be continuous. Then, the null hypothesis that we propose is expressed as:

$$\mathrm{H}_0 : F_X(\cdot) = F_Y(\cdot), \tag{4}$$

versus the general alternative hypothesis

$$\mathrm{H}_a : F_X(\cdot) \neq F_Y(\cdot), \tag{5}$$

which makes no parametric assumption about the shape of these c.p.d.f.s. This procedure is known as the "test of homogeneity between two samples". If the sample sizes n and m are sufficiently large, then this test can be performed by using the Kolmogorov–Smirnov statistic, which compares the empirical c.p.d.f. obtained with each sample (measurement). The sample sizes for this experiment are the number of solar spectral irradiances measured (one for each wavelength recorded by the measurement equipment).

Specifically, if for a real number λ_j in the range [350.0 − 1050.0] (which corresponds to one of the wavelengths measured), we define:

Download English Version:

https://daneshyari.com/en/article/6963960

Download Persian Version:

https://daneshyari.com/article/6963960

Daneshyari.com