ARTICLE IN PRESS

Available online at www.sciencedirect.com

ScienceDirect

Solar Energy xxx (2016) xxx–xxx

SOLAR ENERGY

www.elsevier.com/locate/solener

Computations of diffuse fraction of global irradiance: Part 2 – Neural Networks

R. Tapakis [a], S. Michaelides [b,c], A.G. Charalambides [a,*]

[a] Cyprus University of Technology, Department of Environmental Science and Technology, Lemesos 3603, Cyprus

[b] Cyprus Department of Meteorology, Nicosia, Cyprus

[c] Currently at Cyprus University of Technology, Department of Civil Engineering and Geomatics, Lemesos, 3603, Cyprus and The Cyprus Institute, Nicosia, Cyprus

Received 14 April 2015; received in revised form 7 December 2015; accepted 15 December 2015

Communicated by: Associate Editor Frank Vignola

Abstract

Solar energy is the feedstock for various applications of renewable energy systems, thus, the necessity of calculating and using global tilted irradiance is acknowledged for the computations of the performance and monitoring of Photovoltaic (PV) Parks and other solar energy applications. Thus, the aim of our research is to develop a model for the correlation of diffuse fraction (k_d) and the clearness index (k_t), that can then be used for the evaluation of the diffuse irradiance given the global irradiance.

In a companion paper, existing simple empirical models were reviewed and compared based on 10 years of data from Cyprus and then, analytical approaches for the computation of diffuse fraction were employed, where solar altitude was introduced as an additional parameter in the calculations.

In the present paper, the same dataset was used, and three additional parameters were introduced to the calculations: global irradiance on the horizontal plane, extraterrestrial irradiance on the horizontal plane and the time of the day. These parameters were chosen due to the strong dependence of the diffuse fraction/clearness index correlation to the season/day of the year and time of day. Due to the non-linear influence of these parameters to the k_t–k_d correlations and the additional interaction between them, the employment of analytical methods is not applicable. Thus, non-parametric regression analysis was adopted, using supervised machine learning methodologies such as Artificial Neural Networks, which are able to learn the key information patterns from multivariate input.

Comparing the non-parametric regression to the analytical models developed in the companion paper, it is shown herewith that the accuracy of the models was slightly improved. The statistical indicators MBE, RMSE and R^2 of the best fit model were −4.69%, 21.54% and 0.90 respectively.

Keywords: Solar irradiance; Diffuse fraction; Clearness index; Solar altitude; Neural Networks; Time

1. Introduction

In Part 1 of this work (Tapakis et al., 2015), the necessity of adopting computational models for the calculation of diffuse irradiance at several locations where there is a lack of adequate measurements was presented. These models, usually called Liu–Jordan type are analytical decomposition models correlating diffuse fraction (k_d) to the clearness index (k_t) (Liu and Jordan, 1960). Additionally, in Part 1, existing simple empirical models were reviewed and compared and three analytical approaches for the computation of diffuse fraction were employed. The first was solely based on the measurements of k_t, where new

* Corresponding author at: Corner of Athinon and Anexartisias, 57, 3603 Lemesos, Cyprus. Tel.: +357 25 002306; fax: +357 25 002668.

E-mail address: a.charalambides@cut.ac.cy (A.G. Charalambides).

http://dx.doi.org/10.1016/j.solener.2015.12.042

analytical correlations were developed in a piecewise form. The second regarded the integration of solar altitude (α) into the correlations as an external parameter. The dataset was separated into independent groups according to α, and simple analytical correlations were developed for each sub-dataset. Finally, as a third approach, α was introduced into the computations as an additional parameter and a corresponding non-linear model was developed.

During the processing of the data, it was observed that there is a strong dependence of the k_t–k_d correlations to the season/day of the year and the time of day, since these parameters integrate the air mass traversed by the solar radiation to the computations.

Aiming to achieve better accuracy in the correlations and investigate the influence of the aforementioned parameters, three additional parameters were introduced to the calculations. These parameters were the extraterrestrial irradiance on horizontal plane (G_{oh}), the global irradiance on horizontal plane (G) and the TIME. TIME is defined as the time interval of the measurement from midday, where the time of the measurement was the mean time of the hourly measurements.

G_{oh} encompasses the influence of the day of the year to the computations and G is a quantified indicator of the cloudiness (Tapakis and Charalambides, 2013); these two parameters were indirectly used in the previous computations since they are used during the preprocessing of the data for the calculation of k_t. Knowing both the TIME and α enables the quantification of the influence of the seasons to the computations: during winter, α is at lower angles compared to the summer, thus, the angle of α at winter's noon is equivalent to the angle of α at the morning and evening hours during the summer. Thus, the combination of these two parameters quantifies the variation of k_t–k_d due to the time of the day and the day/season of the year. Overall, the parameters used in the current analysis are: (a) k_t, (b) α, (c) G_{oh}, (d) G and (e) TIME. It has to be mentioned that all five parameters were calculated during the preprocessing of the data (Tapakis et al., 2015), thus, no additional processing was required. However, given that the input to the model should be independent, G is not used in the model because it is conditionally independent if G_{oh} and k_t are used. Thus, if G was included, the neural network should become insensitive to it. Eventually, only four inputs were used in the network: (a) k_t, (b) α, (c) G_{oh} and (d) TIME, and all results for the rest of this paper refer only to these four parameters.

Due to the non-linear influence of these parameters to the k_t–k_d correlations (Soares et al., 2004) and the additional interaction between solar radiation and the terrestrial atmosphere, such as Rayleigh scattering, radiative absorption by ozone and water vapour, as well as aerosol extinction (Kambezidis and Adamopoulos, 1997), the employment of analytical methods was quite difficult. Thus, non-parametric regression analysis was adopted, using supervised machine learning methodologies such as Artificial Neural Networks (ANN), which are able to learn the key information patterns from multivariate input.

Supervised regression can be employed when a training sample is used to train the model; therefore, the model is subject to the specific dataset and adjusted according to its error. Usually, the procedure relies on the training of the algorithm based on the dataset and the testing of the developed model using an independent part of the dataset (which was not used during the training) to provide an independent measure of the network's performance. Additionally, a validation set is usually used to measure network generalisation and to halt training when generalisation stops improving. Corresponding research on diffuse fraction computation using different ANN architectures was performed by Elminir et al. (2007), Rehman and Mohandes (2012) and Soares et al. (2004). Comparative studies of the implementation of different ANN for solar irradiance computations (hourly and daily G) were performed by Tymvios et al. (2005) and Khatib et al. (2012).

The present paper is organised as follows: A brief summary of the dataset, the equipment used in the study, and the preprocessing procedure are described in Section 2. Section 3 presents the theory of the ANN used in the current study. Section 4 presents the statistical results of the performance, the characteristics of the best fit model and a comparison to alternative models with different characteristics. Finally, Section 5 summarises the major findings of the paper. All computations presented in this paper were performed using MATLAB.

2. Solar irradiance data – Measurements and pre-processing

The irradiance was recorded at the main Solar Radiometric Center of the Department of Meteorology of Cyprus, located at Athalassa, Cyprus (35°28′27″ North, 33°23′47″ East, height 165 m amsl). Hourly values of Global Horizontal (H_h), Diffuse (H_d) and Beam Irradiation (H_b) were recorded using Kipp & Zonen pyranometers (model CM21) and pyrheliometer (model CH1) over a thirteen year period, from January 2001 to December 2013. The CH1 pyrheliometer and the CM21 used for diffuse radiation measurements were positioned on a Kipp & Zonen two axis solar tracker where a shade ball was used to block the sun.

Each record of the measured data consisted of four values: the date/time of the measurement and the corresponding hourly values of H, H_d and H_b in kJ/m^2.

A pre-processing procedure was applied to each record. At first, the irradiation measurements were converted to irradiance (W/m^2) and G (W/m^2) was calculated. For each measurement, TIME was computed as the time interval of the measurement from midday, where the time of the measurement was the mean time of the hourly measurements. The values of TIME ranged from −6.5 (05:00–06:00) to +6.5 (19:00–20:00) in discrete hourly values. In the computations, TIME was computed based on local time (daylight saving time was corrected) and was referenced to local noon. Overall, since the measurements are based on hourly values of solar irradiance, the deviation of local to solar