Renewable and Sustainable Energy Reviews xx (xxxx) xxxx–xxxx

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

Estimation methods for global solar radiation: Case study evaluation of five different approaches in central Spain

R. Urraca[a,*], E. Martinez-de-Pison[a], A. Sanz-Garcia[b,c], J. Antonanzas[a], F. Antonanzas-Torres[a]

[a] EDMANS Group, Department of Mechanical Engineering, University of La Rioja, Logroño, Spain
[b] Division of Pharmaceutical Biosciences, University of Helsinki, Viikinkaari, 5 E, P.O. Box 56, 00014 Helsinki, Finland
[c] Institute of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University, 8-1 Kawada-cho, Shinjuku-ku, 162-8666 Tokyo, Japan

ARTICLE INFO

Keywords:
Estimation methods
Daily global irradiation
ERA-Interim
CM-SAF
XGBoost
Kriging

ABSTRACT

Solar radiation can be estimated by a variety of methods in an attempt to overcome the limitations of on-ground records. Novel methods are often appearing but these are rarely compared to others from a different approach. This study surveys the main types of estimation methods for daily Global Horizontal Irradiation (GHI), and then, one characteristic technique per group is selected, discarding possible hybrid approaches: a parametric model based on temperatures and precipitation (Antonanzas model), a statistical model (XGBoost), interpolated ground-based measurements (Ordinary Kriging (OK)), a satellite-based dataset (CM-SAF-SARAH), and a reanalysis dataset (ERA-Interim). The techniques are evaluated in relation to the seasonal variation, the clearness index and the spatial performance at 38 grounds stations in central Spain from 2001 to 2013.

Three different tiers of estimations were obtained being SARAH and OK the best performing methods overall. The SARAH dataset (MAE=1.10 ± 0.13 MJ/m^2, MBE=0.22 ± 0.36 MJ/m^2) generated estimates with the lowest spread, but led to a slight overestimation in low-altitude flat areas. The OK (MAE=1.10 ± 0.25 MJ/m^2, MBE=0.00 ± 0.31 MJ/m^2) outperformed SARAH in these flat areas (high density of stations), but at the expense of a higher variability. Alternatively, SARAH surpassed Ordinary Kriging (OK) when the distance to the closest station exceeded 20–30 km. The ERA-Interim reanalysis and the XGBoost were in the second tier of estimations, and the parametric model yielded the worst results overall. ERA-Interim exhibited a systematic overestimation. The locally trained Antonanzas and XGBoost struggled to model the atmospheric transmissivity, showing large positive errors in spring months and a small underestimation of clear-sky days. Finally, a summary with the strengths and weaknesses of the five methods provides a deeper understanding for the selection of the adequate estimation approach.

1. Introduction

An accurate knowledge of surface solar irradiation is of great interest for current society. It is an essential variable for different branches of knowledge such as energy industry, agriculture, climatology or environmental sciences. In energy production, solar energy industry has been growing since late 90s in response to environmental problems associated with fossil fuel technologies. At the end of 2014, the global power installed was 177 GW for photovoltaics, 406 GW for solar thermal collectors for heating and cooling, and 4.4 GW for concentrated solar power [1], and there is an expected positive trend in solar power installation for the upcoming years [2]. Hence, the

Abbreviation: ANN, Artificial Neural Network; BSRN, Baseline Surface Radiation Network; CLARA, CLoud, Albedo and RAdiation dataset; CM-SAF, Satellite Application Facility on Climate Monitoring; DNI, Direct Normal Irradiance; ECMWF, European Center for Medium-Range Weather Forecast; EUMETSAT, European Organisation for the Exploitation of Meteorological Satellites; GA, Genetic Algorithms; GBM, Gradient Boosting Machine; GEBA, Global Energy Balance Archive; GHI, Global Horizontal Irradiation; GMS, Geostationary Meteorological Satellite; GOES, Geostationary Operational Environmental Satellite; IDW, Inverse Distance Weighting; JMA, Japan Meteorological Agency; LM, Linear Regression; LOO, Leave-One-Out; LSA-SAF, Satellite Application Facility on Land Surface Analysis; LUT, Look-Up Table; MAD, Mean Absolute Deviation; MAE, Mean Absolute Error; MAGIC, Mesoscale Atmospheric Global Irradiance Code; MARS, Meteorological Archival and Retrieval System; MBD, Mean Bias Deviation; MBE, Mean Bias Error; MFG, Meteosat First Generation satellite; MSG, Meteosat Second Generation satellite; NASA, National Aeronautics and Space Administration; NOAA, National Oceanic and Atmospheric Administration; NWP, Numerical Weather Prediction; OK, Ordinary Kriging; RBF, Radial Basis Function; rMAE, relative Mean Absolute Error; rMBE, relative Mean Bias Error; RMSE, Root Mean Squared Error; rRMSE, relative Root Mean Squared Error; RS, Random Search; RTM, Radiative Transfer Model; SARAH, Surface Solar RAdiation data set - Heliosat; SIAR, Sistema de Informacion Agroclimatica para el Regadio; SIS, Surface Incoming Shortwave Irradiance; SMSE, Surface Meteorology and Solar Energy; SoDa, Solar Data; SSI, Surface Solar Irradiance; SSRD, Surface Solar Radiation Downwards; SVM, Support Vector Machine; UK, Universal Kriging; WRDC, World Radiation Data Center; XGBoost, eXtreme Gradient Boosting

* Corresponding author.
E-mail address: ruben.urraca@unirioja.es (R. Urraca).

http://dx.doi.org/10.1016/j.rser.2016.11.222
Received 23 March 2016; Received in revised form 2 September 2016; Accepted 18 November 2016
Available online xxxx
1364-0321/

demand for more reliable resource data to design new energy plants is steady increasing, as this information plays a key role in the selection of suitable locations and in capacity planning, among other issues. In agriculture, solar radiation is a key input to derive primary agricultural variables such as evapotranspiration and irrigation needs [3]. Besides, in climatology, long quality time series of radiation data are demanded to analyze trends in global climate [4].

The straightforward assessment of the solar resource is via ground measurements by pyranometers, which are the type of radiometer most frequently installed in meteorological stations. It records the total amount of shortwave (0.29–2.8 μm) solar irradiance (direct, diffuse and reflected) received by a horizontal surface at ground level. This is the variable typically used for solar resource studies that is herein named Global Horizontal Irradiation (GHI). However, Surface Solar Irradiance (SSI), downwelling surface solar flux or incoming solar radiation are also common denominations for the same term. If well-calibrated, pyranometer records are the most accurate estimates of solar radiation [5]. Some examples of world-wide networks of stations that provide monthly, daily, hourly and even minutely estimates of GHI are the Baseline Surface Radiation Network (BSRN), the World Radiation Data Center (WRDC), or the Global Energy Balance Archive (GEBA). Nevertheless, pyranometers are quite sensitive sensors that require a continuously maintenance and calibration to keep an optimal degree of accuracy, and besides, their high installation costs prevent them to be extensively installed in most developing countries. Therefore, obtaining high-quality pyranometer records with a reasonable spatial density is not always a simple task.

In the absence of ground records, different estimation methods have been developed to determine GHI. These can be not only for a specific location but also for generating spatially-continuous predictions, i.e., solar radiation maps. The initial approaches were based on the use of empirical *parametric models* at locations where other secondary but more commonly recorded variables were available. These models tried to correlate these secondary variables, mainly sunshine duration and temperatures, with GHI through analytic expressions [6]. These simple correlations evolved in late 90s to the more sophisticated *statistical models*, partly due to the advances in computer science. However, both parametric and statistical models are only able to generate discrete estimations in the locations where the secondary variables are recorded. The necessity of spatially-continuous estimates of solar irradiation boosted the search for other approaches. The simplest method is the use of *interpolation models* to generate radiation maps from point records of pyranometers, or even from the estimations of parametric and statistical models. More advanced methods attempt to model the underlying physics in the atmosphere to generate solar maps. *Satellite-based models* are based on the use of images of reflected radiation collected by radiometers in satellites to predict GHI. *Reanalysis models* also generate radiation maps by running Numerical Weather Prediction (NWP) models but with data from the past.

Most of these estimation methods are first validated with high-quality records from pyranometers, and then compared against previous models from the same group. New proposals are rarely compared with methods from a different approach, and most studies lack of a more general discussion of the pros and cons of the technique proposed. Although some recent works have compared different types of estimation methods [3,7], no article to date, to the author's knowledge, includes all existing types: *parametric*, *statistical*, *interpolation*, *satellite and reanalysis*. It is against this background that this study performs a comparison of available methods to estimate GHI, highlighting their strengths and weaknesses, and the situations were they are most suited to. To this end, one characteristic model from each type of estimation method is selected, evaluated against on-ground records (reference data) and then compared to the others. The combination of different estimation methods is not considered in order to analyze the quality of each approach alone. The area selected for the study is Castilla-La Mancha, a region located in the central part of Spain with high integration of photovoltaic and concentrated solar thermal power plants. The ground database is composed by daily GHI records of pyranometers at 38 locations with historical data from 2001 to 2013.

The remainder of the paper is organized as follows. In Section 2, the most representative techniques of each estimation method for GHI are described. In Section 3, both, on-ground records of GHI and estimations obtained from each selected technique are briefly explained. In Section 4, the calibration and evaluation process for each technique is established. In Section 5, results obtained with each technique are compared. Besides, the seasonal and spatial distributions of the residuals are discussed. Finally, in Section 6, main conclusions and final remarks for each estimation method are drawn.

2. Estimation methods of surface solar radiation

2.1. Models based on meteorological and geographical variables - parametric and statistical

Historically, GHI has been widely estimated by correlating with other more readily available variables, e.g., meteorological variables, geographical factors (altitude, latitude, longitude), physical factors (aerosols, concentration of gases) and other solar-related variable (extraterrestrial irradiance (I_{ex}) and solar angles) [6]. Regardless of the input variables used, two distinct group of models are differentiated: parametric models, where the correlation is expressed by means of analytic expressions with a finite number of parameters, and statistical models, which encompass a diverse group of generally non-linear techniques from fields of statistics, artificial intelligence and machine learning.

Parametric models, the most elementary estimation method, have been used since early 20^{th} century. Given the fact that I_{ex} can be obtained through sun geometry equations, nearly all models attempt to predict the atmospheric transmissivity, i.e., the relation between terrestrial and extraterrestrial irradiation. Hence, parametric models are grouped based on the secondary variable they use to predict this atmospheric transmissivity. Sunshine parametric models comprise one of the widest groups, and have been reported to yield the most accurate estimations among all [6]. All sunshine models came from the original *Ångström-Prescott equation* [8,9]. Another highly correlated variable with the atmospheric transmissivity is the cloud cover factor, manually measured in Oktas. This variable is used in the *Black model* [10] and the *Supit-Van Kappel model* [11]. Nevertheless, neither sunshine nor cloud cover are variables widely recorded at the meteorological stations, and besides, these time series contain multiple gaps due to the necessity of human intervention in the measurement process. Consequently, other models have been developed using more extensively measured variables, such as temperatures and precipitations. *Bristow-Campbell model* [12] and *Hargreaves model* [13] are two of the most well-known models based on temperatures, while the *de Jong-Stewart model* [14] and the *Liu-Scott model* [15] used precipitations. With the advent of new sensors and the spread of ground stations, additional variables have been included in parametric models. This is the case of wind speed, relative humidity and geographic factors [16]. For further information about parametric models and even novel hybrid approaches, the reader is referred to recent literature [6,16].

The main drawback of parametric models is that they need to be locally calibrated (in space and time) to obtain fair predictions, penalizing their generalization capacity. Besides, both the use of only one input variable and simple linear correlations limits the accuracy of the predictions obtained. Statistical models emerged in late 90s to overcome these issues, combining several input variables in more complex non-linear relationships. In solar estimation, different algorithms have been tested with this purpose: Artificial Neural Network (ANN) [17–20], Support Vector Machine (SVM) [21–23], regression trees [24] or fuzzy logic [25]. Besides, several hybrid techniques have