Renewable Energy 60 (2013) 604–614

Contents lists available at SciVerse ScienceDirect

Renewable Energy

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

Evaluation and improvement of empirical models of global solar irradiation: Case study northern Spain

F. Antonanzas-Torres [a,*], A. Sanz-Garcia [a], F.J. Martínez-de-Pisón [a], O. Perpiñán-Lamigueiro [b,c]

[a] EDMANS Group, Department of Mechanical Engineering, University of La Rioja, Logroño, Spain
[b] Electrical Engineering Department, EUITI-UPM, Ronda de Valencia 3, 28012 Madrid, Spain
[c] Instituto de Energía Solar, Ciudad Universitaria s/n, Madrid, Spain

ARTICLE INFO

Article history:
Received 21 January 2013
Accepted 8 June 2013
Available online 3 July 2013

Keywords:
Solar global irradiation
Empirical models
Time series
Evapotranspiration

ABSTRACT

This paper presents a new methodology to build parametric models to estimate global solar irradiation adjusted to specific on-site characteristics based on the evaluation of variable importance. Thus, those variables highly correlated to solar irradiation on a site are implemented in the model and therefore, different models might be proposed under different climates. This methodology is applied in a study case in La Rioja region (northern Spain). A new model is proposed and evaluated on stability and accuracy against a review of twenty-two already existing parametric models based on temperatures and rainfall in seventeen meteorological stations in La Rioja. The methodology of model evaluation is based on bootstrapping, which leads to achieve a high level of confidence in model calibration and validation from short time series (in this case five years, from 2007 to 2011).

The model proposed improves the estimates of the other twenty-two models with average mean absolute error (MAE) of 2.195 MJ/m^{2}day and average confidence interval width (95% C.I., $n = 100$) of 0.261 MJ/m^{2}day. 41.65% of the daily residuals in the case of SIAR and 20.12% in that of SOS Rioja fall within the uncertainty tolerance of the pyranometers of the two networks (10% and 5%, respectively). Relative differences between measured and estimated irradiation on an annual cumulative basis are below 4.82%. Thus, the proposed model might be useful to estimate annual sums of global solar irradiation, reaching insignificant differences between measurements from pyranometers.

1. Introduction

Solar irradiation research is a field of rising interest due to its many applications, such as the study of evapotranspiration [1] and optimization of water demand in irrigation, crop forecasting [2] from near-to-present measurements and estimates, the development and reduction of uncertainties in solar energy technologies (generation and internal rate of return) [3], the adjustment of energy policies to promote solar energies, and research on climate change [4]. The high cost of measuring solar irradiation with pyranometers and the scarcity of long, reliable datasets for specific locations has propitiated the progress in estimators such as the analysis of satellite images [4,5], artificial neural networks (ANN) [6,7] and empirically-based parametric models [8–10]; the latter estimating daily global horizontal irradiation (R_s) from other meteorological variables.

Satellite-based R_s estimates are only provided with high resolution for specific areas in the planet, for example, 70S-70N, 70W-70E in the Satellite Application Facility for Climate Monitoring (CM SAF) [11], Helioclim1 and Helioclim3 from SODA [12]. In other areas, resolution from satellite-based estimates is low, such as in some regions of South America and South-East Asia (INPE [13] and the National Renewable Energy Laboratory (NREL) [14] with 40 × 40 km resolution). The NASA Surface meteorology and Solar Energy (SSE) [15] coverage is global but resolution is very low (1 × 1°). Due to the effect of local microclimatic events on R_s, daily and annual divergence within a 40 × 40 km or 1° × 1° cell might be significant [16]. In addition, satellite-based daily estimates are not generally freely accesible in the near present. For instance, the SODA provides R_s from Helioclim1 for the period 1985–2005, Helioclim3 for the year 2005 and from the SSE database for the period 1983–2005. These near-to-present estimates are necessary in different applications such as the estimation of evapotranspiration of previous days to

* Corresponding author.
E-mail address: antonanzas.fernando@gmail.com (F. Antonanzas-Torres).

0960-1481/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.renene.2013.06.008

Nomenclature

BC	Bristow & Campbell model
ΔT	daily range of maximum and minimum temperatures
$\overline{\Delta T_c}$	average ΔT of the *calibration* dataset
ΔT_{i-1}	daily range of maximum and minimum temperatures on day $i-1$
ΔT_m	monthly average of ΔT
$\overline{\Delta T_t}$	average ΔT of the *testing* dataset
h	elevation above sea level
H	daily mean relative humidity
J	Julian day
M	logical variable of rainfall
$\mathrm{MAE_{tes}}$	mean absolute error of testing
$\mathrm{MAE_{val}}$	mean absolute error of validation
$\overline{\mathrm{MAE_{val}}}$	average $\mathrm{MAE_{val}}$ for the whole set of stations
n	length in days of the *validation* database
P	rainfall
P_c	yearly average rainfall in mm for the *calibration* dataset
P_t	yearly rainfall in mm for the *testing* dataset
$p_{\mathrm{sat}}[T_{\max}]$	vapor saturation pressure at $T_{\max}$
R^2	coefficient of determination
R_a	extraterrestrial irradiation
$R_{a,i-30}$	extraterrestrial irradiation on day $i-30$
R_s	daily global solar irradiation
$\overline{R_s}$	monthly mean of daily global irradiation
$\overline{R_{s,c}}$	average R_s for the *calibration* period
$R_{s,\mathrm{est}}$	daily estimated irradiation
$R_{s,\mathrm{meas}}$	daily measured irradiation
$\overline{R_{s,t}}$	average R_s for the *testing* period
$\overline{R_{\mathrm{MAE,val}}}$	average confidence interval width of MAE
$\overline{R_{\mathrm{RMSE,val}}}$	average confidence interval width of RMSE
$\overline{\mathrm{RMSE_{val}}}$	average $\mathrm{RMSE_{val}}$ for the whole set of stations
$\mathrm{RMSE_{tes}}$	root mean square error of testing
T_{avg}	daily average air temperature
$T_{\max}$	daily maximum temperature
$T_{\min}$	daily minimum temperature
θ	Julian angle
W	daily mean wind speed

forecast irrigation. As a result, the empirically-based parametric models stand out because of their high simplicity in estimating near-to-present R_s from measurements of commonly registered variables, generally registered with a higher distribution than the satellite resolution.

Refs. [17] and [18] developed the first parametric models to estimate R_s out of sunshine records and introduced the concept of the atmospheric transmittance that affects incoming extraterrestrial irradiation (R_a). The common figure of most parametric models is that they account for latitude, solar declination, the Julian day (J), and day length by including R_a [19]. Ref. [20] included mean daily cloud coverage to explain R_s. Ref. [21] introduced relative humidity and maximum temperature to estimate the monthly mean of the daily irradiation ($\overline{R_s}$). However, the scarcity of sunshine and cloud cover records limits the usage of these methods to the location of validation.

Refs. [9], [22], and [8] developed the first models in which R_s is estimated through the daily range of maximum and minimum temperatures (ΔT). Note that in these models ΔT behaves as an indicator of atmospheric transmittance, providing information about cloud cover. The higher emissivity of clouds than clear sky makes the maximum air temperature decrease and the minimum temperature increase, and as a result the ΔT decreases [23].

Refs. [24] studied the [9] model with $\overline{R_s}$, distinguishing between inland and coastal locations and obtaining higher accuracy in monthly than in daily estimates [25]. Other authors also modified the [9] model, introducing elevation [26], or modifying the square root by a Neperian logarithm [27] (the latter attributing it to [25]).

Rainfall (P) was introduced as an explanatory variable directly [10,28] or as a binary variable (M) equal to 1 in days with some rainfall (denoted as rainy days) and 0 in days without any rainfall recorded (non-rainy days) [29–31]. Refs. [30,31] rejected using ΔT in his model, considering P sufficient to explain R_s. Ref. [30] also rejected R_a and applied Fourier series based on the Julian angle (θ), corresponding to the angle in radians of the J.

Ref. [8] (hereinafter BC) calculated ΔT as the difference between the maximum temperature of the day and the average of the minimum temperatures of the current day and the following day. Ref. [32] modified the BC model, calculating ΔT related to rainfall. Ref. [19] studied the influence of ΔT on estimations, calculated as the difference between the maximum ($T_{\max}$) and minimum temperatures ($T_{\min}$) and as ΔT as per BC and evaluated it with sixteen BC and [9] derived models. Eventually, better estimations were achieved with ΔT as the difference between $T_{\max}$ and $T_{\min}$. The BC equation has also been modified by considering some parameters as constants [1,19,33,34]. The last of this papers attributed two new models to [33] and [35]. Additionally [33], concluded that [25] and BC models perform better for $\overline{R_s}$ than for daily values. Ref. [36] and latter [35] (who referred it as BC) included the monthly mean of the daily ΔT to smooth the results of the BC model. Ref. [36] also developed a model in which the daily average temperature was introduced. Refs. [37,38] also modified the BC model, introducing the R_a as a function of the atmospheric transmittance. Indeed, several papers have proved the efficacy of the BC model by comparing it with their own models or with other models, e.g. Refs. [1,19,23,28,29,32–35, 39–42].

Most of parametric models to estimate R_s have been derived from the [9] and the BC models by adding other variables that were proved to achieve better estimates where validated. However, a variable which might be correlated with R_s in a site, might not have such a dependency in other site [26]. This paper proposes the evaluation of variable importance as a method to adjust general models, i.e., the BC model. New models are then built by including important variables, obtained by on-site specific relationships between predictors and R_s.

Several papers have already evaluated models according to test errors, assessing the capacity of generalization under unproven data [23,35,39]. Nevertheless, models might generate low test errors for a specific time series while still being unstable under slight variations in the calibration data [43]. This paper also proposes an evaluation including stability and accuracy under different initial conditions as model selection criteria, and implements it on twenty-four parametric models (including two new models built on the method of evaluation of variable importance) in seventeen meteorological stations in La Rioja (Spain). The estimates of the best performing model are also compared with the CMSAF SIS satellite-derived database.

Table 1 summarizes the twenty-four models studied.

2. Meteorological data

The assessment is performed in La Rioja, a 5028 km^2 region of Spain with significant climatic differences mainly due to differences