Agricultural and Forest Meteorology 310 (2021) 108627

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Modelling Photosynthetic Active Radiation (PAR) through meteorological indices under all sky conditions

A. García-Rodríguez, D. Granados-López, S. García-Rodríguez, M. Díez-Mediavilla, C. Alonso-Tristán *

Research Group Solar and Wind Feasibility Technologies, SWIFT, Electromechanical Engineering Department, Avda, Cantabria s/n, 09006, Burgos, Spain

ARTICLE INFO

Keywords:
PAR
Modelling
CIE standard sky classification

ABSTRACT

In this study, ten-minute meteorological data-sets recorded at Burgos, Spain, are used to develop models of Photosynthetic Active Radiation (*PAR*) following two different procedures: multilinear regression and Artificial Neural Networks. Ten Meteorological Indices (MIs) are chosen as inputs to the models: clearness index (k_t), diffuse fraction (k_d), direct fraction (k_b), Perez's clear sky index (ε), brightness index (Δ), cloud cover (*CC*), air temperature (*T*), pressure (*P*), solar azimuth cosine (*cosZ*), and horizontal global irradiation (*RaGH*). The experimental data are clustered according to the sky conditions, following the CIE standard sky classification. A previous feature selection procedure established the most adequate MIs for modelling *PAR* in clear, partial and overcast sky conditions. *RaGH* was the common MI used by all models and for all sky conditions. Additional variables were also included: the geometrical parameter, *cosZ*, and three variables related to the sky conditions, k_t, ε, and Δ. Both modelling methods, multilinear regression and ANN, yielded very high determination coefficients (R^2) with very close results in the models for each of the different sky conditions. Slight improvements can be observed in the ANN models. The results underline the equivalence of multilinear regression models and ANN models of *PAR* following previous feature selection procedures.

1. Introduction

Photosynthetic Active Radiation, *PAR*, is the region of the solar spectrum between 400 and 700 nm (Alados et al., 2000) used by plants in the photosynthesis process. *PAR* knowledge can provide key inputs for modelling biomass and forestry production (Aguiar et al., 2012; Landsberg and Waring, 1997), plant growth (Liu et al., 2016), natural greenhouse illumination (Alados et al., 1996) and the calculation of the euphotic depth of the oceans (Kirk, 1979), among others. Usually, Photosynthetic Photon Flux Density, Q_p ($\mu mol{\cdot}s^{-1}{\cdot}m^{-2}$), is measured and converted into energy units using the McCree's conversion factor ($4.57\ \mu mol{\cdot}J^{-1} \pm\ 3\%$) depending on climatic factors (Akitsu et al., 2015)). Spatial distribution and long-term trends of *PAR* have been reported in many early studies developed in Europe (Ferrera-Cobos et al., 2020; Jacovides et al., 2003; Jacovides et al., 2007; Leuchner et al., 2011), China (Hu et al., 2018; Hu et al., 2010; Liang and Xia, 2005; Wang et al., 2015), United States (Yu et al., 2015) or Africa (Finch et al., 2004).

Alternative procedures have been developed for calculating*PAR*, due to the scarcity of *PAR* data from direct measurements at ground meteorological stations. These procedures include other meteorological and climatic variables, mainly horizontal solar global irradiation (*RaGH*). Many studies have established the *PAR*/*RaGH* ratio at between 0.45 and 0.50 (Meek et al., 1984; Monteith, 1973; Moon, 1940; Stanhill and Fuchs, 1977; Szeicz, 1974; Tsubo and Walker, 2005), depending on the location, solar elevation or atmospheric aerosols, and water vapor concentrations. Empirical models based on linear regressions have been developed for *PAR* estimates. Turbidity, cloud cover, atmospheric water and aerosols content, clearness and sky brightness, diffuse fraction, dew point temperature, and solar zenith angle have, among others, been used as data input for multilinear models around the world (Al-Shooshan, 1997; Alados et al., 1996; Bat-Oyun et al., 2012; González and Calbó, 2002; Janjai et al., 2015; Peng et al., 2015; Wang et al., 2016; Yu et al., 2015; Zhang et al., 2000). However, the resulting models are strongly location dependent. Their use at other locations is not immediate, and requires new experimental studies (Al-Shooshan, 1997; Alados and Alados-Arboledas, 1999; Ferrera-Cobos et al., 2020; Nwokolo and Amadi, 2018). Satellite data have also been used for

* Corresponding author.
E-mail address: catristan@ubu.es (C. Alonso-Tristán).

https://doi.org/10.1016/j.agrformet.2021.108627
Received 3 May 2021; Received in revised form 16 July 2021; Accepted 21 August 2021
Available online 2 September 2021

Fig. 1. Experimental facility on the roof of the Higher Polytechnic School of Burgos University, Spain.

Table 1
Technical specifications of pyranometers and pyrheliometer.

Model	SR11	DR01	ML-020P
Measurement Range	0 to 3000 (W/m^2)	0 to 4000 (W/m^2)	0–3000 ($\mu mol/s{\cdot}m^2$)
Calibration uncertainty	< 1.8 % (k = 2)	< 1.2 % (k = 2)	< 2 % (k = 2)
Spectral Range	285 to 3000 × 10^{-9} m	200 to 4000 × 10^{-9} m	400 to 700 nm
Sensitivity (nominal)	15 × 10^{-6} V/(W/m^2)	10 × 10^{-6}V/(W/m^2)	0.15 × 10^{-6} $V/(\mu mol/s{\cdot}m^2)$
Operating temperature range	−40°C to 80°C	−40°C to 80°C	−10°C to 50°C
Temperature response	<± 2% (-10°C to 40°C)	< ± 1 % (-10°C to 40°C)	±1.1% (-10°C to 50°C)

Table 2
Sky Camera Technical specifications.

Model	SONA201-D
Sensor	CMOS-2.3MP
Vision Angle	<180° (fisheye lens)
Operating temperature	−40°C to 55°C
Image format	RAW

estimating *PAR* (Gao et al., 2011; Janjai and Wattan, 2011; Li et al., 2015; Vindel et al., 2018), although those models need to be validated from ground measured *PAR* data.

Over recent years, machine learning algorithms (ML) have been developed as a useful tool for modelling meteorological and climatic data (Huntingford et al., 2019; Torres et al., 2011). Artificial Neural Networks (ANN) have been used for modelling *PAR*, with different meteorological variables as input. Horizontal global solar irradiance, solar azimuth angle, clearness index and different parameters to take into account the atmospheric water vapor content such as the vapour content, the wet bulb temperature or the dew point temperature, are the most common inputs for developing an ANN for *PAR* estimations (Foyo-Moreno et al., 2017; Jacovides et al., 2015; Yu and Guo, 2016).

The sky conditions have a relatively low influence on the *PAR* /*RaGH* ratio. Escobedo *et al.* (Escobedo et al., 2009) concluded that the sky conditions had no influence on *PAR* levels within a variation of ± 8%. This conclusion might appear to contradict the definition of the *PAR* models in relation to sky conditions (Alados et al., 2000; Bosch et al., 2009; Dye, 2004; Serrano and Boscà, 2011). Most studies have determined that the *PAR*/*RaGH* ratio presents the highest values for cloud-covered skies and the lowest for clear skies (Blackburn and Proctor, 1983; García-Rodríguez et al., 2020; Stigter and Musabilha, 1982; Yamashita and Yoshimura, 2018) where the definition of the sky type is based on different criteria in each of the reviewed works (García-Rodríguez et al., 2020).

Independently of the *PAR* modelling procedure, *RaGH* has proven to be sufficient in itself as a factor to obtain accurate *PAR* data (±5% accuracy), with no consideration given to sky conditions (Ferrera-Cobos et al., 2020) using different time intervals (hourly, daily, monthly basis). However, the inclusion of other meteorological variables in the model improved the estimations of *PAR*. The objective of this work is the establishment of *PAR* models based on meteorological variables (or meteorological indices (MIs)) that are easily accessible at ground meteorological facilities. Its main novelties include the influence of sky conditions on the models, paying attention to the CIE standard sky classification (ISO, 2004), and using data collected at ten-minute intervals.

The following MIs are included in the study: clearness index (k_t), diffuse fraction (k_d), direct fraction (k_b), Perez's clearness sky index (ε), brightness index (Δ), cloud cover (*CC*), air temperature (*T*), pressure (*P*), solar azimuth cosine (*cosZ*), and horizontal global irradiation (*RaGH*). First, a feature selection procedure was applied, to identify related features and remove the irrelevant or less important ones. After the feature selection procedure, two different strategies were used for modelling *PAR*: multilinear regression and ANN. Analysis and comparisons of both models were conducted to study the influence of sky conditions on the accuracy of the model. The experimental data for this study were collected in an experimental campaign that ran from April 2019, to February 2021, in Burgos, Spain.

The paper will be structured as follows: after the Introduction Section, the experimental facility and the measurement campaign as well the quality filters applied to the experimental data will be described in Section 2. This Section will also include the definition and description of the meteorological indices used for modelling *PAR*. The CIE standard sky classification in Burgos, Spain, during the experimental campaign will be introduced in Section 3, as well the statistical analysis of the experimental data. In Section 4, the feature selection algorithm will be described. In Sections 5 and 6, the *PAR* modelling will be introduced using multilinear regression models and ANNs, respectively. Finally, the main results and some conclusions will be advanced in Section 7.

2. Experimental section

The ground meteorological facility where the experimental data used in this study were recorded, is located on the flat roof of the Higher Polytechnic School building (EPS) of Burgos University, Spain (42°21′04″ N, 3°41′20″ W, 856 m above mean sea level) and it is shown in Fig. 1. A complete description of the meteorological facility can be found in previous papers (García-Rodríguez et al., 2020; Granados-López et al., 2020; Granados-López et al., 2021; Suárez-García et al., 2020).

The following parameters are required for calculation of the ten MIs detailed in this study: solar horizontal global irradiance, *RaGH*; horizontal diffuse irradiance, *RaDH*; beam irradiance, *RaB*; cloud cover, *CC*; *PAR* data; air temperature, *T*; pressure, *P*; and the sky conditions under CIE standard. *RaGH*, *RaDH* and *RaB* were measured using a Hukseflux pyranometers, model SR11, and a Hukseflux pyrheliometer, model DR01, respectively (Hulseflux, Delft, The Netherlands). The beam irradiance sensor was installed on a sun tracker, model Sun-Tracker 3000, from Geónica (Geónica, Madrid, Spain). The diffuse irradiance sensor was protected from direct sunlight by a shadow hat. An ML-020P quantum sensor was used to measure Q_p. Cloud cover, *CC*, was

Table 3
Sky scanner technical specifications.

Model	MS-321LR Sky Scanner
FOV	11°
Luminance	0 to 50 kcd/m^2
Radiance	0 to 300 W/m^2
A/D Convertor	16 bits
Calibration Error	2%

measured by a sky camera model SONA201D (Sieltec Canary Islands, Spain) and the sky luminance and irradiance distribution, needed to establish the CIE standard sky conditions, was determined with a commercial MS-321LR sky scanner manufactured by EKO instruments (EKO Instruments Europe B.V., Den Haag, The Netherlands). Technical specifications of all instruments are shown in Tables 1–3

All meteorological and radiometric data were recorded every five minutes (averages from 30 seconds). Q_p data ($\mu mol{\cdot}s^{-1}{\cdot}m^{-2}$) were converted into *PAR* data ($W{\cdot}m^{-2}$) using McCree's conversion factor (4.57 $\mu mol{\cdot}J^{-1}$). The sky scanner was adjusted on a monthly basis for taking measurements from sunrise to sunset. It completed a full scan in four minutes and started a new scan every 10 min. The first and last measurements of the day were discarded, as well as measurements higher than 50 $kcd{\cdot}m^{-2}$ and lower than 0.1 $kcd{\cdot}m^{-2}$, following the specifications of the equipment. Only simultaneous data-sets could be used. Experimental data were analyzed and then filtered using conventional quality criteria (Gueymard and Ruiz-Arias, 2016). If a data-set failed to pass the quality criteria, then all the simultaneous data-sets were rejected. The original data-set counted 71600 datum, 36% of which were eliminated after the filtering procedure.

Five of the meteorological indices (MIs) used in this study were directly obtained from the experimental measurements: cloud cover (*CC*), from the sky camera; air temperature and pressure (T, P); horizontal global irradiation (*RaGH*); and *PAR*, measured by the pyranometers and the quantum sensor, respectively. Solar azimuth cosine (*cosZ*) was calculated from the geometrical data of the location, using well-known mathematical relationships Iqbal, 1983). The other MIs -clearness index (Iqbal, 1983), k_t, horizontal difuse fraction (Erbs et al., 1982), k_d; beam fraction (Suárez-García et al., 2020), k_b; and Perez's clearness index, ε; and brightness factor, Δ (Perez et al., 1990) were calculated using Eqs. (1) to ((6):

$$k_t = \frac{RaGH}{B_{sc}{\cdot}\varepsilon_0{\cdot}cosZ_s} \tag{1}$$

where, B_{sc} is the extraterrestrial irradiance constant (1361.1 W/m^2(Gueymard, 2018)); Z_s is the angle between sky zenith, and sun, and ε_0 is the average value of the orbital eccentricity of the Earth, calculated from eq. 2:

$$\varepsilon_0 = 1 + 0.033{\cdot}\cos\left(2{\cdot}\pi{\cdot}\frac{d_n}{365}\right) \tag{2}$$

where d_n is the day of the year.

$$k_d = \frac{RaDH}{RaGH} \tag{3}$$

$$k_b = \frac{RaB{\cdot}sen\gamma_s}{RaGH} \tag{4}$$

where γ_s is the solar elevation;

$$\varepsilon = \frac{\frac{RaDH+RaB}{RaDH} + k{\cdot}Z_s^3}{1 + k{\cdot}Z_s^3} \tag{5}$$

where k is 1.04 for Z_s expressed in rad (or $5.56{\cdot}10^{-6}$ if Z_s is expressed in degrees); and

$$\Delta = \frac{m{\cdot}RaDH}{B_{sc}{\cdot}\varepsilon_0} \tag{6}$$

where m is the optical air mass calculated using the Kasten model (Kasten, 1993).

3. CIE standard sky classification in Burgos

Following the analysis, great variability was observed in the experimental data, due to the different environmental and atmospheric conditions during the extensive experimental campaign. In many studies, the CIE standard sky classification for homogeneous skies has been considered an adequate representation of empirical sky conditions (Alshaibani, 2011; Li and Cheung, 2006; Li et al., 2008; Torres et al., 2010a; Torres et al., 2010b; Tregenza, 2004). Sky types of the same category have the same well-defined sky luminance pattern. The general formula for defining the relative pattern of luminance for any sky type is a combination of a gradation function, dependent on two parameters,

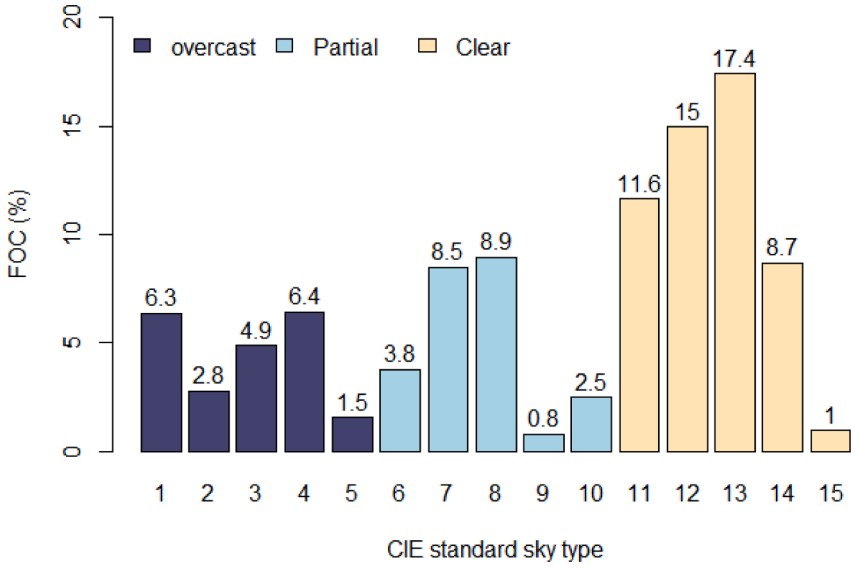

Fig. 2. Frequency of occurrence (FOC, %) of CIE standard sky types in Burgos, Spain, between April 2019 and February 2021.

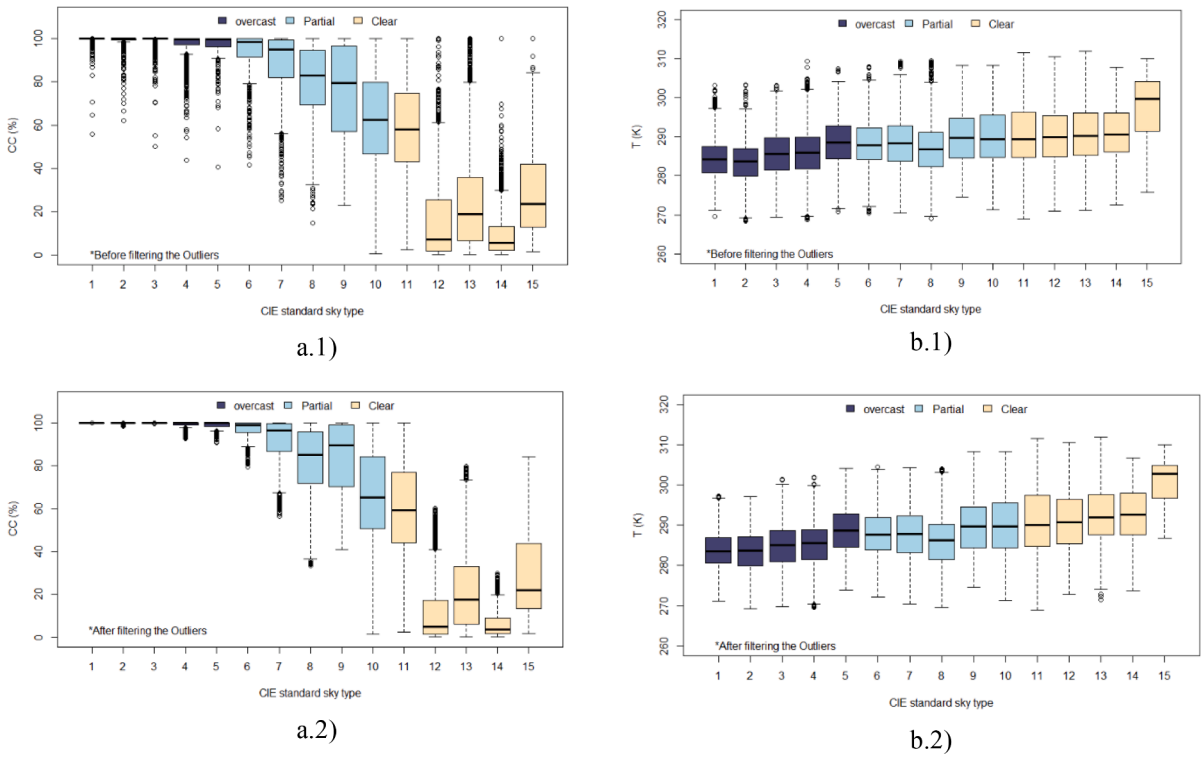

Fig. 3. Statistical analysis of air temperature, T (a), and cloud cover, CC (b), vs. CIE standard sky type: before (1) and after (2) filtering the outliers.

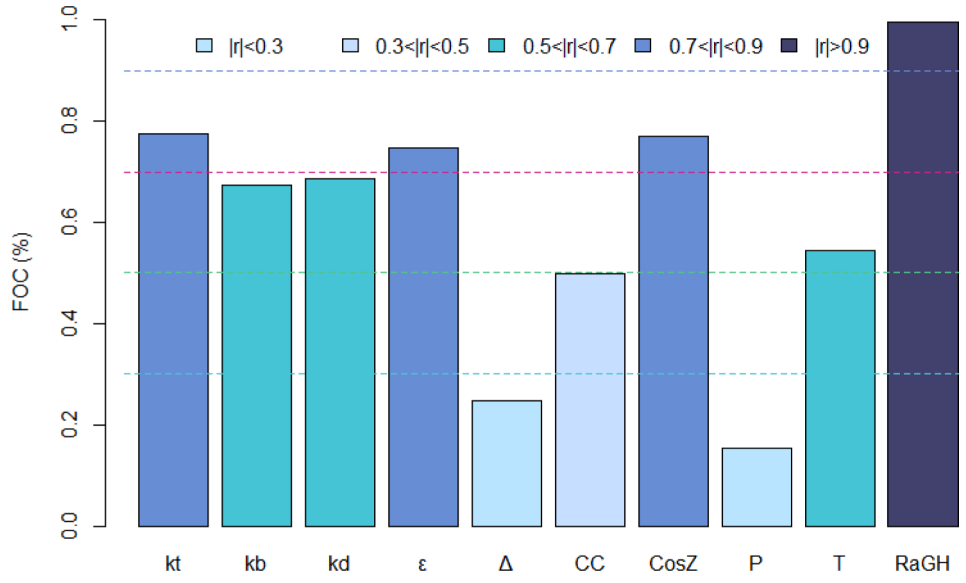

Fig. 4. The Pearson Correlation Coefficient ($r(PAR,\ MI_i)$) calculated for the 10 MIs used in this study.

Table 4
The Pearson correlation coefficient or Pearson's $r(PAR,\ MI_i)$ taking into account the sky conditions defined by the CIE standard sky classification (overcast, partial and clear).

	$\lvert r(PAR,\ MI_i)\rvert$				
CIE sky type	**[1-0.9]**	**(0.9-0.7]**	**(0.7-0.5]**	**(0.5-0.3]**	**(0.3,0]**
Overcast	*RaGH*	$k_t,\ \Delta$	*cosZ*	$k_b,\ \varepsilon$	$P,\ T$
Partial	*RaGH*	*cosZ*	$k_t,\ \varepsilon$	k_b, T	$CC,\ P$
Clear	*RaGH*, *cosZ*	ε	k_t	$k_b,\ T$	

and the indicatrix function, which considers the scatter of luminance with regard to the direction of sunrays, modelled as a function of three adjustable parameters. The gradation function modifies the luminance value between the horizon and the local zenith, assigning the highest luminance value to the zenith with cloudy skies and in reverse to clear skies. The indicatrix function shows the dispersion in the atmosphere of sunlight. The maximum luminance appears near the solar position, decreasing rapidly with the distance to the sun. Each of the functions takes six different forms and the combination yields 36 sky types from which 15 were selected: five overcast, five partly cloudy, and five clear sky types. Once the sky types are identified, the basic solar irradiance and daylight illuminance on the surfaces of interest can be obtained

Table 5
Multilinear regression models of *PAR*.

Sky conditions	Multilinear regression model	R^2	nRMSE (%)	nMAE (%)	nMBE (%)
Clear skies (MLR1)	$PAR = 0.3806 \cdot RaGH + 0.524 \cdot \varepsilon + 33.247 \cdot cosZ - 2.646$	0.991	3.93	2.72	$-2 \cdot 10^{-6}$
Partial skies (MLR2)	$PAR = 0.3958 \cdot RaGH + 18.282 \cdot cosZ - 2.508$	0.976	7.81	5.04	$-2 \cdot 10^{-6}$
Overcast skies (MLR3)	$PAR = 0.4335 \cdot RaGH - 7.726 \cdot k_t - 9.078 \cdot \Delta + 4.065$	0.985	6.62	4.44	$-5 \cdot 10^{-6}$

through simple mathematical expressions (Li et al., 2013). Therefore, the CIE standard sky taxonomy was selected for atmospheric conditions reference. A complete description of the CIE standard sky classification and the procedure to obtain the CIE standard sky classification from the sky scanner measurements can be found in previous works (Granados-López et al., 2021; Suárez-García et al., 2020; Suárez-García et al., 2018). Fig. 2 shows the Frequency Of Occurrency (FOC) of the different CIE standard sky types in Burgos during the experimental campaign, which extended from April, 2019 to February, 2021. As can be seen, in Burgos, the clear sky conditions were prevalent, as sky types 11 to 15 were the most frequent over the period under study. Sky type 13 (Cloudless polluted with a broader solar corona) had the highest FOC (17.4%) followed by sky types 7 (Partly cloudy with a uniform gradation and a brighter circumsolar effect) and 8 (Partly cloudy, rather uniform with a clear solar corona).

Once the reference sky conditions had been established, The MIs included in the study were analyzed from the sky conditions to identify the correlation between the MIs and the CIE standard sky conditions. The box-plot of the MIs, with respect to the CIE standard sky type, gave the following conclusions: air temperature, pressure, and *cosZ* were not influenced by the sky type, while *CC*, k_t, k_d, k_b, ε, and Δ were greatly affected by the sky conditions. A new quality test of the experimental data was introduced, discarding any simultaneous data corresponding to outliers in the box-plots. Fig. 3 shows the results for *CC* and *T*, where the influence of the sky type on the MI value is almost negligible in the case of *T* and very important in the case of *CC*. The discarded data reached 10%.

4. Feature selection

Feature Selection (FS) is the identification of related features within a set of data and the removal of irrelevant or less important features that contribute little or nothing to the definition of the target variable, so as to achieve models of greater accuracy. FS is one of the core concepts of ML that will impact on the performance of the developed model, improving its precision and reducing its complexity and overfitting as well as its runtime.

In this work, the Pearson criterion was used to analyze the degree of correlation of each of the selected MIs to *PAR*. The Pearson criterion is based on the Pearson correlation coefficient, *r*. For the case under study, if *PAR* and one selected MI are strongly correlated, the Pearson coefficient is 1 (direct correlation) or -1 (inverse correlation). However, a Pearson coefficient near 0 implies a weak or null correlation. The Thumb rule (Mukaka, 2012) established five r intervals for the correlation: direct ($1 \geq |r(PAR,\ MI_i)| \geq 0.9$), strong ($0.9 > |r(PAR,\ MI_i)| \geq 0.7$), moderate ($0.7 > |r(PAR,\ MI_i)| \geq 0.5$), weak ($0.5 > |r\ (PAR,\ MI_i)| \geq 0.3$), and negligible ($|r(PAR,\ MI_i)| < 0.3$).

Fig. 4 shows the Pearson coefficient, $r(PAR,\ MI_i)$, calculated for the 10 MIs used in this study, regardless of the sky conditions. A direct correlation was obtained between *PAR* and *RaGH* and a strong correlation between *PAR* and ε, k_t, and *cosZ*. The *r* coefficient with the other MIs was moderate, weak or negligible, so, these MIs were discarded as inputs for modelling *PAR*.

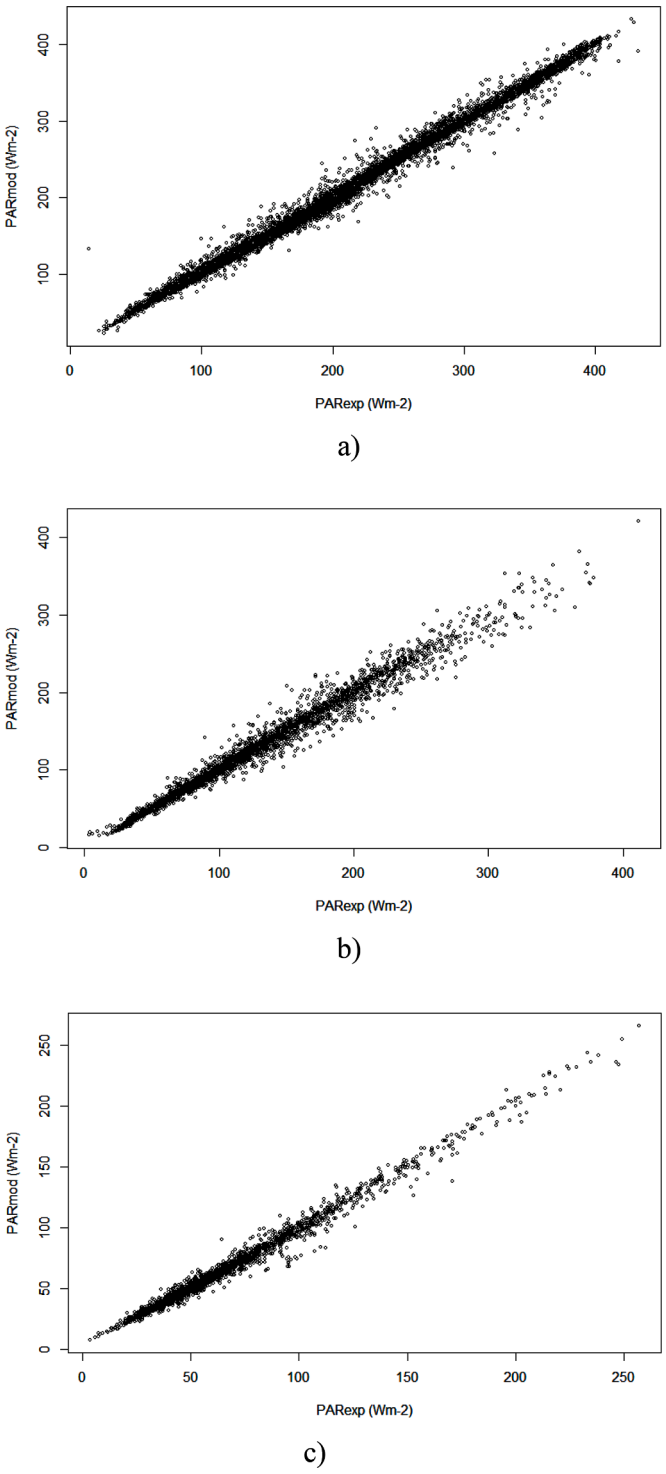

Fig. 5. Correlogram of the multilinear regression model of *PAR*: a) Clear skies; b) Partial skies; c) Overcast skies.

As Fig. 4 highlights, *RaGH*, the geometric parameter solar azimuth cosine, *cosZ*, Perez's clearness index, ε, and the clearness index, k_t, showed the strongest correlation to *PAR* ($|r(PAR,\ MI_i)| > 0.7$). These results agree with the literature, in that *RaGH* (or alternatively k_t) is the most widely used parameter for modelling *PAR* (Nwokolo and Amadi, 2018). Air pressure, *P*, the brightness factor, Δ, and cloud cover, *CC*, presented weak correlations with *PAR*.

The data were clustered, taking into account the CIE standard sky

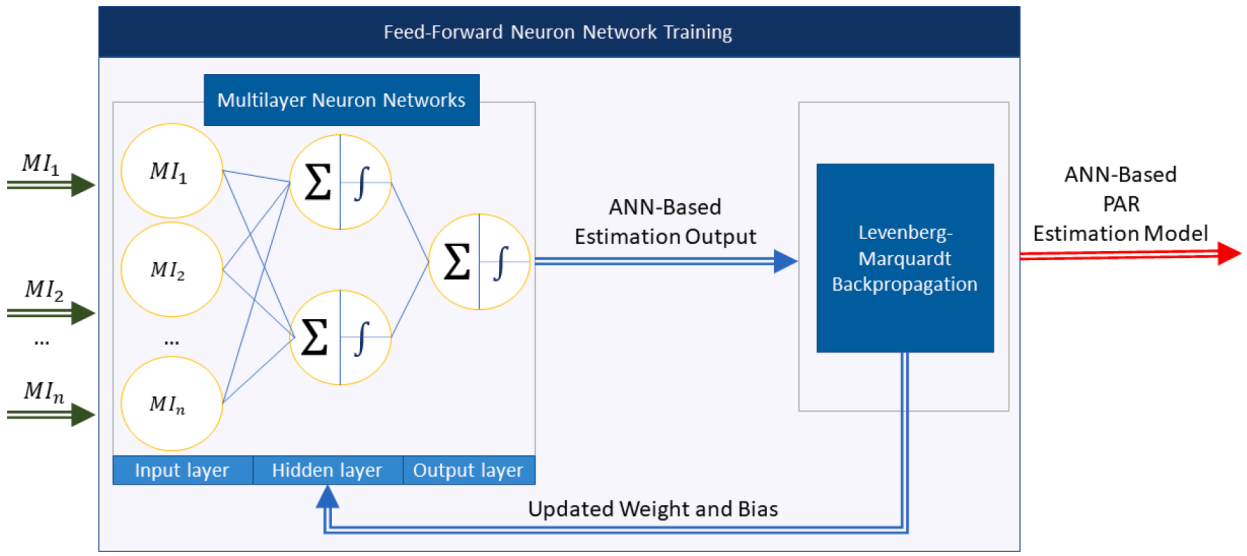

Fig. 6. ANN system architecture using the Levenberg-Marquardt Backpropagation algorithm.

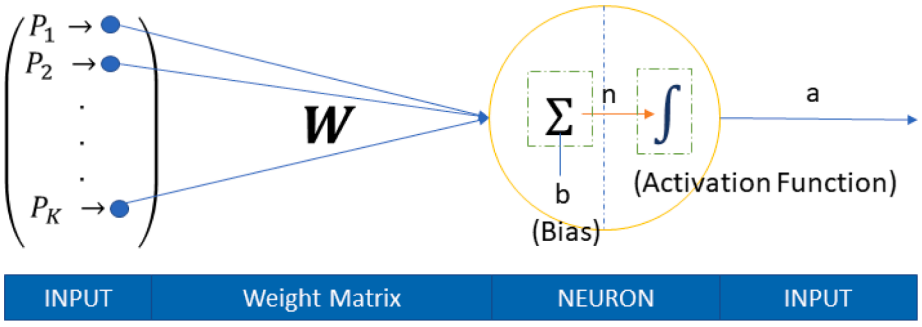

Fig. 7. Scheme of a neuron with k inputs. The weighting matrix and the bias are used for calculating the neuron output thought the activation function.

Table 6
Statistical results of the ANN models.

Sky conditions	R^2	*nRMSE* (%)	*nMAE* (%)	*nMBE* (%)
Clear skies (ANN1)	0.992	3.846	2.59	$-2{\cdot}10^{-6}$
Partial skies (ANN2)	0.977	7.795	5.00	$-3{\cdot}10^{-6}$
Overcast skies (ANN3)	0.987	6.466	4.34	$-5{\cdot}10^{-6}$

classification for clear sky (CIE sky types 1 to 5), and partial (CIE sky types 6 to 10) and overcast sky conditions. The Pearson correlation coefficient or Pearson's $r(PAR,\ MI_i)$ was calculated taking into account the sky classification and the results are shown in Table 4.

RaGH was the MI with the strongest correlation to *PAR*, regardless of sky conditions. *cosZ* only showed a very strong correlation to *PAR* for clear skies. MIs with strong correlations to *PAR* were ε for clear skies, *cosZ*, for partial skies, and k_t, and Δ in overcast sky conditions. The brightness factor, Δ, that presented weak correlations to *PAR* when data were not clustered, highlighted its strong correlation in overcast sky conditions. In the following sections, the *PAR* is modelled by following different procedures for each of the CIE standard sky conditions that were developed using the MIs with the Pearson correlation coefficient $|r(PAR,\ MI_i)|>0.7$.

5. Multilinear regression models

Three multilinear regression models, one for each sky type, were formulated to estimate *PAR*, based on the previous feature selection process shown in Table 4. The experimental data-set, taking into account the CIE standard classification, was randomly divided into two groups: the first one, comprising 85% of the data, was used to fit the models. The other group, with the remaining 15% of the data, was used to validate the models. The statistics used for this validation were as follows: the corresponding determination coefficient (R^2), the normalized mean bias error (*nMBE*), the normalized mean absolute error, *nMAE*, and the root mean square error (*nRMSE*). The expressions for these last three statistical terms are:

$$nMBE = \frac{1}{\overline{PAR}_{exp}} \frac{\sum_{i=1}^{n} \left(PAR_{mod} - PAR_{exp}\right)}{n} \times 100\ (\%) \tag{7}$$

$$nRMSE = \frac{1}{\overline{PAR}_{exp}} \sqrt{\frac{\sum_{i=1}^{n} \left(PAR_{mod} - PAR_{exp}\right)^2}{n}} \times 100\ (\%) \tag{8}$$

$$nMAE = \frac{1}{\overline{PAR}_{exp}} \frac{\sum_{i=1}^{n} \left|PAR_{mod} - PAR_{exp}\right|}{n} \times 100\ (\%) \tag{9}$$

where n is the number of the experimental data used for fitting the models; PAR_{mod} are the modelled values of *PAR*; and PAR_{exp} is the experimental value of *PAR*. As Table 5 shows, the results of fitting the models presented good correlation with the experimental data, given that $R^2>0.97$ and *nRMSE* was lower than 8% and *nMAE* lower than 5%. The small and negative values of *nMBE* indicate that the models present a good fit although tend to underestimate *PAR* values. The graph in Fig. 5 shows the multilinear regression models for each of the sky

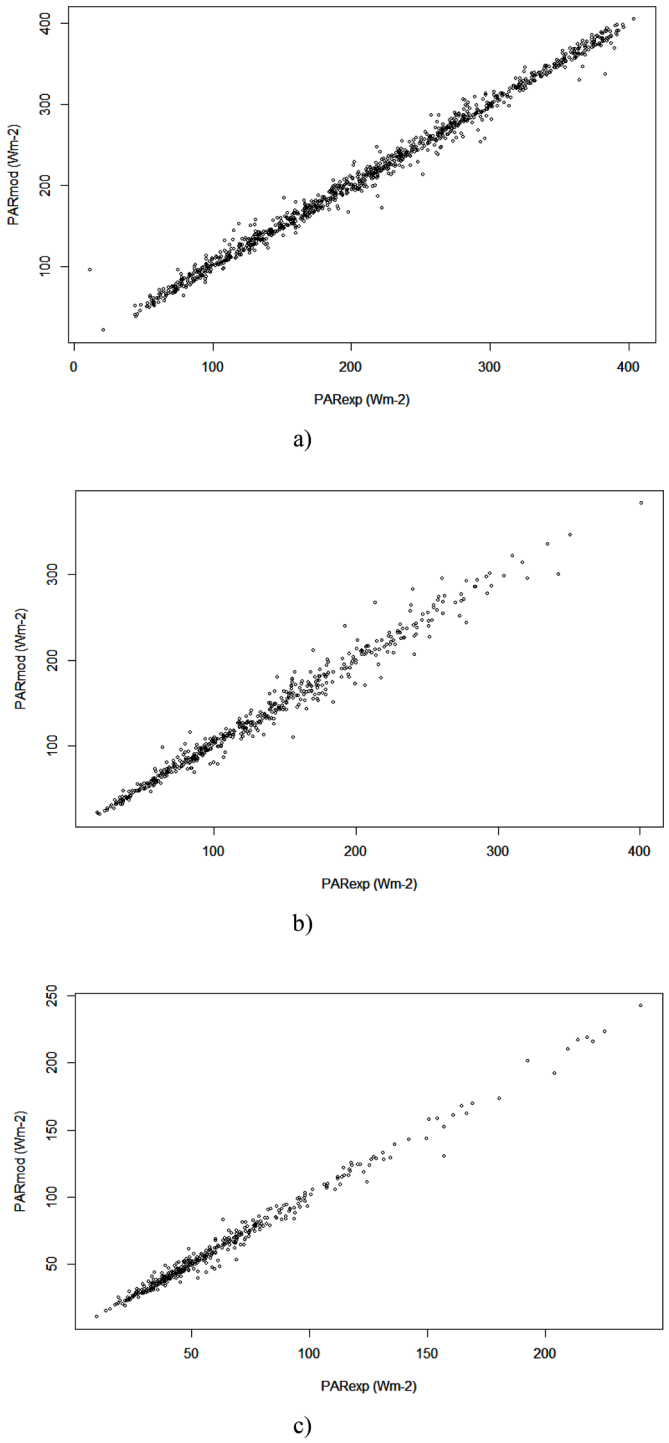

Fig. 8. Correlogram of the ANN model of *PAR*: a) Clear skies; b) Partial skies; c) Overcast skies.

conditions.

6. Artificial neural network for modelling *PAR*

This section proposes an Artificial Neural Network (ANN) trained with the Levenberg-Marquardt Back-Propagation (LMBP) algorithm (Du and Stephanus, 2018) that is used to model *PAR* for each of the CIE standard sky conditions (clear, partial and overcast skies). ANN training uses an iterative process to assign the correct weight that each neuron must set to each input to obtain the desired output value, *PAR* data in this case. Including the dependency between the input and output variables is not necessary, because the ANN is training to learn this dependency. ANN works as a 'black box' and only the final result is known and not the intermediate processes that are followed. It is therefore not possible to obtain the relationship between the input variables and their relative weight in the final result. Fig. 6 illustrates the system architecture used in this work.

A three-layer configuration has been chosen: the input layer, where the selected MIs for the model were introduced in the system; the hidden layer, or the information processing centers and the output layer, where the result is obtained. Each processing center (neuron) adjusts to the other neurons in an interactive process. The fit of the weights (weighting matrix, *W*) in each iteration is done using the Levenberg Marquardt algorithm (Lv et al., 2017). Fig. 7 shows the structure of a neuron where the input information is calculated from eq. 10:

$$n = \sum_{j=1}^{K} W_j P_j + b \tag{10}$$

W_j are the components of the weighting matrix, P_j are the input variables and b is the bias. The neuron generates an output, a, in Fig. 7, through the activation function ($f(n)$) given, in this work, by Eq. 11:

$$f(n) = \frac{1}{1 + e^{-n}} \tag{11}$$

The input data-set of an ANN is divided into three groups for its training process: a training group (with 70% of data), a validation data-set (15% of data) and a test data-set (15% of data). The training and validation groups have been established by dividing the fitting dataset used in the MLR modelling. The training group is used to determine the weighted matrix and the bias in an iterative process. The training is over when the results of the performance of the resulting model, calculated using the validation set, reach the desired quality. The test data group is used to calculate the performance of the model. The test data-set matched the one used for the test of the multilinear regression models fitted in the previous section.

Three ANNs were developed and tested in this work, one for each sky type, based on the feature selection previously performed and shown in Table 4. The statistic used for the validation of the models were the corresponding determination coefficient (R^2), the normalized mean bias error (*nMBE*), and the normalized root mean square error (*nRMSE*), previously defined. Table 6 summarizes the statistical results of the ANNs, and Fig. 8 shows a graph of these results.

The slope and intercept ($W{\cdot}m^{-2}$) of the correlograms shown in Figs. 5 and 8 give a practically perfect fit for all sky conditions, as the results show in Table 7.

Table 7
Results of the correlograms of the different models.

Sky conditions	R^2	Slope	Intercept ($W{\cdot}m^{-2}$)	Sky conditions	R^2	Slope	Intercept($W{\cdot}m^{-2}$)
Clear skies (MLR1)	0.992	1.0009	-0.6753	Clear skies (ANN1)	0.992	1.0004	-0.5484
Partial skies (MLR2)	0.978	0.9922	0.6845	Partial skies (ANN2)	0.978	0.9924	0.6078
Overcast skies (MLR3)	0.987	0.9990	-0.2836	Overcast skies (ANN3)	0.988	1.0002	-0.3550

7. Conclusions

Two different procedures, multilinear regression and ANNs, have been used to develop *PAR* models using Meteorological Indices (MIs). A previous feature selection procedure has been performed with the Pearson's correlation coefficient that pointed to the most influential variables, so that the irrelevant ones could be discarded. As its main novelty, sky conditions were included in the modelling of *PAR*, using the CIE standard sky classification as the criterion for the definition of clear, partial, and overcast skies. The experimental data-set collected at ten-minute intervals, was extended for twenty-two months. In addition to the traditional quality filters for solar radiation measurements, a second filtering procedure has been applied in statistical terms and according to the clustering of the data following the CIE standard sky classification, which guarantees the quality and homogeneity of the experimental data for each of the established sky conditions.

A feature selection procedure has been applied before the modelling of *PAR*, for selecting the most influential variables and discarding the most redundant ones. In the feature selection procedure, a maximum of three MIs were selected as input for the models. *RaGH* was the common MI used by all models and for all sky conditions. The additional variables were the geometrical parameter, *cosZ*, and three variables related to the sky conditions, k_t, ε, and Δ. These variables have been used in other works for modelling *PAR* (Aguiar et al., 2012; Alados-Arboledas et al., 2000; Alados et al., 1996; López et al., 2001; Wang et al., 2014) with different time intervals and following different strategies, obtaining comparable results to this work.

Both modelling methods, multilinear regression and ANN, have obtained very high determination coefficients (R^2) with very close results in the models for each of the different sky conditions. Slight improvements were observed in the ANN models. The lowest *nRMSE* values were obtained for clear skies models while modelling of partial conditions yielded the highest values. *nMBE* values were practically insignificant in all cases, although its negative value showed than all models tend to underestimate *PAR* values.

Regarding the ANN models of *PAR*, the use of Levenberg-Marquardt Back-Propagation (LMBP) algorithm with three layers (input, hidden layer and output) was the chosen configuration, based on literature data (Ferrera-Cobos et al., 2020) and previous experience.

In this study, the equivalence of multilinear regression models and ANN models of *PAR* following a feature selection procedure has been highlighted. The main advantage of the multilinear regression models is knowledge of the relationship between the variables: in the case of *PAR*, the *RaGH* coefficient indicated that *PAR* value is higher under overcast sky conditions than in partial and clear skies, as previous works have demonstrated (García-Rodríguez et al., 2020). The transferability of the models to other locations and its local adaptation will be discussed in future works.

CRediT authorship contribution statement

A. García-Rodríguez: Investigation, Methodology, Formal analysis, Validation. **D. Granados-López:** Investigation, Methodology, Formal analysis, Validation. **S. García-Rodríguez:** Investigation, Methodology, Software, Visualization. **M. Díez-Mediavilla:** Conceptualization, Supervision, Funding acquisition, Project administration. **C. Alonso-Tristán:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the Regional Government of Castilla y León, under projects BU021G19 and INVESTUN/19/BU/0004 and the Spanish Ministry of Science & Innovation under the I+D +i state program "Challenges Research Projects" (Ref. RTI2018-098900-B-I00). Diego Granados López expresses his thanks to the Junta de Castilla y León for economic support (PIRTU Program, ORDEN EDU/556/2019).

References

Aguiar, L.J.G., et al., 2012. Modeling the photosynthetically active radiation in South West Amazonia under all sky conditions. Theor. Appl. Climatol. 108 (3-4), 631–640.
Akitsu, T., Kume, A., Hirose, Y., Ijima, O., Nasahara, K.N., 2015. On the stability of radiometric ratios of photosynthetically active radiation to global solar radiation in Tsukuba. Japan. Agric. For. Meteorol. 209-210, 59–68.
Al-Shooshan, A.A., 1997. Estimation of photosynthetically active radiation under an arid climate. J. Agric. Eng. Res. 66 (1), 9–13.
Alados-Arboledas, L., Olmo, F., Alados, I., Perez, M., 2000. Parametric models to estimate photosynthetically active radiation in Spain. Agric. For. Meteorol. 101 (2-3), 187–201.
Alados, I., Alados-Arboledas, L., 1999. Validation of an empirical model for photosynthetically active radiation. Int. J. Climatol. 19 (10), 1145–1152.
Alados, I., Foyo-Moreno, I.y., Alados-Arboledas, L., 1996. Photosynthetically active radiation: measurements and modelling. Agric. For. Meteorol. 78 (1-2), 121–131.
Alados, I., Olmo, F.J., Foyo-Moreno, I., Alados-Arboledas, L., 2000. Estimation of photosynthetically active radiation under cloudy conditions. Agric. For. Meteorol. 102 (1), 39–50.
Alshaibani, K., 2011. Finding frequency distributions of CIE Standard General Skies from sky illuminance or irradiance. Light. Res. Technol. 43 (4), 487–495.
Bat-Oyun, T., Shinoda, M., Tsubo, M., 2012. Effects of cloud, atmospheric water vapour and dust on photosynthetically active radiation and total solar radiation in a Mongolian grassland. J. Arid Land 4 (4), 349–356.
Blackburn, W.J., Proctor, J.T.A., 1983. Estimating photosynthetically active radiation from measured solar irradiance. Sol. Energy 31 (2), 233–234.
Bosch, J.L., López, G., Batlles, F.J., 2009. Global and direct photosynthetically active radiation parameterizations for clear-sky conditions. Agric. For. Meteorol. 149 (1), 146–158.
Du, Y.-C., Stephanus, A., 2018. Levenberg-Marquardt neural network algorithm for degree of arteriovenous fistula stenosis classification using a dual optical photoplethysmography sensor. Sensors 18 (7), 2322.
Dye, D.G., 2004. Spectral composition and quanta-to-energy ratio of diffuse photosynthetically active radiation under diverse cloud conditions. J. Geophys. Res. Atmos. 109 (10), D10203, 1-12.
Erbs, D.G., Klein, S.A., Duffie, J.A., 1982. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. Solar Energy 28 (4), 293–302.
Escobedo, J.F., Gomes, E.N., Oliveira, A.P., Soares, J., 2009. Modeling hourly and daily fractions of UV, PAR and NIR to global solar radiation under various sky conditions at Botucatu. Brazil. Appl. Energy 86 (3), 299–309.
Ferrera-Cobos, F., Vindel, J., Valenzuela, R., González, J., 2020. Models for estimating daily photosynthetically active radiation in oceanic and mediterranean climates and their improvement by site adaptation techniques. Adv. Space Res. 65 (8), 1894–1909.
Finch, D., Bailey, W., McArthur, L., Nasitwitwi, M., 2004. Photosynthetically active radiation regimes in a southern African savanna environment. Agric. For. Meteorol. 122 (3-4), 229–238.
Foyo-Moreno, I., Alados, I., Alados-Arboledas, L., 2017. A new conventional regression model to estimate hourly photosynthetic photon flux density under all sky conditions. Int. J. Climatol. 37, 1067–1075.
Gao, Z., Xie, X., Gao, W., Chang, N.B., 2011. Spatial analysis of terrain-impacted photosynthetic active radiation (PAR) using MODIS data. GISCI. Remote Sens. 48 (4), 501–521.
García-Rodríguez, A., García-Rodríguez, S., Díez-Mediavilla, M., Alonso-Tristán, C., 2020. Photosynthetic Active Radiation, Solar Irradiance and the CIE Standard Sky Classification. Appl. Sci. 10 (22), 8007.
González, J.A., Calbó, J., 2002. Modelled and measured ratio of PAR to global radiation under cloudless skies. Agric. For. Meteorol. 110 (4), 319–325.
Granados-López, D., Díez-Mediavilla, M., Dieste-Velasco, M.I., Suárez-García, A., Alonso-Tristán, C., 2020. Evaluation of the vertical sky component without obstructions for daylighting in Burgos, Spain. Appl. Sci. 10 (9), 3095.
Granados-López, D., Suárez-García, A., Díez-Mediavilla, M., Alonso-Tristán, C., 2021. Feature selection for CIE standard sky classification. Sol. Energy 218, 95–107.
Gueymard, C.A., 2018. A reevaluation of the solar constant based on a 42-year total solar irradiance time series and a reconciliation of spaceborne observations. Sol. Energy 168, 2–9.
Gueymard, C.A., Ruiz-Arias, J.A., 2016. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. Sol. Energy 128, 1–30.
Hu, B., et al., 2018. Trends of photosynthetically active radiation over China from 1961 to 2014. Int. J. Climatol. 38 (10), 4007–4024.

Hu, B., Wang, Y., Liu, G., 2010. Long-term trends in photosynthetically active radiation in Beijing. Adv. Atmos. Sci. 27 (6), 1380–1388.
Huntingford, C., et al., 2019. Machine learning and artificial intelligence to aid climate change research and preparedness. Environ. Res. Lett. 14 (12), 124007.
Iqbal, M., 1983. An introduction to solar radiation.
ISO, 2004. Spatial Distribution of Daylight-CIE Standard General Sky. ISO-15469:2004 (E). Geneve, Switzerland.
Jacovides, C.P., Tymvios, F.S., Asimakopoulos, D.N., Theofilou, K.M., Pashiardes, S., 2003. Global photosynthetically active radiation and its relationship with global solar radiation in the Eastern Mediterranean basin. Theor. Appl. Climatol. 74 (3-4), 227–233.
Jacovides, C.P., Tymvios, F.S., Assimakopoulos, V.D., Kaltsounides, N.A., 2007. The dependence of global and diffuse PAR radiation components on sky conditions at Athens, Greece. Agric. For. Meteorol. 143 (3-4), 277–287.
Jacovides, C.P., Tymvios, F.S., Boland, J., Tsitouri, M., 2015. Artificial neural network models for estimating daily solar global UV, PAR and broadband radiant fluxes in an eastern Mediterranean site. Atmos. Res. 152, 138–145.
Janjai, S., Wattan, R., 2011. Development of a model for the estimation of photosynthetically active radiation from geostationary satellite data in a tropical environment. Remote Sens. Environ. 115 (7), 1680–1693.
Janjai, S., Wattan, R., Sripradit, A., 2015. Modeling the ratio of photosynthetically active radiation to broadband global solar radiation using ground and satellite-based data in the tropics. Adv. Space Res. 56 (11), 2356–2364.
Kasten, F., 1993. Discussion on the relative optical air mass. Light. Res. Technol. 25 (3), 129–130.
Kirk, J.T.O., 1979. Spectral distribution of photosynthetically active radiation in some South-eastern Australian Waters. Mar. Freshwater Res. 30 (1), 81–91.
Landsberg, J.J., Waring, R.H., 1997. A generalised model of forest productivity using simplified concepts of radiation-use efficiency, carbon balance and partitioning. Forest Ecol. Manag. 95 (3), 209–228.
Leuchner, M., Hertel, C., Menzel, A., 2011. Spatial variability of photosynthetically active radiation in European beech and Norway spruce. Agric. For. Meteorol. 151 (9), 1226–1232.
Li, D.H.W., Chau, N.T.C., Wan, K.K.W., 2013. Predicting daylight illuminance and solar irradiance on vertical surfaces based on classified standard skies. Energy 53, 252–258.
Li, D.H.W., Cheung, G.H.W., 2006. Average daylight factor for the 15 CIE standard skies. Light. Res. Technol. 38 (2), 137–152.
Li, D.H.W., Lam, T.N.T., Cheung, K.L. and Tang, H.L., 2008. An analysis of luminous efficacies under the CIE standard skies. 33(11): 2357-2365.
Li, L., et al., 2015. A method for estimating hourly photosynthetically active radiation (PAR) in China by combining geostationary and polar-orbiting satellite data. Remote Sens. Environ. 165, 14–26.
Liang, F., Xia, X.A., 2005. Long-term trends in solar radiation and the associated climatic factors over China for 1961-2000. Ann. Geophys. 23 (7), 2425–2432.
Liu, Z., Wu, C., Xu, S., 2016. Analyzing the potential of different fractions of photosynthetically active radiation (FPARS) to estimate gross primary production. Int. Geosci. Remote Sens. Symp. (IGARSS) 4355–4358.
López, G., Rubio, M.A., Martínez, M., Batlles, F.J., 2001. Estimation of hourly global photosynthetically active radiation using artificial neural network models. Agric. For. Meteorol. 107 (4), 279–291.
Lv, C., et al., 2017. Levenberg–Marquardt backpropagation training of multilayer neural networks for state estimation of a safety-critical cyber-physical system. IEEE Trans. Industr. Inform. 14 (8), 3436–3446.
Meek, D.W., Hatfield, J.L., Howell, T.A., Idso, S.B., Reginato, R.J., 1984. Generalized relationship between photosynthetically active radiation and solar radiation. J. Agron. 76 (6), 939–945.
Monteith, J.L., 1973. Principles of environmental physics.
Moon, P., 1940. Proposed standard solar-radiation curves for engineering use. J. Franklin Inst. 230 (5), 583–617.
Mukaka, M., 2012. Statistics corner: a guide to appropriate use of correlation in medical research. Malawi Med. J. 24 (3), 69–71.
Nwokolo, S.C., Amadi, S.O., 2018. A global review of empirical models for estimating photosynthetically active radiation. Tr. Ren. Energy 4 (2), 236–327.
Peng, S., et al., 2015. Observation and estimation of photosynthetically active radiation in Lhasa (Tibetan Plateau). Adv. Space Res. 55 (6), 1604–1612.
Perez, R., Ineichen, P., Seals, R., Michalsky, J., Stewart, R., 1990. Modeling daylight availability and irradiance components from direct and global irradiance. Sol. Energy 44 (5), 271–289.
Serrano, M.A., Boscà, J.V., 2011. Validation of a method to estimate direct normal irradiance of UVA and PAR bands from global horizontal measurements for cloudless sky conditions in Valencia, Spain, by a measurement campaign. Theor. Appl. Climatol. 103 (1), 95–101.
Stanhill, G., Fuchs, M., 1977. The relative flux density of photosynthetically active radiation. J. Appl. Ecol. 14 (1), 317–322.
Stigter, C.J., Musabilha, V.M.M., 1982. The conservative ratio of photosynthetically active to total radiation in the tropics (Dar es Salaam). J. Appl. Ecol. 19 (3), 853–858.
Suárez-García, A., Díez-Mediavilla, M., Granados-López, D., González-Peña, D., Alonso-Tristán, C., 2020. Benchmarking of meteorological indices for sky cloudiness classification. Sol. Energy 195, 499–513.
Suárez-García, A., Granados-López, D., González-Peña, D., Díez-Mediavilla, M., Alonso-Tristán, C., 2018. Seasonal caracterization of CIE standard sky types above Burgos, northwestern Spain. Sol. Energy 169, 24–33.
Szeicz, G., 1974. Solar radiation for plant growth. J. Appl. Ecol. 11 (2), 617–636.
Torres, A.F., Walker, W.R., McKee, M., 2011. Forecasting daily potential evapotranspiration using machine learning and limited climatic data. Agric. Water Manag. 98 (4), 553–562.
Torres, J.L., de Blas, M., García, A., Gracia, A., de Francisco, A., 2010a. Sky luminance distribution in Pamplona (Spain) during the summer period. J. Atmos. Sol.-Terr. Phys. 72 (5-6), 382–388.
Torres, J.L., de Blas, M., García, A., Gracia, A., de Francisco, A., 2010b. Sky luminance distribution in the North of Iberian Peninsula during winter. J. Geophys. Res. Solid Earth 72 (16), 1147–1154.
Tregenza, P.R., 2004. Analysing sky luminance scans to obtain frequency distributions of CIE Standard General Skies. Light. Res. Technol. 36 (4), 271–279.
Tsubo, M., Walker, S., 2005. Relationships between photosynthetically active radiation and clearness index at Bloemfontein, South Africa. Theor. Appl. Climatol. 80 (1), 17–25.
Vindel, J.M., et al., 2018. Modeling Photosynthetically active radiation from satellite-derived estimations over Mainland Spain. Remote Sens. 10 (6), 849.
Wang, L., et al., 2015. Modeling and analysis of the spatiotemporal variations of photosynthetically active radiation in China during 1961-2012. Renew. Sustain. Energy Rev. 49, 1019–1032.
Wang, L., Gong, W., Hu, B., Zhu, Z., 2014. Analysis of photosynthetically active radiation in Northwest China from observation and estimation. Int. J. Biometeorol. 59 (2), 193–204.
Wang, L., Kisi, O., Zounemat-Kermani, M., Hu, B., Gong, W., 2016. Modeling and comparison of hourly photosynthetically active radiation in different ecosystems. Renew. Sustain. Energy Rev. 56, 436–453.
Yamashita, M., Yoshimura, M., 2018. Influence of sky conditions on estimation of photosynthetic photon flux density for agricultural ecosystem. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 42, 3.
Yu, X., Guo, X., 2016. Hourly photosynthetically active radiation estimation in Midwestern United States from artificial neural networks and conventional regressions models. Int. J. Biometeorol. 60 (8), 1247–1259.
Yu, X., Wu, Z., Jiang, W., Guo, X., 2015. Predicting daily photosynthetically active radiation from global solar radiation in the Contiguous United States. Energy Convers. Manag. 89, 71–82.
Zhang, X., Zhang, Y., Zhoub, Y., 2000. Measuring and modelling photosynthetically active radiation in Tibet Plateau during April-October. Agric. For. Meteorol. 102 (2-3), 207–212.