AGU
ADVANCING EARTH AND SPACE SCIENCES

JGR Atmospheres

Key Points:
- Diffuse component of photosynthetically active radiation can be modeled even when no continuous measurements are available
- Diffuse photosynthetically active radiation can be modeled from the total solar spectrum with high performance
- Artificial neural networks and empirical models have similar results modeling from total range

Correspondence to:
I. Lozano,
ismael.lozano@helsinki.fi

Author Contributions:
Conceptualization: G. Sánchez-Hernández, I. Foyo-Moreno
Data curation: I. Lozano
Formal analysis: I. Lozano, I. Alados, I. Foyo-Moreno
Investigation: I. Lozano, I. Alados, G. Sánchez-Hernández, J. L. Guerrero-Rascado, I. Foyo-Moreno
Methodology: I. Alados, G. Sánchez-Hernández, I. Foyo-Moreno
Resources: J. L. Guerrero-Rascado, I. Foyo-Moreno
Supervision: I. Foyo-Moreno
Visualization: I. Lozano
Writing – original draft: I. Lozano, I. Alados, I. Foyo-Moreno
Writing – review & editing: I. Lozano, I. Alados, J. L. Guerrero-Rascado, I. Foyo-Moreno

Improving the Estimation of the Diffuse Component of Photosynthetically Active Radiation (PAR)

I. Lozano[1,2] , **I. Alados[3,4]**, **G. Sánchez-Hernández[4,5]**, **J. L. Guerrero-Rascado[4,6]** , **and I. Foyo-Moreno[4,6]**

[1]Department of Forest Sciences, University of Helsinki, Helsinki, Finland, [2]Institute for Atmospheric and Earth System Research, University of Helsinki, Helsinki, Finland, [3]Department of Applied Physics II, University of Málaga, Málaga, Spain, [4]Andalusian Institute for Earth System Research, Granada, Spain, [5]Department of Physics, University of Jaén, Jaén, Spain, [6]Department of Applied Physics, University of Granada, Granada, Spain

Abstract Most weather forecasting models are not able to accurately reproduce the great variability existing in the measurements of the diffuse component of photosynthetically active radiation (PAR; 400–700 nm) under all sky conditions. Based on the well-known relationship between the diffuse fraction (k) and the clearness index (k_t), this study addresses improvements in estimations by proposing adaptations of previous models, which were previously applied only to the total solar irradiance (TSI; 280–3,000 nm). In order to reproduce this variability, additional parameters were introduced. The models were tested employing a multisite database gathered at the Mediterranean basin. Since Artificial Neural Network (ANN) models are not limited to fixed coefficients to predict the diffuse fraction of PAR (k_{PAR}), these types of models are more accurate than empirical ones, reaching determination coefficients (r^2) up to 0.998. However, the simpler linear model proposed by Foyo-Moreno et al. (2018), https://doi.org/10.1016/j.atmosres.2017.12.012 shows a similar performance to the ANN models, directly predicting the diffuse component of PAR (PAR$_{Diffuse}$) from TSI$_{Diffuse}$, with a r^2 up to 0.997. Results obtained here also determine that the most important variables for estimating PAR$_{Diffuse}$ are k_t or $k_{t,PAR}$, and the apparent solar time (AST). Therefore, PAR$_{Diffuse}$ can be modeled using TSI measured in most radiometric stations, reaching r^2 up to 0.858 for empirical models and 0.970 for ANN models. This modified approach will allow for the very accurate construction of long-term data series of PAR$_{Diffuse}$ in regions where continuous measurements of PAR are not available.

Plain Language Summary Increasing and deepening the knowledge of photosynthetically active radiation (PAR; 400–700 nm wavelength) is important since it plays a key role as one of Earth's climate drivers, since PAR represents approximately 50% of sun's energy. PAR also plays a main role modulating emissions of greenhouse gases (GHGs) in biomass production (agriculture), among other issues. The knowledge of the diffuse component of PAR is especially crucial since it can be directly related to the light use efficiency (LUE) of plants, explaining somehow its effects over the aforementioned issues. However, at most measurement stations direct PAR measurements are not available. In fact, in those where PAR is measured it is rare to find diffuse PAR measurements, which limits the understanding of PAR effects. Modeling the diffuse component of PAR is an effective way to fill this knowledge gap but there are few models in the scientific literature that do it. In this study, we analyze previous models applied to other wavelengths in addition to new models and Artificial Neural Network approximation to estimate the diffuse component of PAR. Our main finding allows us to construct long-term data series of diffuse PAR even when no PAR measurements are available.

1. Introduction

Photosynthetically active radiation (PAR) commonly refers to solar radiation in the visible wavelength range (400–700 nm; McCree, 1972). It owes its name to the fact that algae, vascular plants, and photosynthetic bacteria use it to produce biomass from light energy. PAR is one of the key factors driving the photosynthesis process and, therefore, very important for organic carbon production and storage (Bauerle et al., 2014). Numerous recent studies emphasize the importance of increasing our knowledge of PAR, for example, the relationship between changes in PAR radiation intensity and efficiency in wastewater treatment (Verma et al., 2021), the regulatory effect of PAR on greenhouse gases (GHGs) emissions from crops (Keane et al., 2017), and the influence of PAR on the eutrophication of marine and continental waters due to the regulation of atmospheric N_2 fixation (Dou et al., 2021; Gradoville et al., 2021), among others.

As in the spectral range of total solar irradiance (TSI; 280–3,000 nm), the global PAR irradiance (PAR_{Global}) reaching the Earth's surface can be divided into a direct and a diffuse component. Precisely, the diffuse component of PAR ($PAR_{Diffuse}$) is especially important since plant photosynthesis tends to increase under diffuse light conditions (Gui et al., 2021; Mercado et al., 2009). Mercado et al. (2009) evaluated the impact of changes in diffuse radiation of PAR on the global land carbon sink. Other studies also reported that an increase in $PAR_{Diffuse}$ causes an increased light use efficiency by plants (Gu et al., 2002; Kanniah et al., 2012; Zhou et al., 2021), since the diffuse radiation tends to produce less canopy photosynthetic saturation (Gu et al., 2002). Precisely, since this diffuse component can promote vegetation photosynthesis, this is known as the diffuse fertilisation effect (DFE; Gui et al., 2021).

However, despite the interest of the applications that derive from deepening the knowledge of PAR, several authors have shown that there is a scarcity of experimental measurements of PAR radiation on the surface, and especially of its diffuse component (Ferrera-Cobos et al., 2020a; Niu et al., 2019; Wang et al., 2016). Surface PAR is not a common observation at meteorological stations. This is particularly true in the Mediterranean area (Di Biagio et al., 2009). Therefore, considering the scarcity of PAR radiation measurements, different models have been proposed in the literature for its estimation. Notable among these models are those that estimate PAR from TSI from surface measurements, from satellite data averages (e.g., Foyo-Moreno et al., 2017; Hao et al., 2019; Peng et al., 2015; Vindel et al., 2018), or from radiation spectral measurements (Trisolino et al., 2016). García-Rodríguez et al. (2022) also proposed two different ways to model PAR from meteorological indices: firstly, by a multilinear regression model, and secondly by using artificial neural networks (ANN). Another very common approach is to consider the PAR-TSI fraction as a constant when estimating PAR radiation (e.g., Janjai et al., 2015; Yu et al., 2015). However, all this diversity of models to obtain PAR values have no equivalent to evaluate the diffuse component of PAR, for which a much more modest number of models have been proposed (e.g., Foyo-Moreno et al., 2018; Jacovides et al., 2010; Lozano et al., 2022).

From this contextualization, the aim of this study is to carry out a comparative analysis of several empirical and ANN models, used to estimate the diffuse component of PAR with the goal to improve its estimations, which can be applied to optimize crop management practices, increase prediction accuracy, and optimize solar panel performance, and so on. The basis of some of the models is the $k_t - k$ relationship for PAR, but also others estimate $PAR_{Diffuse}$ directly. The article will be structured as follows. The experimental setup and data quality are described in Section 2. Section 3 includes the description of the empirical and ANN models together with the statistical analysis, while Section 4 presents the results and discussion, in terms of a comparative analysis between ANN and empirical models. Finally, the main conclusions are given in Section 5. Glossary is a list of abbreviations to assist the reader.

2. Experimental Setup and Data Control Quality

An experimental dataset with surface measurements of global and diffuse PAR and TSI in Granada (Spain, 37.16°N 3.61°W, 680 m a.s.l., years 1994 and 1995), Almería (Spain, 36.83°N 2.41°W, 21 m a.s.l., years 1993–1995), and Renon (Italy, 46.42°N 11.28°E 1735 m a.s.l., years 2014 and 2015) have been used in this work. The data were acquired at 1-min intervals for each station and hourly values were generated from them. Therefore, data were stored at 1-hr average intervals. Previous studies using 1-min data have shown that the functional forms analyzed in this study can also be applied at higher temporal resolution despite differences in dispersion, for example, Engerer (2015) and Yang and Gueymard (2020) employ the $k - k_t$ relationship in TSI with a temporal resolution of 1-min. The period of study of each station (2 or 3 years) includes a wide range of seasonal conditions and solar zenith angles.

In Granada and Almería, the instrumentation used to measure PAR were two Licor Li-190SA quantum sensors (Lincoln, NE, ISA), and two radiometers Kipp and Zonen CM-11 (Delf. Netherlands) to measure TSI, being one of each kind of instrument mounted on a polar axis shadowband in order to measure $PAR_{Diffuse}$ and TSI. The Li-190SA has a relative error lower than 5% meanwhile the CM-11 directional error is lower than 10 Wm^{-2} for solar zenith angles (θ_z) up to 80 (Kipp & Zonen, 2000). In Renon, PAR was measured by a BF2 sunshine sensor (Delta-T Devices, Burwell, United Kingdom). This device uses an array of silicon photodiodes and a shading pattern on the radiometer dome to determine the diffuse component. The BF2 has a relative error lower than 15%; more details can be found in Foyo-Moreno et al. (2018). The diffuse measurements in both spectral ranges were corrected following the Batlles et al. (1995) method, and a conversion factor of 4.57 µmol m^{-2} s^{-1}/Wm^{-2}
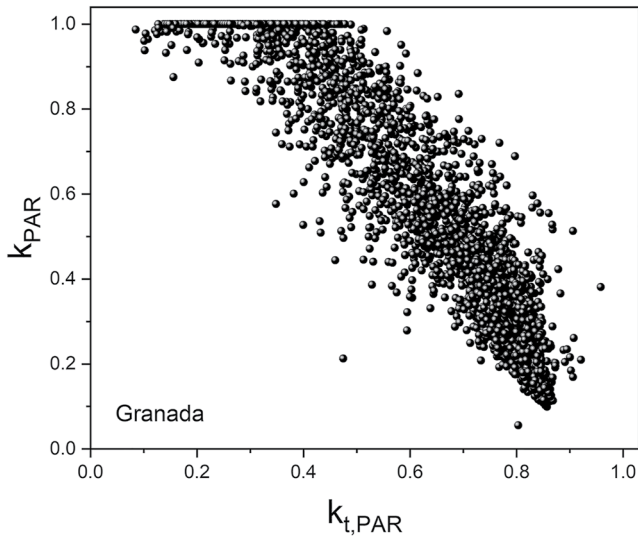
Figure 1. Relationship between k_{PAR} and $k_{t,PAR}$ for the experimental measurements at Granada for the 2-year period analyzed (1994 and 1995).

(McCree, 1972) was used to convert to energy units the PAR measurements. Further details for the instrumentation can be found in Foyo-Moreno et al. (2018) and Lozano et al. (2022, 2021).

An in-depth quality control analysis was performed to detect and remove low-accurate and anomalous data. Firstly, only those measurements recorded at $\theta_z < 80°$ have been used in order to avoid the cosine response error in solar radiation measurements. Secondly, we employed two tests based on the clearness index (k_t), defined as the ratio between global irradiance and extraterrestrial global irradiance both on a horizontal surface and the diffuse fraction (k; defined as the ratio between diffuse and global irradiance). The expressions to estimate the clearness index ($k_{t,PAR}$) and the diffuse fraction (k_{PAR}) in the PAR range are:

$$k_{t,PAR} = \frac{PAR_{Global}}{I_{TOA,PAR}} \tag{1}$$

$$k_{PAR} = \frac{PAR_{Diffuse}}{PAR_{Global}} \tag{2}$$

where $I_{TOA,PAR}$ is the PAR irradiance on the top of the atmosphere (TOA), and can be computed by the product of the eccentricity correction factor of the Earth's orbit (E_0), the cosine of θ_z, and the solar constant for the PAR range ($I_{SC,PAR} = 531.8$ Wm^{-2}; Gueymard, 2018). PAR_{Global} is the sum of the components of PAR, that is, the direct and diffuse components. Therefore, $0 < k_{PAR} < 1$ and $0 < k_{t,PAR} < 1$ constraints were applied. Thirdly, based on the dependence between the solar radiation and θ_z, it is possible to parametrize two envelopes (upper and lower) of the data using a linear equation. The upper envelope corresponds to the maximum values, clear skies for global PAR and overcast skies for diffuse PAR, meanwhile the lower envelope corresponds to the minimum values (Foyo-Moreno et al., 2017, 2018). In order to avoid outliers and extreme values, these envelopes have been computed according to the 1st and 99th percentiles. The experimental data were then considered to be the data that remained between these two envelopes, both for PAR_{Global} and $PAR_{Diffuse}$. Finally, a visual inspection was performed to detect anomalous data (mainly due to voltage malfunctioning and outliers) after the performance of the above tests. The final dataset was 2,578 rows (of data) for Granada, 4,471 for Almería and 4,714 for Renon.

3. Methodology

To estimate $PAR_{Diffuse}$ two types of approaches have been widely applied until now. The first one uses the well-known dependence of k on k_t. This relationship has been widely studied in the scientific literature in the TSI spectrum. This relation is used to estimate the diffuse component of the solar radiation due to there being much less uncertainty than the relation between the absolute values of diffuse and global irradiances (e.g., Badarinath et al., 2007a, 2007b; Meloni et al., 2006). In fact, this relationship is also employed in several studies to make estimations in other solar spectral ranges such as the TSI or ultraviolet (UV) spectral ranges (e.g., Ridley et al., 2010; Sánchez et al., 2017). Figure 1 shows the relationship between those non-dimensional indexes for PAR, that is, k_{PAR} and $k_{t,PAR}$. Subsequently $PAR_{Diffuse}$ is obtained by multiplying the estimated ratio (k_{PAR}) by PAR_{Global}. The second approach for estimating $PAR_{Diffuse}$ uses the absolute values of the irradiances involved in them, thus the estimation of $PAR_{Diffuse}$ is direct.

To carry out the analysis of both approaches, the dataset described in the previous section was employed. This data set has been averaged into hourly timesteps and randomly divided into two subsets: (a) a subset containing 75% of the data, to fit the models and obtain the empirical coefficients as well as to train the ANN, and (b) a subset composed of the 25% of data remaining for the validation and comparison of the empirical models and the ANN models.

3.1. Empirical Models

Most of the models analyzed in this study are based on relationships proposed in the literature to estimate the diffuse component of TSI. All of them present functional forms that are easy to fit in order to obtain their

empirical coefficients. In addition, most of them involve independent variables that only require PAR_{Global} data and solar geometry factors, calculated from the date and time at which each measurement is recorded and, therefore, favoring their application worldwide to generate long-term time series.

The first model (M1) analyzed in this work is based on the model proposed by Reindl et al. (1990). Reindl et al. (1990) analyzed a multilinear relationship between k in the shortwave radiation vs. several variables, including k_t and solar elevation. Their main two findings are: (a) k_t is the most important predictor of the diffuse fraction in TSI range for cloudy skies (medium and low values of k_t), and (b) under clear skies (high values of k_t) the importance of k_t dramatically decreases and the solar position becomes more relevant. According to this, the model has been applied to the PAR range, predicting k_{PAR} using, as input variables, $k_{t,PAR}$ and the cosine of θ_z, which form the expression:

$$k_{PAR} = a_1 + a_2 k_{t,PAR} + a_3 \cos \theta_z \tag{3}$$

where a_i are the fitting coefficients.

The second model (M2) is an adaptation to the PAR range of the model proposed by Ridley et al. (2010), based on a logistic function that tries to reproduce the differences in the behavior of k_{PAR} for different intervals of $k_{t,PAR}$. Jacovides et al. (2010) and Kathilankal et al. (2014) have analyzed different versions of this logistic model in the PAR spectral range including different independent variables. In particular, in our work the original and most complete version of the model is analyzed in the PAR range, as follow:

$$k_{PAR} = \frac{1}{1 + e^{\left(b_1 + b_2 k_{t,PAR} + b_3 AST + b_4 \alpha + b_5 K'_{t,PAR} + b_6 \Psi_{PAR}\right)}} \tag{4}$$

where b_i are the fitting coefficients, AST is the apparent solar time (in hour), computed following the Iqbal (1983) equation. α is the solar elevation in degrees, $k'_{t,PAR}$ is the daily clearness index and Ψ_{PAR} is the so-called persistence index. All these variables are introduced in order to reproduce the diffuse fraction variability. The solar elevation, α, accounts for the increase in Rayleigh scattering as α decreases, while AST considers the differences in the atmosphere between morning and afternoon. $k'_{t,PAR}$ is a measure of the PAR daily variability, primarily associated with clouds. Finally, Ψ_{PAR} is a measure of the atmospheric stability in terms of k_t: the closer the Ψ_{PAR} is to k_t, the clearer the atmosphere, and the opposite for overcast skies. This variable is evaluated with respect to the previous and next time (time−1, and time+1, respectively). These two last variables in the PAR range are defined as:

$$K'_{t,PAR} = \frac{\sum_{i=1}^{24} PAR_{Global}}{\sum_{i=1}^{24} I_{TOA,PAR}} \tag{4a}$$

$$\Psi_{PAR} = \frac{k_{t,PAR}^{time-1} + k_{t,PAR}^{time+1}}{2} \tag{4b}$$

The third model (M3) is the adaptation of the Ridley's model suggested by Lozano et al. (2022) who analyzed the relationship between k_{PAR} and k_t concluding in their preliminary analysis that k_{PAR} can be predicted directly from TSI. Therefore, the resulting expression is the following:

$$k_{PAR} = \frac{1}{1 + e^{\left(c_1 + c_2 k_t + c_3 AST + c_4 \alpha + c_5 K'_t + c_6 \Psi\right)}} \tag{5}$$

where c_i are the fitting coefficients, k'_t is the daily clearness index and Ψ is the persistence index, and are defined as:

$$K'_t = \frac{\sum_{i=1}^{24} TSI_{Global}}{\sum_{i=1}^{24} I_{TOA}} \tag{5a}$$

Table 1
Summary of the Empirical Models Employed to Estimate k_{PAR} (M1 to M5) or $PAR_{Diffuse}$ (M6) Directly

Model	Output	Predictors	Functional form	Equation
M1	k_{PAR}	$K_{t,PAR}$ $\cos\theta_z$	Linear: $f(x) = a + bx$	$k_{PAR} = a_1 + a_2 k_{t,PAR} + a_3 \cos\theta_z$
M2	k_{PAR}	$K_{t,PAR}$ AST α $K'_{t,PAR}$ Ψ_{PAR}	Logistic: $f(x) = \frac{1}{1+e^{g(x)}}$	$k_{PAR} = \frac{1}{1+e^{\left(b_1 + b_2 k_{t,PAR} + b_3 AST + b_4 \alpha + b_5 K'_{t,PAR} + b_6 \Psi_{PAR}\right)}}$
M3	k_{PAR}	K_t AST α K'_t Ψ	Logistic: $f(x) = \frac{1}{1+e^{g(x)}}$	$k_{PAR} = \frac{1}{1+e^{\left(c_1 + c_2 k_t + c_3 AST + c_4 \alpha + c_5 K'_t + c_6 \Psi\right)}}$
M4	k_{PAR}	$K_{t,PAR}$ AST α $K'_{t,PAR}$ Ψ_{PAR}	Polynomial: $f(x) = a + bx + cx^2 + dx^3$	$k_{PAR} = d_1 + d_2 k_{t,PAR} + d_3 k_{t,PAR}^2 + d_4 k_{t,PAR}^3 + d_5 \alpha + d_6 AST + d_7 K'_{t,PAR} + d_8 \Psi_{PAR}$
M5	k_{PAR}	K_t AST α K'_t Ψ	Polynomial: $f(x) = a + bx + cx^2 + dx^3$	$k_{PAR} = e_1 + e_2 k_t + e_3 k_t^2 + e_4 k_t^3 + e_5 \alpha + e_6 AST + e_7 K'_t + e_8 \Psi$
M6	$PAR_{Diffuse}$	$TSI_{Diffuse}$ I_{TOA} $\cos\theta_z$	Linear: $f(x) = a + bx$	$PAR_{Diffuse} = f_1 \frac{TSI_{Diffuse}}{I_{TOA}} \cos\theta_z$

$$\Psi = \frac{k_t^{time-1} + k_t^{time+1}}{2} \qquad (5b)$$

where I_{TOA} is the TSI irradiance on TOA, computed as described above for PAR solar range:

$$I_{TOA} = E_0 \cdot \cos\theta_z \cdot I_{SC} \qquad (6)$$

where I_{SC} is the solar constant ($I_{SC} = 1,361.1$ Wm^{-2}; Gueymard, 2018).

The fourth model (M4) is inspired by de Miguel's et al. (2001) model, who proposed a third-order polynomial expression with k_t as the only independent variable. As a result they obtained a curve reproducing the behavior of k_{PAR} in all the $k_{t,PAR}$ range. However, a model that uses an unique variable to predict the behavior of k_{PAR} is unable to reproduce all its variability. For this reason, as a novelty, this study proposes, in addition to applying the model to the PAR range, to add two independent variables to reproduce the k_{PAR} variability. Thus, the variables included in this new model are the solar position (in terms of α) and the persistence index (Equation 4b). The resulting expression is:

$$k_{PAR} = d_1 + d_2 k_{t,PAR} + d_3 k_{t,PAR}^2 + d_4 k_{t,PAR}^3 + d_5 \alpha + d_6 AST + d_7 K'_{t,PAR} + d_8 \Psi_{PAR} \qquad (7)$$

where d_i are the fitting coefficients.

The fifth model (M5) is the same model as M4, but in this case using k_t as predictor of k_{PAR}, instead of using $k_{t,PAR}$:

$$k_{PAR} = e_1 + e_2 k_t + e_3 k_t^2 + e_4 k_t^3 + e_5 \alpha + e_6 AST + e_7 K'_t + e_8 \Psi \qquad (8)$$

where e_i are the fitting coefficients.

Finally, the sixth model (M6), proposed by Foyo-Moreno et al. (2018), has been analyzed. Unlike the previous models, M6 directly estimates the $PAR_{Diffuse}$, instead of k_{PAR}. To this aim, M6 is based on the decomposition of the TSI_{Global} into its direct and diffuse components, from the following expression:

$$PAR_{Diffuse} = f_1 \frac{TSI_{Diffuse}}{I_{TOA}} \cos\theta_z \qquad (9)$$

where f_1 is the fitting coefficient, and $TSI_{Diffuse}$ is the diffuse component of TSI. This empirical model is based on the previous model proposed by Foyo-Moreno et al. (2017) and developed from experimental measurements recorded at Granada.

Table 1 is a summary of the empirical models. It should be highlighted that models M3, M5 and M6 allow for estimating the diffuse component in the PAR range even when global PAR irradiance measurements are not available because those models directly employ global or diffuse TSI.

3.2. Neural Network Models

In addition to the empirical models, several ANN models have been used to estimate k_{PAR} and $PAR_{Diffuse}$ directly. ANN have been widely used in many research fields (e.g., Desai & Shah, 2021; Ghritlahre & Prasad, 2018;

Table 2
Relative Importance of the Variables Evaluated Modeling k_{PAR} and $PAR_{Diffuse}$ With ANN, for Granada (GR), Almería (AL), and Renon (RE)

Modeled variable	Site	Predictor variable				
		$k_{t,PAR}$	AST	Ψ_{PAR}	$K'_{t,PAR}$	$\cos\theta_z$
k_{PAR}	GR	0.56	0.26	0.08	0.05	0.04
	AL	0.53	0.23	0.09	0.06	0.09
	RE	0.50	0.24	0.04	0.16	0.06
$PAR_{Diffuse}$	GR	0.26	0.28	0.07	0.06	0.33
	AL	0.23	0.28	0.05	0.06	0.39
	RE	0.30	0.22	0.06	0.15	0.27

Ziółkowski et al., 2021) and have taken a relevant role in solar radiation modeling in recent years (e.g., Ağbulut et al., 2021; Elsheikh et al., 2019; Kamadinata et al., 2019). In the present work, a multilayer perceptron (MLP) has been used to model k_{PAR} and $PAR_{Diffuse}$. A MLP is a feedforward ANN organized in at least three layers of fully connected neurons which employ a non-linear activation function. As an ANN a MLP is characterized by three fundamental elements: the training algorithm, the activation function and its architecture which determines the connections between neurons. This particular type of ANN is composed of a set of input neurons (input layer), equal to the variables used to estimate k_{PAR} or $PAR_{Diffuse}$ and a set of one or more hidden layers of neurons (hidden layers). More details on solar irradiance predictions using a MLP model can be followed in Alados et al. (2004, 2007). In our study, an automatic architecture selection was used in order to select the best architecture and to build a network with one hidden layer. The automatic architecture selection helped to specify the best number of units (neurons) for the hidden layer, ultimately setting it as the minimum number of hidden neurons with the maximum possible performance. An ANN with one hidden layer was employed in several previous works to estimate solar radiation (e.g., Alsina et al., 2016; Hasni et al., 2012; Kamadinata et al., 2019). Finally, a layer of output neurons (output layer), which is equal to the modeled variable (k_{PAR} or $PAR_{Diffuse}$) was used.

The activation functions used are hyperbolic tangent expressed as:

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{10}$$

where x is the input vector. The weight values were initialized as randomised values, and in order to minimize the mismatch between measured values and the computed values the learning algorithm modified the weights in the so-called forward-propagation phase. The error in each neuron, per iteration, is calculated as follow:

$$SE = \sum_{i=1}^{N} \left(z_i - z_i^* \right)^2 \tag{11}$$

where SE is the sum of the square error, z_i represents the real value and z_i^* is the estimated value. A more detailed description of this algorithm can be found in Alados et al. (2007, 2004) where the same type of MLP was employed for modeling radiometric UV erythemal irradiance.

The ANNs have been built using the same input variables as employed in empirical models described in Section 3.1, selecting the input variables by two predictor importance tests, for each location, the first in order to predict k_{PAR}, and the second one to predict $PAR_{Diffuse}$ directly (Table 2).

A predictor importance test is a statistical analysis of the relative importance of the predictors on the studied variable, and it is a crucial part of the pre-processing analysis to build an ANN, since it allows a reduction in the number of inputs into the model, restricting them to those that have proven to have a greater influence when performing the predictions. In our study, the predictor importance test was performed with the IBM spss software; this test is a part of the MLP app. The first predictor importance test points out $k_{t,PAR}$ is the most important factor when predicting k_{PAR}, for all the sites, with at least 0.5 relative importance followed by AST (0.24–0.26). The other variables have less relative importance, varying between 0.04 and 0.09 for the three sites except for Renon in which $K'_{t,PAR}$ reaches approximately 0.16 of relative importance. The high relative importance as a predictor of k_{PAR} is not surprising given the relationship between k_{PAR} and $k_{t,PAR}$ shown in Figure 1, which is one of the basis of most models in the scientific literature. This is because the knowledge of the atmosphere transparency is essential but not sufficient to reproduce k_{PAR} variability.

Taking into account the results of this importance analysis, we have evaluated two ANNs to predict k_{PAR}, in each of the three locations of study (Granada, Almeria and Renon): (a) ANN1 from $k_{t,PAR}$ and AST, (b) ANN2 using as parameters k_t instead of $k_{t,PAR}$, in order to be able to compute k_{PAR} even if there is a lack of global PAR measurements, and AST.

From the second predictor importance test performed for $PAR_{Diffuse}$ as the output variable, we obtain three variables with a high and similar relative importance: $k_{t,PAR}$, AST and $\cos\theta_z$. These results allow us to build an ANN

from these variables: (a) ANND1 using $k_{t,PAR}$, AST and $\cos\theta_z$, (b) ANND2 similar to ANND1 but using k_t instead of $k_{t,PAR}$, and (c) ANND3 built by using $TSI_{Diffuse}$ as a unique predictor. The advantage of using this parameter is that TSI is a variable more frequently measured in the radiometric stations than PAR, and in those in which the diffuse component of TSI is measured, this variable allows reproduction of the solar radiation interactions through the atmosphere.

3.3. Fitting and Validation Statistics

In order to analyze and compare the models, a statistical analysis has been performed including the mean bias error (MBE), the coefficient of determination (r^2) and the relative root mean square error (rRMSE). The MBE provides us with information about whether the model overestimates (positive values) or underestimates (negative values) with respect to the experimental values. The lower the value for both rRMSE and MBE, the better the behavior of our model (Ma & Iqbal, 1984). r^2 is an estimate of the total variance explained by the model, while the rRMSE allows us to make a term-to-term comparison between the experimental and estimated values of the diffuse fraction, and quantifies the differences between the estimated and experimental values. These statistics can be obtained by the following expressions:

$$MBE = \frac{1}{N}\sum_{i=1}^{N}\left(k_{PAR,i} - k_{PAR,i}^*\right) \tag{12}$$

$$r^2 = 1 - \frac{\sum_{i=1}^{N}\left(k_{PAR,i} - k_{PAR,i}^*\right)^2}{\sum_{i=1}^{N}\left(k_{PAR,i} - \overline{k_{PAR}}\right)^2} \tag{13}$$

$$rRMSE = \frac{100}{k_{PAR}}\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(k_{PAR,i} - k_{PAR,i}^*\right)^2} \tag{14}$$

where $k_{PAR,i}$ is the experimental diffuse fraction, and $k_{PAR,i}^*$ is the estimated diffuse fraction.

4. Results and Discussion

4.1. Modeling k_{PAR}

Table 3 shows the fittings coefficients obtained for the analyzed models. p-value have been obtained for all the coefficients in order to determine their statistical significance. The coefficient is considered to be significant when p-value <0.05 or p-value <0.01, marked with * or **, respectively. For Renon, all the coefficients for the five models are significant, except the coefficient e_4 in M5 (for $k_{t,PAR}^3$). M1 coefficients are significant for Granada and Almería, and M2 has significant coefficients in Almería as well. For Granada $K'_{t,PAR}$ and K'_t are not significant for any model while for Almeria K'_t is not significant. On the other hand, most of the coefficients for the polynomial expression in models M4 and M5 are not significant for both Granada and Almeria. Finally, α is not significant in M2 and M3 models for Granada. These findings show that for different models and locations k_t and the solar position could be irrelevant in contrast with the result obtained by Reindl et al. (1990) for TSI in which they found that the solar position and/or k_t are the main factors needed to model k, depending on the atmospheric conditions in terms of k_t.

Table 4 shows the values of the statistics previously defined in Section 3.3 for all the models analyzed. The values are very similar for both subsets of data (fitting and validation datasets), which confirms the good results of the models and allows us to focus on analyzing the validation dataset. The MBE values are close to 0, which indicates the absence of overestimation or underestimation of these models. In general, all the models work remarkably well, M1 showing the lowest performance with r^2 ranging from 0.550 in Almería to 0.679 in Granada. All the other models have a similar statistical behavior in each location. It is remarkable that the best performance of all the models (except for M1) is obtained in Renon. In this location the best performance was observed counterintuitively in model M5, which employs k_t to model k_{PAR} instead of using $k_{t,PAR}$, with a r^2 of 0.858, and indeed this model has the lowest rRMSE values, 18.3%, although M2 to M5 show quite similar performance. In Granada

Table 3
Fitting Coefficients for the Analyzed Empirical Models (M1 to M5) for Granada (GR), Almería (GR), and Renon (RE)

Site	Model	Coefficient							
		1	2	3	4	5	6	7	8
GR	M1	$1.317 \pm 0.013^{**}$	$-0988 \pm 0.019^{**}$	$-0.237 \pm 0.015^{**}$					
	M2	$-7.84 \pm 0.12^{**}$	$7.16 \pm 0.13^{**}$	$0.203 \pm 0.006^{**}$	0.0005 ± 0.0009	-0.27 ± 0.18	$1.04 \pm 0.13^{**}$		
	M3	$-7.35 \pm 0.11^{**}$	$7.58 \pm 0.13^{**}$	$0.178 \pm 0.006^{**}$	0.0016 ± 0.0008	-0.24 ± 0.20	$1.16 \pm 0.14^{**}$		
	M4	$1.77 \pm 0.08^{**}$	-0.4 ± 0.4	-1.1 ± 0.8	0.4 ± 0.4	$-0.039 \pm 0.001^{**}$	$-0.00041 \pm 0.0015^{**}$	0.05 ± 0.03	$-0.196 \pm 0.024^{**}$
	M5	$1.63 \pm 0.07^{**}$	-0.1 ± 0.4	$-2.0 \pm 0.9^{*}$	0.8 ± 0.6	$-0.0345 \pm 0.0010^{**}$	$-0.00059 \pm 0.00015^{**}$	0.05 ± 0.03	$-0.22 \pm 0.03^{**}$
AL	M1	$1.157 \pm 0.011^{**}$	$-0.832 \pm 0.019^{**}$	$-0.188 \pm 0.015^{**}$					
	M2	$-7.33 \pm 0.13^{**}$	$6.67 \pm 0.10^{**}$	$0.215 \pm 0.005^{**}$	$-0.0035 \pm 0.0008^{**}$	$-0.46 \pm 0.17^{**}$	$0.93 \pm 0.10^{**}$		
	M3	$-7.02 \pm 0.12^{**}$	$6.88 \pm 0.11^{**}$	$0.184 \pm 0.004^{**}$	$-0.0019 \pm 0.0007^{**}$	0.26 ± 0.19	$0.94 \pm 0.11^{**}$		
	M4	$1.80 \pm 0.08^{**}$	-0.6 ± 0.4	-1.0 ± 0.7	0.4 ± 0.4	$-0.0441 \pm 0.0009^{**}$	$0.00060 \pm 0.00015^{**}$	$0.10 \pm 0.03^{**}$	$-0.19 \pm 0.02^{**}$
	M5	$1.78 \pm 0.06^{**}$	-0.9 ± 0.4	-0.3 ± 0.8	-0.2 ± 0.5	$-0.0383 \pm 0.0009^{**}$	$-0.00035 \pm 0.00014^{**}$	0.05 ± 0.04	$-0.19 \pm 0.022^{**}$
RE	M1	$1.19 \pm 0.01^{**}$	$-0.980 \pm 0.013^{**}$	$-0.182 \pm 0.015^{**}$					
	M2	$-3.07 \pm 0.09^{**}$	$5.2 \pm 0.1^{**}$	$-0.168 \pm 0.005^{**}$	$0.0112 \pm 0.0009^{**}$	$1.99 \pm 0.13^{**}$	$0.93 \pm 0.11^{**}$		
	M3	$-0.72 \pm 0.09^{**}$	$5.36 \pm 0.09^{**}$	$-0.209 \pm 0.006^{**}$	$0.0086 \pm 0.0010^{**}$	$-5.15 \pm 0.19^{**}$	$1.54 \pm 0.11^{**}$		
	M4	$0.852 \pm 0.017^{**}$	$-0.32 \pm 0.11^{**}$	$-1.55 \pm 0.27^{**}$	$1.22 \pm 0.19^{**}$	$0.0329 \pm 0.0009^{**}$	$-0.00114 \pm 0.00015^{**}$	$-0.38 \pm 0.02^{**}$	$-0.118 \pm 0.018^{**}$
	M5	$0.489 \pm 0.014^{**}$	$-1.50 \pm 0.09^{**}$	$0.50 \pm 0.22^{*}$	0.21 ± 0.15	$0.0387 \pm 0.0008^{**}$	$-0.00045 \pm 0.00013^{**}$	$1.33 \pm 0.03^{**}$	$-0.188 \pm 0.015^{**}$

Note. Significance of coefficients are also shown: p-value <0.05 (*) or p-value <0.01 (**).

Table 4
Statistical Performance of the Models for Estimating the Diffuse Fraction of the PAR Range Corresponding to the Fit Data Set, on the Left, and the Validation Data Set, on the Right, for Granada (GR), Almería (AL), and Renon (RE)

Site	Model	Fitting			Validation		
		r^2	MBE	rRMSE (%)	r^2	MBE	rRMSE (%)
GR	M1	0.678	$5 \cdot 10^{-5}$	26.8	0.679	0.002	27.3
	M2	0.829	0.003	19.5	0.805	0.002	21.3
	M3	0.833	0.003	19.3	0.806	0.003	21.2
	M4	0.821	−0.002	19.9	0.801	−0.002	21.9
	M5	0.827	−0.002	19.5	0.804	$3 \cdot 10^{-4}$	21.7
AL	M1	0.552	$3 \cdot 10^{-5}$	33.9	0.550	−0.005	34.7
	M2	0.768	−0.002	24.4	0.767	−0.007	24.9
	M3	0.777	−0.002	23.9	0.771	−0.005	24.7
	M4	0.742	−0.003	25.1	0.752	−0.005	25.9
	M5	0.756	−0.002	24.5	0.758	−0.004	25.4
RE	M1	0.655	$4 \cdot 10^{-5}$	27.8	0.646	−0.005	27.7
	M2	0.839	0.018	19.0	0.821	0.015	19.6
	M3	0.820	0.026	20.1	0.842	0.028	18.5
	M4	0.829	−0.013	20.2	0.812	−0.019	21.1
	M5	0.852	−0.011	18.3	0.858	−0.012	18.3

and Almería, all the models have pretty similar performance, with Almería having the lowest performance of the three locations. In Almería, the minimum values were obtained with M4 giving a $r^2 = 0.752$, and the maximum values were from M3 with $r^2 = 0.771$. It is important to note that models M2 to M5 employ all the same variables to obtain k_{PAR}, although there are two differences between them. Firstly, the functional form in M2 and M3 are based on the logistic equation while M4 and M5 are based on a polynomial equation. Although the behavior of the models is quite similar regardless of the functional form analyzed, in Granada and Almeria the logistic models behave slightly better, while in Renon the polynomial model does. Secondly, in the models M2 and M4 k_{PAR} use measurements in the PAR range of solar radiation, meanwhile M3 and M5 employ the same variables but from measurements in the TSI range. Lozano et al. (2022) suggested the possibility to model PAR directly from TSI_{Global} measurements, and these findings confirmed that there are no statistical differences when modeling $PAR_{Diffuse}$ from TSI or PAR measurements.

Figure 2 shows the values of k_{PAR} versus $k_{t,PAR}$ for models M1, M2 and M4 and versus k_t for models M3 and M5, evaluated in Granada. Similar results are observed for Almería and Renon (not shown here). As expected, despite the relatively high values of r^2 and the good statistical results (low MBE and relatively low rRMSE), M1 is unable to reproduce the variability in k_{PAR}. Its apparent good performance is related to the use of the two main factors to model solar radiation (clearness index and solar position). However, a linear model needs to add other variables in the equation to improve the results. In that sense, the rest of the models have a better development because instead of using a linear equation the first two models (M2 and M3) employ a logistic equation meanwhile the latter two (M4 and M5) use a third order polynomial expression with the same variables as in the previous two models.

When comparing models, the logistic and the polynomial models behave quite similarly, as shown in the statistics summarized in Table 4. Focusing on models that include the clearness index in the PAR range (M2 and M4), their r^2 only differs up to 2% in all the locations (validation dataset), and the difference in rRMSE ranges from 0.6% in Granada to 1.5% in Renon. For models including the clearness index in the TSI range (M3 and M5), the differences in their r^2 are 1% in all the locations, and the difference in its rRMSE ranges from 0.2% in Renon to 0.7% in Almería. Finally, focusing on the differences between modeling k_{PAR} from PAR measurements or from TSI range measurements (Model M2 or M4 vs. M3 or M5) the differences in r^2 and rRMSE are also low, and up to 4% and to 2.8%, respectively, both found in Renon. On the other hand, there are only slight differences between modeling k_{PAR} from logistic or polynomial models when considering all the locations. The site-dependant model that
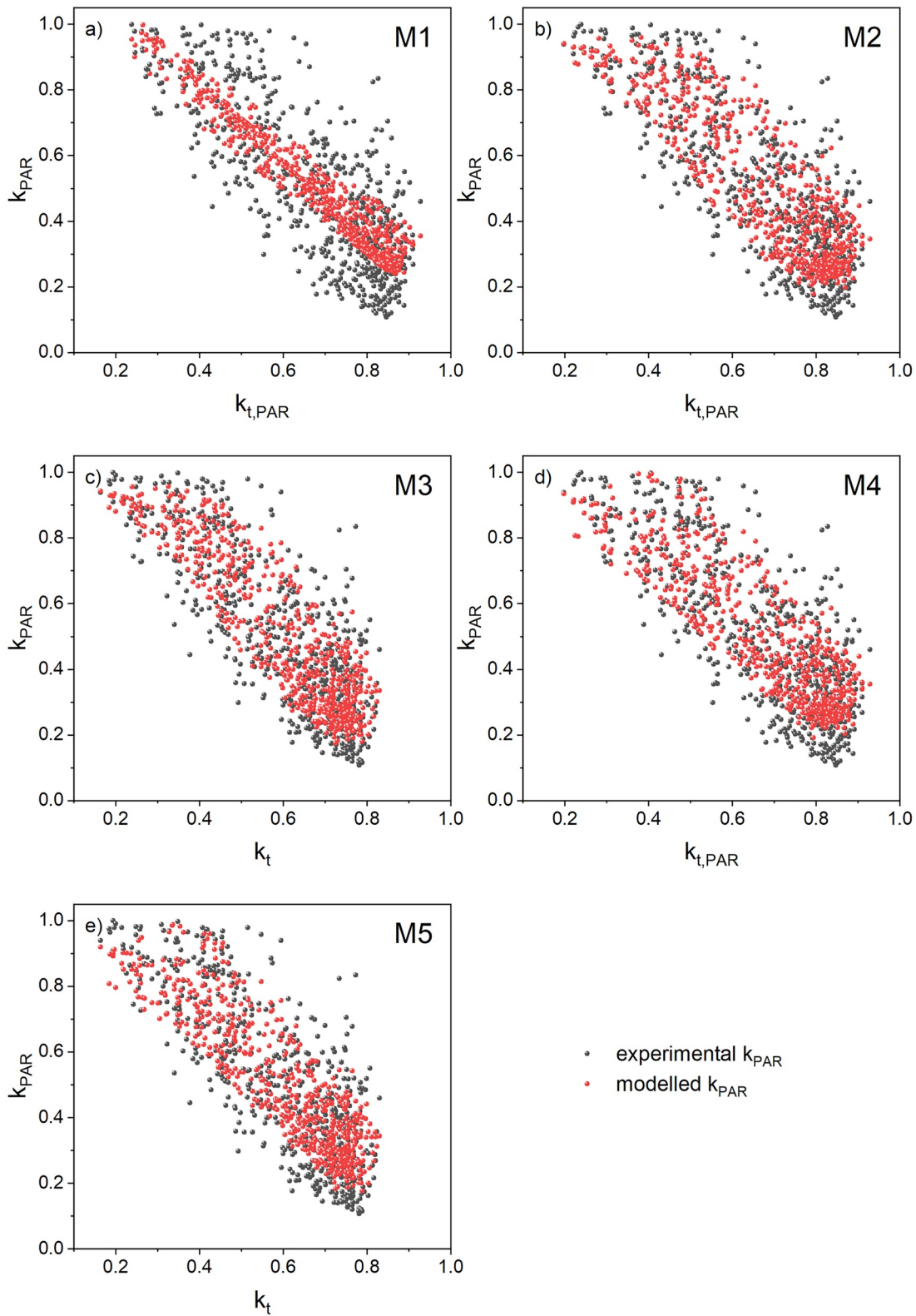
Figure 2. Hourly PAR diffuse fraction (k_{PAR}) modeled (red) and experimental (black) versus hourly clearness index for PAR range ($k_{t,PAR}$), for the empirical models (a) M1, (b) M2, (d) M4, or versus hourly clearness index (k_t), for the empirical models (c) M3, and (e) M5. The validation dataset for Granada was used to build the figures.

Table 5
Statistical Performance for Validation of the ANN Models for Estimating the Diffuse Fraction of the PAR Range, for Granada (GR), Almería (AL), and Renon (RE)

Site	Model	r^2	MBE	rRMSE (%)
GR	ANN1	0.967	0.001	21.0
	ANN2	0.965	−0.001	20.4
	ANN3	0.966	−0.006	20.7
AL	ANN1	0.956	−0.003	24.5
	ANN2	0.951	−0.003	24.9
	ANN3	0.960	−0.009	23.7
RE	ANN1	0.976	−0.007	17.1
	ANN2	0.970	−0.002	19.1
	ANN3	0.980	−0.002	15.6

best fits the experimental measurements was the polynomial using TSI range measurements (M5) reaching a $r^2 = 0.858$ and a rRMSE = 18.3% in Renon.

It is important to point out that there are very few comparative studies of models for k_{PAR} like the one carried out in this work. Thus, for example, in the study carried out by Jacovides et al. (2010) in Greece, only different functional forms are analyzed with $k_{t,PAR}$ as the only independent variable, without considering the variability of the diffuse fraction. Jacovides et al. (2010) analyzed the versions of the models by Reindl et al. (1990) and Ridley et al. (2010) applied them on the PAR range, and obtained a r^2 of 0.87 and 0.90 and a rRMSE of 32.5% and 27.1%, respectively. Kathilankal et al. (2014) also evaluated the same simplified version of the Ridley et al. (2010) model, in the United States, obtaining a r^2 of 0.76 and a rRMSE of 30.6%. Both obtain r^2 values close to those obtained in this work, which confirms the high percentage of variance explained by the functional forms analyzed. However, the rRMSE values are higher than those obtained in this study for models, which shows that it is necessary to introduce additional variables to reproduce the variability of the diffuse fraction derived from different atmospheric conditions.

Table 5 shows the values of the statistics in the validation of the ANN models for the diffuse fraction, k_{PAR}, following the same procedure for the empirical models in the above analysis. All the ANN models have a good performance with its r^2 ranging between 0.967 (ANN1) in Granada and 0.980 (ANN3) in Renon. Again, the percentage of explained variance and the statistics for these ANN models are within the range of other authors, even if we compare the metrics obtained in our work for the diffuse component or k_{PAR} with other studies for the PAR_{Global}. For example, Ferrera-Cobos et al. (2020b) obtained r^2 ranging between 0.992 and 0.998 and a rRMSE between 1.86% and 9.97%, while García-Rodríguez et al. (2022) found values of r^2 from 0.994 to 0.997 and rRMSE values from 4.62% to 6.87%, in addition, López et al. (2001) evaluated a wide range of ANN models with different weather and solar radiation measurements as inputs, as well as PAR_{Global}, obtaining a high range of variation in r^2 (0.337–0.999) and rRMSE (2.0%–44.3%). The fact that our statistics in estimating the diffuse component are within the variation range of other authors predicting global PAR highlight the good performance of these models. MBE is close to zero implying that, for the average amount, there is no overestimation or underestimation of the data. The performance of the three models is very similar in all locations, with very slight differences between themselves, r^2 differs up to 0.01 between the three models in Renon, and the maximum difference between the rRMSE is also 3.5% in Renon.

ANN1 and ANN2 have quite similar results with very important implications, in those situations in which there are no measurements of PAR solar range, TSI through k_t can be used to predict diffuse PAR as can be seen from ANN2. These findings from ANN models are therefore in agreements with the previous findings for the empirical models. A similar conclusion was found by Lozano et al. (2022) with a preliminary evaluation of Ridley et al. (2010) empirical model to obtain k_{PAR} from k_t. Figure 3 shows the values of k_{PAR} versus $k_{t,PAR}$ for model ANN1 and k_{PAR} versus k_t for model ANN2 evaluated in Granada. Similar results are observed for Almería and Renon (not shown here), and despite the fact that ANN models have better statistic performance, it is interesting to see that for high values of k_{PAR} and medium values of $k_{t,PAR}$ or k_t, the empirical logistic and polynomial models better represent the scattering of the cloud points.

4.2. Modeling PAR$_{Diffuse}$

A similar analysis than the developed above for k_{PAR} has been carried out to estimate PAR$_{Diffuse}$. In this section, we analyze one empirical model to estimate PAR$_{Diffuse}$ directly (M6) which is proposed in the literature by Foyo-Moreno et al. (2018). Table 6 shows the statistics of model M6 for both the fitting, and the validation datasets, as well as the fitting coefficient for each location. A quite good performance of this model can be observed by the similar values of the statistics between both subsets of the data. The low values of MBE suggest a slight underestimation at all the cities. However, this is the empirical model with the highest r^2 for the validation data, ranging from 0.981 in Renon to 0.997 in Granada. The estimation of the error for this model (rRMSE) is also the lowest compared with the other empirical models, ranging from 6.1% (Granada) to 15.6% (Renon). Therefore, these rRMSE values are in the same order as those for the above ANN models, highlighting the behavior of this
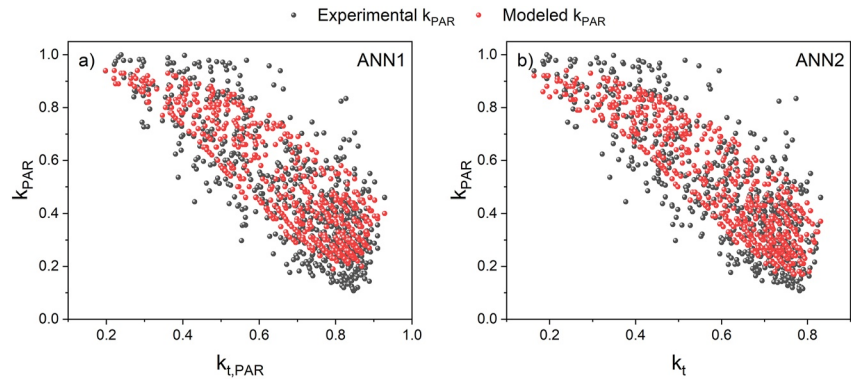
Figure 3. Hourly PAR diffuse fraction (k_{PAR}) modeled (red) and experimental (black) versus (a) hourly clearness index for PAR range ($k_{t,PAR}$; ANN1), or versus (b) hourly clearness index (k_t; ANN2), with the validation dataset in Granada.

model. In addition, it is important to emphasize that this model employs only TSI range measurements to predict the diffuse component of PAR, with the only disadvantage of needing the diffuse component of solar radiation.

Three ANN models have been evaluated by directly estimating the $PAR_{Diffuse}$, and its statistics are summarized in Table 7. As expected, ANND3 has the better performance, since this model uses as an input the $TSI_{Diffuse}$, with r^2 values ranging from 0.989 in Renon to 0.998 in Granada, and rRMSE between 5.1% (Granada) and 11.7% (Renon). However, it is remarkable that despite ANN's better general performance compared to empiric models, overall the ANND3 model and Foyo-Moreno et al. (2018) model (M6) have as similar behavior as can be seen from its statistics. On the other hand, the uncertainty of directly modeling $PAR_{Diffuse}$ instead of using ratios is not observed in ANN models nor in M6 neither, where ANND1 and ANND2 are only slightly worse than the results when modeling from k_{PAR}, furthermore as seen in the previous section for k_{PAR} modeling no improvements were found when modeling $PAR_{Diffuse}$ from $k_{t,PAR}$ instead of k_t.

Figure 4 shows the modeled $PAR_{Diffuse}$ from the empirical model M6 and from ANND1, ANND2 and ANND3, at Granada, including the 1:1 line as a reference. The M6 model shows the estimated $PAR_{Diffuse}$ values had a very low dispersion with respect to the experimental values, and a high r^2 value of 0.997 in Granada. M6 shows less dispersion than ANND1 or ANND2, only ANND3 has a similar good performance. This is a consequence of the fact that its main input variable is diffuse radiation in the TSI range, which is fundamentally affected by the same scattering processes as the PAR interval. This result shows the possibility of obtaining very precise $PAR_{Diffuse}$ values from the values of this component in the TSI spectrum, in a similar way to how PAR_{Global} radiation is obtained (Alados & Alados Arboledas, 1999; Foyo-Moreno et al., 2017).

5. Conclusions

Most of the empirical models that estimate the diffuse component of the photosynthetically active radiation ($PAR_{Diffuse}$) are based on the relationship between the diffuse fraction (k) and clearness index (k_t) and these can be adapted to the PAR range. Thus, this work presents an exhaustive evaluation of different empirical and ANN models which estimate $PAR_{Diffuse}$ using the relationship between k_{PAR} and $k_{t,PAR}$, and also directly estimate $PAR_{Diffuse}$. Our analysis is therefore applicable to other parts of the world after calibrating/fitting the models with

Table 6
Statistical Performance for M6 Model Estimating the Diffuse Component of PAR Corresponding to the Fit Data Set, on the Left, and the Validation Data Set, on the Right, for Granada (GR), Almería (AL), and Renon (RE)

	Fitting				Validation		
Site	r^2	MBE (Wm^{-2})	rRMSE (%)	Coefficient (f)(Wm^{-2})	r^2	MBE (Wm^{-2})	rRMSE (%)
GR	0.996	−1.54	6.8	621.8 ± 0.9	0.997	−2.23	6.1
AL	0.991	−1.50	10.4	632.8 ± 1.0	0.990	−1.58	11.1
RE	0.982	−4.10	15.4	535.7 ± 1.2	0.981	−3.97	15.6

Note. The fitting coefficient of the model is also shown for each location.

Table 7
Statistical Performance for Validation of the ANND Models for Estimating the Diffuse PAR, for Granada (GR), Almería (AL), and Renon (RE)

Site	Model	r^2	MBE (Wm^{-2})	rRMSE (%)
GR	ANND1	0.951	−0.24	24.0
	ANND2	0.952	−0.59	23.6
	ANND3	0.998	−0.55	5.1
AL	ANND1	0.940	0.11	27.0
	ANND2	0.941	−0.05	26.8
	ANND3	0.990	−0.38	11.1
RE	ANND1	0.945	−1.23	26.7
	ANND2	0.939	−1.04	28.0
	ANND3	0.989	0.25	11.7

local inputs. However, the majority of the models are not able to reproduce the great variability that exists in the relationship between k_{PAR} and $k_{t,PAR}$. This work proposes improvements to reproduce this variability and estimate PAR$_{Diffuse}$, which can be applied to optimize crop management practices, increase prediction accuracy, and optimize solar panel performance, and so on. The following main conclusions are derived from this study.

- Empirical models were accurately estimating k_{PAR} when they were based on functional forms that reproduce the existing great variability in the $k_t - k$ relation, which can be a logistic (M2 and M3 models) or a polynomial function (M4 and M5 models), with r^2 ranging from 0.77 to 0.84 for logistic models, and from 0.76 to 0.86 for empirical ones, at our three study sites.

- This work proposes improvements adapting existing models with the TSI range to the PAR range and adding more variables to reproduce the existing variability between $k_t - k$ relation. In fact, the proposal is based on different functional forms adding variables such as α or θ_z, AST, k_t and Ψ, at both ranges (PAR and TSI).

- However, only the empirical models can completely reproduce the variability of experimental measurements if the diffuse component is considered in another range (e.g., in TSI range), since it is an easy way to reproduce
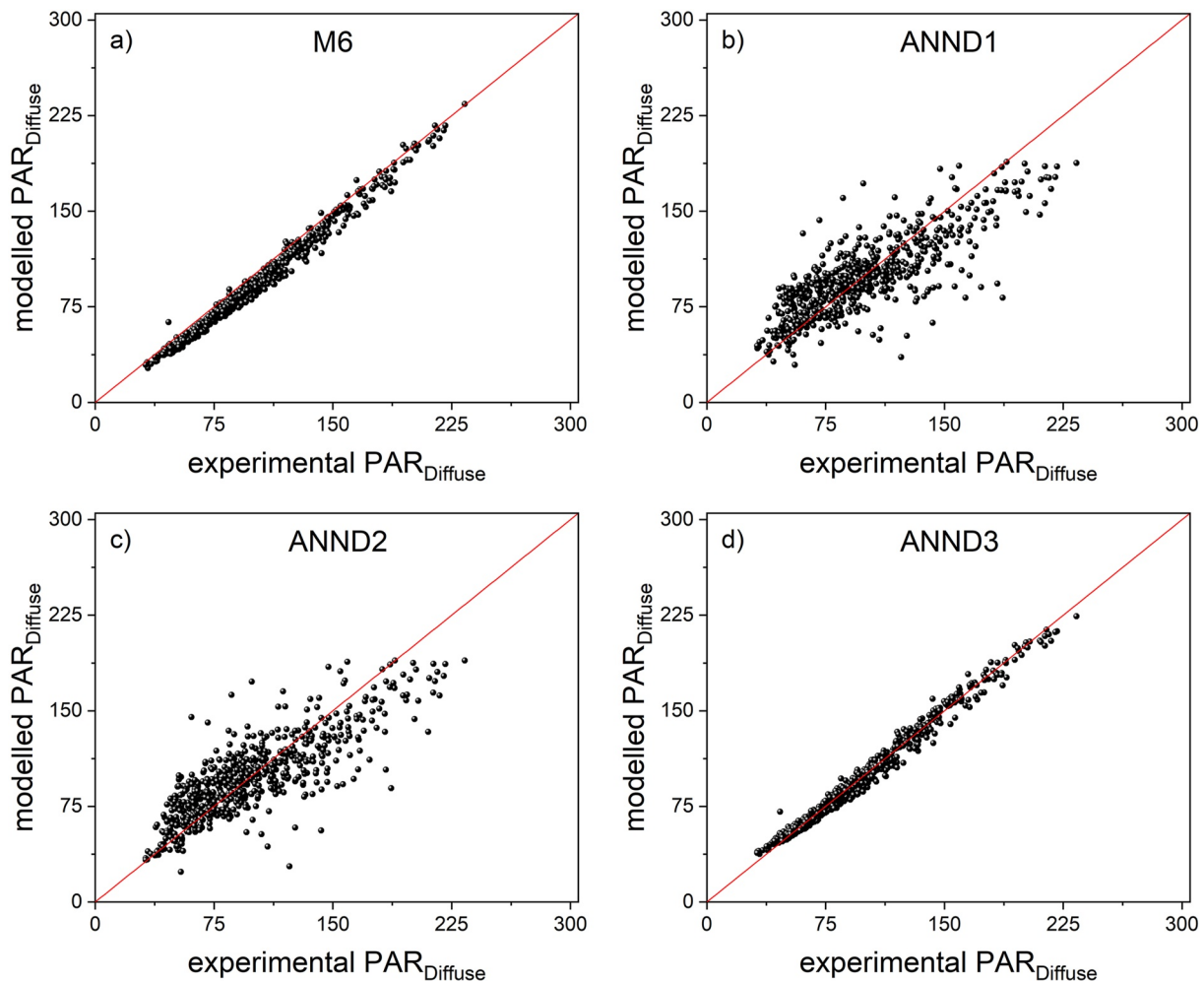

Figure 4. Modeled diffuse photosynthetically active radiation (PAR$_{Diffuse}$), for (a) the empirical model M6, and the ANN models (b) ANND1, (c) ANND2 and (d) ANND3 in Granada, represented against the experimental PAR for the validation dataset. 1:1 line is also shown (red line).

the atmospheric attenuation process for solar radiation. Therefore, the M6 model reached a r^2 up to 0.996 or 0.997, when evaluated from the fitting and validation datasets, respectively.

- ANN models were the best models for estimating k_{PAR}, with r^2 ranging from 0.90 to 0.98. However, the polynomial model M5 reached a r^2 quite close to these ANN models, 0.86 in Renon, which suggests that the polynomial model might be an alternative to ANN models.

- From our $PAR_{Diffuse}$ modeling analysis, it can be concluded that there is no advantage in applying ANN instead of the Foyo-Moreno model, especially when considering the simplicity of this empirical linear model.

- Finally, besides this comparative study among empirical and ANN models two main findings were found: (a) contrary to empirical models, using many variables is not related to an increase in the performance of the models, (b) although modeling k_{PAR} or $PAR_{Diffuse}$ from measurements in the PAR solar range has accurate results, estimations from the TSI range has an even better performance, reaching values of r^2 up to 0.86 when modeling k_{PAR} from empirical models, and up to 0.998 or 0.997 for ANN or empirical models modeling $PAR_{Diffuse}$ directly. These two findings allow us to reproduce diffuse PAR from clearness index and AST, even if PAR measurements are not available, with very precise results.

Glossary

α	Solar elevation in degrees
Ψ	Persistence index
Ψ_{PAR}	Persistence index of PAR
θ_z	Solar zenith angle
ANN	Artificial neural networks
AST	Apparent solar time
DFE	diffuse fertilisation effect
E_0	Eccentricity correction factor of the Earth's orbit
GHGs	Greenhouse gases
I_{SC}	Solar constant
$I_{SC,PAR}$	Solar constant for the PAR range
I_{TOA}	Irradiance at the top of the atmosphere
$I_{TOA,PAR}$	PAR irradiance at the top of the atmosphere
k	Diffuse fraction of total solar irradiance
k_{PAR}	Diffuse fraction in the PAR range
k_t	Clearness index
$k_{t,PAR}$	Clearness index in the PAR range
K'_t	Daily clearness index
$K'_{t,PAR}$	Daily clearness index in PAR range
$k_t^{time-1} \mid k_t^{time+1}$	Clearness index at interval time+1, or time −1
$k_{t,PAR}^{time-1} \mid k_{t,PAR}^{time+1}$	Clearness index of PAR at interval time+1, or time −1
m a.s.l.	Meters above sea level
MBE	Mean bias error
MLP	Multilayer perceptron
PAR	Photosynthetically active radiation (400–700 nm)
$PAR_{Global} \mid PAR_{Diffuse}$	Global solar irradiance, and diffuse component, both for PAR
r^2	Coefficient of determination
rRMSE	Relative root mean square error
TOA	Top of the atmosphere
TSI	Solar irradiance from 280 – 3000 nm
$TSI_{Diffuse}$	Diffuse component in the TSI spectral range
UV	Ultraviolet solar spectral range

Data Availability Statement

Raw measurements from this study are available on request. The datasets of processed measurements used in this manuscript (to produce figures and tables) are publicly available: CSV data tables (Lozano et al., 2023a), Lozano et al., 2023b and modelled and measured diffuse and diffuse fraction of PAR range (Lozano et al., 2023b). Analysis was performed using Matlab R2022b (The MathWorks Inc., 2022).

References

Ağbulut, Ü., Gürel, A. E., & Biçen, Y. (2021). Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison. *Renewable and Sustainable Energy Reviews*, *135*, 110114. https://doi.org/10.1016/j.rser.2020.110114

Alados, I., & Alados Arboledas, L. (1999). Validation of an empirical model for photosynthetically active radiation. *International Journal of Climatology*, *19*, 1145–1152. https://doi.org/10.1002/(sici)1097-0088(199908)19:10<1145::aid-joc428>3.0.co;2

Alados, I., Gomera, M. A., Foyo-Moreno, I., & Alados-Arboledas, L. (2007). Neural network for the estimation of UV erythemal irradiance using solar broadband irradiance. *International Journal of Climatology*, *27*(13), 1791–1799. https://doi.org/10.1002/joc.1496

Alados, I., Mellado, J. A., Ramos, F., & Alados-Arboledas, L. (2004). Estimating UV erythemal irradiance by means of neural networks. *Photochemistry and Photobiology*, *80*(2), 351–358. https://doi.org/10.1111/j.1751-1097.2004.tb00095.x

Alsina, E. F., Bortolini, M., Gamberi, M., & Regattieri, A. (2016). Artificial neural network optimisation for monthly average daily global solar radiation prediction. *Energy Conversion and Management*, *120*, 320–329. https://doi.org/10.1016/j.enconman.2016.04.101

Badarinath, K. V. S., Kharol, S. K., Kaskaoutis, D. G., & Kambezidis, H. D. (2007a). Case study of a dust storm over Hyderabad area, India: Its impact on solar radiation using satellite data and ground measurements. *Science of the Total Environment*, *384*(1–3), 316–332. https://doi.org/10.1016/j.scitotenv.2007.05.031

Badarinath, K. V. S., Kharol, S. K., Kaskaoutis, D. G., & Kambezidis, H. D. (2007b). Influence of atmospheric aerosols on solar spectral irradiance in an urban area. *Journal of Atmospheric and Solar-Terrestrial Physics*, *69*(4–5), 589–599. https://doi.org/10.1016/j.jastp.2006.10.010

Batlles, F. J., Olmo, F. J., & Alados-Arboledas, L. (1995). On shadowband correction methods for diffuse irradiance measurements. *Solar Energy*, *54*(2), 105–114. https://doi.org/10.1016/0038-092X(94)00115-T

Bauerle, W. L., Daniels, A. B., & Barnard, D. M. (2014). Carbon and water flux responses to physiology by environment interactions: A sensitivity analysis of variation in climate on photosynthetic and stomatal parameters. *Climate Dynamics*, *42*(9–10), 2539–2554. https://doi.org/10.1007/s00382-013-1894-6

De Miguel, A., Bilbao, J., Aguiar, R., Kambezidis, H., & Negro, E. (2001). Diffuse solar irradiation model evaluation in the North Mediterranean Belt area. *Solar Energy*, *70*(2), 143–153. https://doi.org/10.1016/S0038-092X(00)00135-3

Desai, M., & Shah, M. (2021). An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural network (MLP) and Convolutional neural network (CNN). *Clinical eHealth*, *4*, 1–11. https://doi.org/10.1016/j.ceh.2020.11.002

Di Biagio, C., Di Sarra, A., Meloni, D., Monteleone, F., Piacentino, S., & Sferlazzo, D. (2009). Measurements of Mediterranean aerosol radiative forcing and influence of the single scattering albedo. *Journal of Geophysical Research*, *114*(D6), D06211. https://doi.org/10.1029/2008JD011037

Dou, M., Liang, L., Han, Y., Jia, R., & Zhang, Y. (2021). Eutrophication model driven by light and nutrients condition change in sluice-controlled river reaches. *Environmental Science & Pollution Research*, *28*(43), 61647–61664. https://doi.org/10.1007/s11356-021-15002-1

Elsheikh, A. H., Sharshir, S. W., Abd Elaziz, M., Kabeel, A. E., Guilan, W., & Haiou, Z. (2019). Modeling of solar energy systems using artificial neural network: A comprehensive review. *Solar Energy*, *180*, 622–639. https://doi.org/10.1016/j.solener.2019.01.037

Engerer, N. A. (2015). Minute resolution estimates of the diffuse fraction of global irradiance for southeastern Australia. *Solar Energy*, *116*, 215–237. https://doi.org/10.1016/j.solener.2015.04.012

Ferrera-Cobos, F., Vindel, J. M., Valenzuela, R. X., & González, J. A. (2020a). Analysis of Spatial and temporal variability of the PAR/GHI ratio and PAR modeling based on two satellite estimates. *Remote Sensing*, *12*(8), 1262. https://doi.org/10.3390/rs12081262

Ferrera-Cobos, F., Vindel, J. M., Valenzuela, R. X., & González, J. A. (2020b). Models for estimating daily photosynthetically active radiation in oceanic and mediterranean climates and their improvement by site adaptation techniques. *Advances in Space Research*, *65*(8), 1894–1909. https://doi.org/10.1016/j.asr.2020.01.018

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. *International Journal of Climatology*, *37*(S1), 1067–1075. https://doi.org/10.1002/joc.5063

Foyo-Moreno, I., Alados, I., & Alados-Arboledas, L. (2018). A new empirical model to estimate hourly diffuse photosynthetic photon flux density. *Atmospheric Research*, *203*, 189–196. https://doi.org/10.1016/j.atmosres.2017.12.012

García-Rodríguez, A., García-Rodríguez, S., Granados-López, D., Díez-Mediavilla, M., & Alonso-Tristán, C. (2022). Extension of PAR models under local all-sky conditions to different climatic Zones. *Applied Sciences*, *12*(5), 2372. https://doi.org/10.3390/app12052372

Ghritlahre, H. K., & Prasad, R. K. (2018). Exergetic performance prediction of solar air heater using MLP, GRNN and RBF models of artificial neural network technique. *Journal of Environmental Management*, *223*, 566–575. https://doi.org/10.1016/j.jenvman.2018.06.033

Gradoville, M. R., Cabello, A. M., Wilson, S. T., Turk-Kubo, K. A., Karl, D. M., & Zehr, J. P. (2021). Light and depth dependency of nitrogen fixation by the non-photosynthetic, symbiotic cyanobacterium UCYN-A. *Environmental Microbiology*, *23*(8), 4518–4531. https://doi.org/10.1111/1462-2920.15645

Gu, L. H., Baldocchi, D., Verma, S. B., Black, T. A., Vesala, T., Falge, E. M., & Dowty, P. R. (2002). Advantages of diffuse radiation for terrestrial ecosystem productivity. *Journal of Geophysical Research*, *107*(D6), 4050. https://doi.org/10.1029/2001JD001242

Gueymard, C. A. (2018). Revised composite extraterrestrial spectrum based on recent solar irradiance observations. *Solar Energy*, *169*, 434–440. https://doi.org/10.1016/j.solener.2018.04.067

Gui, X., Wang, L., Su, X., Yi, X., Chen, X., Yao, R., & Wang, S. (2021). Environmental factors modulate the diffuse fertilization effect on gross primary productivity across Chinese ecosystems. *Science of the Total Environment*, *793*, 148443. https://doi.org/10.1016/j.scitotenv.2021.148443

Hao, D., Asrar, G. R., Zeng, Y., Zhu, Q., Wen, J., Xiao, Q., & Chen, M. (2019). Estimating hourly land surface downward shortwave and photosynthetically active radiation from DSCOVR/EPIC observations. *Remote Sensing of Environment*, *232*, 111320. https://doi.org/10.1016/j.rse.2019.111320

Hasni, A., Sehli, A., Draoui, B., Bassou, A., & Amieur, B. (2012). Estimating global solar radiation using artificial neural network and climate data in the South-western region of Algeria. *Energy Procedia, Terragreen 2012: Clean Energy Solutions for Sustainable Environment (CESSE)*, *18*, 531–537. https://doi.org/10.1016/j.egypro.2012.05.064

Iqbal, M. (1983). An introduction to solar-radiation. *Bulletin of the American Meteorological Society*, *65*(10). Article 10.

Jacovides, C. P., Boland, J., Asimakopoulos, D. N., & Kaltsounides, N. A. (2010). Comparing diffuse radiation models with one predictor for partitioning incident PAR radiation into its diffuse component in the eastern Mediterranean basin. *Renewable Energy*, *35*(8), 1820–1827. https://doi.org/10.1016/j.renene.2009.11.015

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. *Advances in Space Research*, *56*(11), 2356–2364. https://doi.org/10.1016/j.asr.2015.09.020

Kamadinata, J. O., Ken, T. L., & Suwa, T. (2019). Sky image-based solar irradiance prediction methodologies using artificial neural networks. *Renewable Energy*, *134*, 837–845. https://doi.org/10.1016/j.renene.2018.11.056

Kanniah, K. D., Beringer, J., North, P., & Hutley, L. (2012). Control of atmospheric particles on diffuse radiation and terrestrial plant productivity: A review. *Progress in Physical Geography: Earth and Environment*, *36*(2), 209–237. https://doi.org/10.1177/0309133311434244

Kathilankal, J. C., O'Halloran, T. L., Schmidt, A., Hanson, C. V., & Law, B. E. (2014). Development of a semi-parametric PAR (Photosynthetically Active Radiation) partitioning model for the United States, version 1.0. *Geoscientific Model Development*, *7*(5), 2477–2484. https://doi.org/10.5194/gmd-7-2477-2014

Keane, B. J., Ineson, P., Vallack, H. W., Blei, E., Bentley, M., Howarth, S., et al. (2017). Greenhouse gas emissions from the energy crop oilseed rape (Brassica napus); the role of photosynthetically active radiation in diurnal N2O flux variation. *GCB Bioenergy*, *10*(5), 306–319. https://doi.org/10.1111/gcbb.12491

Kipp, Z., & Zonen, B. V. (2000). *CNR1 net radiometer instruction manual*. Kipp & Zonen.

López, G., Rubio, M. A., Martínez, M., & Batlles, F. J. (2001). Estimation of hourly global photosynthetically active radiation using artificial neural network models. *Agricultural and Forest Meteorology*, *107*(4), 279–291. https://doi.org/10.1016/S0168-1923(01)00217-9

Lozano, I., Alados, I., Sánchez-Hernández, G., Guerrero-Rascado, J. L., & Foyo-Moreno, I. (2023a). CSV data tables [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7941129

Lozano, I., Alados, I., Sánchez-Hernández, G., Guerrero-Rascado, J. L., & Foyo-Moreno, I. (2023b). Modeled and measured diffuse and diffuse fraction of PAR range [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7941471

Lozano, I. L., Sánchez-Hernández, G., Guerrero-Rascado, J. L., Alados, I., & Foyo-Moreno, I. (2021). Aerosol radiative effects in photosynthetically active radiation and total irradiance at a Mediterranean site from an 11-year database. *Atmospheric Research*, *255*, 105538. https://doi.org/10.1016/j.atmosres.2021.105538

Lozano, I. L., Sánchez-Hernández, G., Guerrero-Rascado, J. L., Alados, I., & Foyo-Moreno, I. (2022). Analysis of cloud effects on long-term global and diffuse photosynthetically active radiation at a Mediterranean site. *Atmospheric Research*, *268*, 106010. https://doi.org/10.1016/j.atmosres.2021.106010

Ma, C. C. Y., & Iqbal, M. (1984). Statistical comparison of solar radiation correlations Monthly average global and diffuse radiation on horizontal surfaces. *Solar Energy*, *33*(2), 143–148. https://doi.org/10.1016/0038-092X(84)90231-7

McCree, K. J. (1972). Test of current definitions of photosynthetically active radiation against leaf photosynthesis data. *Agricultural Meteorology*, *10*, 443–453. https://doi.org/10.1016/0002-1571(72)90045-3

Meloni, D., Di Sarra, A., Pace, G., & Monteleone, F. (2006). Aerosol optical properties at Lampedusa (Central Mediterranean). 2. Determination of single scattering albedo at two wavelengths for different aerosol types. *Atmospheric Chemistry and Physics*, *6*(3), 715–727. https://doi.org/10.5194/acp-6-715-2006

Mercado, L. M., Bellouin, N., Sitch, S., Boucher, O., Huntingford, C., Wild, M., & Cox, P. M. (2009). Impact of changes in diffuse radiation on the global land carbon sink. *Nature*, *458*(7241), 1014–1017. https://doi.org/10.1038/nature07949

Niu, Z., Wang, L., Niu, Y., Hu, B., Zhang, M., & Qin, W. (2019). Spatiotemporal variations of photosynthetically active radiation and the influencing factors in China from 1961 to 2016. *Theoretical and Applied Climatology*, *137*(3–4), 2049–2067. https://doi.org/10.1007/s00704-018-2727-7

Peng, S., Du, Q., Lin, A., Hu, B., Xiao, K., & Xi, Y. (2015). Observation and estimation of photosynthetically active radiation in Lhasa (Tibetan Plateau). *Advances in Space Research*, *55*(6), 1604–1612. https://doi.org/10.1016/j.asr.2015.01.002

Reindl, D. T., Beckman, W. A., & Duffie, J. A. (1990). Diffuse fraction correlations. *Solar Energy*, *45*, 1–7. https://doi.org/10.1016/0038-092X(90)90060-P

Ridley, B., Boland, J., & Lauret, P. (2010). Modelling of diffuse solar fraction with multiple predictors. *Renewable Energy*, *35*(2), 478–483. https://doi.org/10.1016/j.renene.2009.07.018

Sanchez, G., Serrano, A., & Cancillo, M. L. (2017). Modeling the erythemal surface diffuse irradiance fraction for Badajoz, Spain. *Atmospheric Chemistry and Physics*, *17*(20), 12697–12708. https://doi.org/10.5194/acp-17-12697-2017

The MathWorks Inc. (2022). *MATLAB version: 9.13.0 (version R2022b)*. The MathWorks Inc. Retrieved from https://www.mathworks.com

Trisolino, P., Di Sarra, A., Meloni, D., & Pace, G. (2016). Determination of global and diffuse photosynthetically active radiation from a multi-filter shadowband radiometer. *Applied Optics*, *55*(29), 8280–8286. https://doi.org/10.1364/AO.55.008280

Verma, S., Chowdhury, R., Das, S. K., Franchetti, M. J., & Liu, G. (2021). Sunlight intensity, photosynthetically active radiation modelling and its application in algae-based wastewater treatment and its Cost estimation. *Sustainability*, *13*(21), 11937. https://doi.org/10.3390/su132111937

Vindel, J. M., Valenzuela, R. X., Navarro, A. A., Zarzalejo, L. F., Paz-Gallardo, A., Souto, J. A., et al. (2018). Modeling photosynthetically active radiation from satellite-derived estimations over Mainland Spain. *Remote Sensing*, *10*(6), 849. https://doi.org/10.3390/rs10060849

Wang, L., Kisi, O., Zounemat-Kermani, M., Hu, B., & Gong, W. (2016). Modeling and comparison of hourly photosynthetically active radiation in different ecosystems. *Renewable and Sustainable Energy Reviews*, *56*, 436–453. https://doi.org/10.1016/j.rser.2015.11.068

Yang, D., & Gueymard, C. A. (2020). Ensemble model output statistics for the separation of direct and diffuse components from 1-min global irradiance. *Solar Energy*, *208*, 591–603. https://doi.org/10.1016/j.solener.2020.05.082

Yu, X., Wu, Z., Jiang, W., & Guo, X. (2015). Predicting daily photosynthetically active radiation from global solar radiation in the Contiguous United States. *Energy Conversion and Management*, *89*, 71–82. https://doi.org/10.1016/j.enconman.2014.09.038

Zhou, H., Yue, X., Lei, Y., Zhang, T., Tian, C., Ma, Y., & Cao, Y. (2021). Responses of gross primary productivity to diffuse radiation at global FLUXNET sites. *Atmospheric Environment*, *244*, 117905. https://doi.org/10.1016/j.atmosenv.2020.117905

Ziółkowski, J., Oszczypała, M., Małachowski, J., & Szkutnik-Rogoż, J. (2021). Use of artificial neural networks to predict fuel consumption on the basis of technical parameters of vehicles. *Energies*, *14*(9), 2639. https://doi.org/10.3390/en14092639