Predictive models for photosynthetic active radiation irradiance in temperate climates

Yazan J.K. Musleh [*], Tasmiat Rahman

University of Southampton, University Road, Southampton, SO17 1BJ, Hampshire, United Kingdom

ABSTRACT

This research evaluated 10 different empirical models designed for predicting Photosynthetically Active Radiation (PAR) at higher latitudes, addressing atmospheric conditions specific to these regions. The research introduces the Musleh-Rahman (MR) model, which substitutes Diffuse Horizontal Irradiance (DHI) with Clear Direct Normal Irradiance (DNI), Ozone and Aerosol Optical Depth at 550 nm (AOD550) sourced for satellite reanalysis data, achieving a Mean Bias Deviation (MBD) of 0.19 % and Root Mean Square Error (RMSE) of 12.42 W/m^2. Furthermore, when applied to six untested locations, results demonstrate that the MR model outperformed the best performing empirical model with an MBD improvement of 3.68 % and an RMSE of 4.28 W/m^2, whereas, when compared to machine learning models, the Light Gradient Boost Model (LGBM), had an MBD of -3.85 %. The MR model also maintained consistency across seasonal and density evaluations, attaining an R^2 value as high as 0.9709, thereby highlighting the significant benefits of integrating satellite-sourced atmospheric data into PAR prediction models. Moreover, the research illustrated that substituting DHI with Clear DNI, Ozone, and AOD550 not only reduces MBD and boosts R^2 values but also amplifies the model's applicability and accuracy in capturing early PAR peaks and reducing overestimations through precise adjustments in Ozone and AOD550 levels. This highlights the benefits of incorporating satellite-derived atmospheric data into PAR predictions models.

1. Introduction

The concept of the food-energy-water nexus underscores the interconnectedness of water, energy, and food security, emphasizing the urgent need for beneficial strategies that align agricultural and energy policies [1]. AgriPhotovoltaics (AgriPV) presents a notable solution, allowing for both crop cultivation and energy production to occur on the same piece of land simultaneously [2]. This coexistence potentially enhances crop yields through the shared use of sunlight, depending on climatic conditions [3] and crop varieties [4]. In certain instances, this arrangement proves mutually advantageous, as the evaporation from crops can lower the operating temperatures of PV modules [5,6]. While AgriPV has garnered increasing interest [7,8], challenges persist due to the lack of comprehensive parameters and models. Despite its potential, the AgriPV sector remains niche, constrained by scarce data and installations that could broaden its impact. Moreover, the growing interest in PV systems has led to land competition, prompting investigations into land dual-use as a solution, especially in regions where land is at a premium [7]. Precise calculation of available irradiance for crops is

essential to determine the viability of AgriPV, utilizing metrics such as Levelized Cost of Electricity (LCOE) or Land Equivalent Ratios (LER). Consequently, this necessitates modeling, which hinges on accurate assessments or predictions of irradiance components relevant to photovoltaics and agriculture [9,10].

Tackling the task of accurate AgriPV potential requires accurate modelling of Photosynthetically Active Radiation (PAR). PAR constitutes a key component of incoming Global Horizontal Irradiance (GHI), reaching the Earth's surface, spanning wavelengths from 400 to 700 nm. The integration of PV panels into agricultural lands necessitates ensuring that crop yields do not fall below acceptable thresholds, thus underlining the importance PAR estimations [11]. Photosynthesis rates hinge on the energy accessible to plants and are expressed either in terms of flux density (i.e., μml photons/m^2/s) or in terms of flux density (i.e., irradiance W/m^2) [12]. Photodetectors are used to record PAR, though these measurements are less prevalent compared to GHI data collection, where the latter is gathered either using pyranometers at meteorological stations or may be estimated through satellite imagery [13].

The comprehensive analysis by Nwokolo et al. provides a detailed examination of empirical approaches for predicting PAR [14]. Following said publication, progress has been made in forecasting PAR by utilizing GHI along with various parameters. Wang demonstrated the impact of sky clarity on PAR across China, introducing a validated model in two central locations [15]. Similarly, Peng et al. developed a versatile PAR model suitable for temperate climates in China [16]. Wang further elaborated on this by incorporating solar geometry into their research, covering 39 different locations on both hourly and daily scales [17]. Furthermore, Aguiar ventured into modeling PAR in Southwest Amazonia, employing methods ranging from simple ratios to complex formulas integrating sky clarity and precipitable water, suitable for all sky conditions [18]. Ferrera-Cobos explored 22 models for estimating daily PAR in oceanic and Mediterranean climates, employing site adaptation techniques and variables like relative humidity, temperature, and extraterrestrial irradiance [19]. Their research compared regression models and a machine learning (ML) approach, finding negligible differences in performance. Similarly, Proutsos et al. examined a high-altitude Mediterranean forest in Greece, developing models for estimating Ultraviolet radiation and PAR based on hourly data, highlighting the importance of optical thickness and temperature [20]. Similarly, Escobedo et al. differentiated parts of the electromagnetic spectrum, in Brazilian climates using daily observations [21]. Furthermore, Akitsu et al. explored models incorporating GHI, precipitable water, and pressure alongside sky clarity under all sky conditions to demonstrate the interplay between said parameters and PAR [22]. These models, however, showed reduced accuracy during winter and in aerosol-dense atmospheres; a common feature in European temperate climates.

Predictive models are crucial for estimating outcomes across various forecasting methodologies, whether in regression modeling or ML [61, 62]. Rogers et al. utilized PAR measurements to derive leaf area index via four distinct methodologies, noting issues with irregular temporal intervals in data collection, which included linear, logistic regression, and least squares regression methods [53]. Yildiz et al. provided an overview of various regression and machine learning models, such as regression trees, support vector regression, and artificial neural networks [54]. Furthermore, Junior et al. explored four machine learning techniques, including k-Nearest Neighbors (kNN), XGBoost (XGB), and Light Gradient Boosting Machine (LGBM), both individually and in a stacked configuration, emphasizing the importance of hyperparameter tuning to prevent overfitting [55]. Murphy et al. applied several ML and regression models, including logistic and linear regression [56]. Building on that, the use of Artificial Neural Networks (ANN) in modeling different parameters in an agricultural setting was discussed as a superior method to address non-linearity and time variance in greenhouse systems [57]. Expanding on ANN, Long Short-term Memory (LSTM) neural networks have been implemented in various forecasting disciplines to support different United Nations Sustainability Development Goals, from PV [58,59] to freshwater accessibility [60].

Research indicates that models incorporating Diffuse Horizontal Irradiance (DHI) or employing Perez coefficients [23], which modify DHI application, tend to show enhanced performance [24–27]. Garcia-Rodriguez et al. incorporated meteorological indices for PAR modeling, using the International Commission on Illumination (CIE) [28] standard sky classification alongside Perez brightness and clearness coefficients for all sky conditions [29]. Dependence on DHI and related metrics introduces complications due to their infrequent measurement. In cases of unavailability, the requirement for specialized decomposition models becomes apparent, as these models significantly vary based on temporal resolution and geographic location. Lu et al. emphasize the importance of advancing PAR prediction methodologies across various global climates [11].

Consequently, there is an imperative need to derive PAR estimates from GHI and other available parameters. Conversion ratios from GHI to PAR under clear skies reveal dependencies on the sky's optical path, but this relationship grows complex under overcast conditions or varying weather, involving numerous parameters for accurate PAR conversion. Numerous investigations across diverse climatic conditions have delved into the conversion of GHI to PAR, however, the applicability of these models to different environments, especially within temperate climates, has been minimally examined [30]. Furthermore, these models are needed to be used to predict its diffuse component [11]. This gap is notable because models tested in diverse climates may not account for the distinct sun path effects critical for model accuracy in higher latitude, temperate European climates. These latitudinal differences significantly impact sun angles, affecting model precision. Additionally, atmospheric variations, including aerosol optical depth, differ by location, influencing the scattering and absorption of solar radiation [31] and, by extension, PAR values. Furthermore, the effect of cloud cover on GHI and subsequently PAR is significant, with models often neglecting regional variations in cloudiness caused by local weather patterns. This leads to potential inaccuracies in predicting solar irradiance [32]. Moreover, the reliance on limited datasets, focusing mainly on daily or hourly solar irradiance for specific climates, underscores the need for more comprehensive models that consider the diverse factors influencing PAR in a sub-hourly context.

This work embarks on a detailed evaluation of 10 recent empirical models designed for PAR irradiance prediction and introduces a new model for PAR prediction tailored to temperate climates. This innovative approach does not rely on the conventional use of DHI, instead, it leverages new inputs such as Clear-sky irradiance, Aerosol Optical Depth, and Ozone profile estimations from satellite imagery. Furthermore, the work explores the use of 4 ML models for PAR estimation, that are tuned with optimized hyperparameters using 9 locations. A detailed statistical analysis is conducted to evaluate the prediction accuracy and scalability of all models across 6 unseen locations. Attention centers on their precision in accurately predicting unfamiliar locales, underscoring scalability, and applicability across diverse European contexts [33]. This enhances the accessibility of PAR estimations, thereby supporting the market expansion of AgriPV, where PAR data can be integrated into LCOE or LER calculations.

2. Methodology

1. Data

The Integrated Carbon Observation System (ICOS) network [34], a European research infrastructure, encompasses multiple stations across Europe, each adhering to strict protocols and standards for data measurement. For GHI measurements, these stations employ Spectrally Flat Class A pyranometers, in accordance with the International Organization for Standardization (ISO) 9060:1990 [35]. PAR measurements are conducted using photodetectors to measure PPF, adhering to specific requirements as outlined in Carra et al. [36]. The measurement of PAR utilizes a conversion factor of 4.57 μ mol/m^2/s to translate these measurements into watts per square meter (W/m^2), aligning with methodologies established in previous studies [6,11].

A total of 14 locations from ICOS were selected. One additional location from the Baseline Surface Radiation Network (BSRN) in Tõravere is included [37], resulting in 15 locations altogether as per Fig. 1 and Table 1. Given the scarcity of simultaneous measurements of PAR and GHI, these specific weather stations that record both parameters concurrently and adhere to the aforementioned standards were chosen. The open-source dataset was selected for its reliability and relevance to the study of PAR. All data was recorded minutely but averaged over a 30-min period. For model training and validation, 9 of the 15 locations are highlighted, with the remaining 6 reserved exclusively for model testing. This distinction ensures a thorough evaluation of model performance on unseen data, particularly focusing on their applicability and scalability within European climates. Such an approach underscores the importance of a rigorous evaluation process,
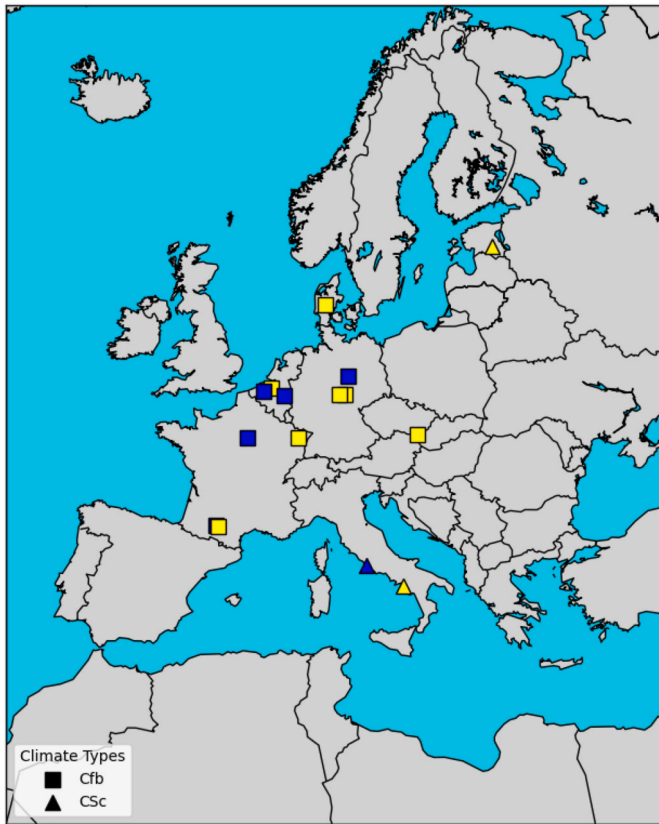
Fig. 1. The spatial layout of the 15 chosen locations includes 9 sites marked in yellow for the purpose of model training and validation, and 6 sites indicated in blue dedicated to evaluating the models against new data. Among these, 12 sites are classified under Temperate Oceanic Climates (Cfb), while 3 sites fall within the Cool Summer Mediterranean Climate category (Csc) as per the Köppen–Geiger climate classification [39].

considering the models' robustness, which is critical [32]. However, the independence of some locations was limited due to proximity, potentially introducing bias in the development of ML models and mathematical regression analyses. Therefore, the dataset was partitioned based on randomly sampled days, with an 80–20 percentage split for training and validation, respectively, following practices as recommended in the literature [38].

Data from locations in Table 1 were synchronized to Coordinated Universal Time (UTC), incorporating the solar zenith angle (referred to

as SZA) and extraterrestrial irradiance (E_a), alongside the solar constant, established at 1361.1 W/m^2 according to Ref. [40]. This information was obtained using the updated Solar Position Algorithm (SPA) [41], which demonstrates a minimal uncertainty of only 0.0003° spanning from the year 2000–6000. Consequently, this allows for the quantification of sky clarity through the sky clearness index, K_t, as outlined in equation (1).

$$K_t = \frac{GHI}{E_a} \tag{1}$$

Although definitive quality control protocols for GHI and DHI measurements have not been established, let alone PAR, insights from the operation of the BSRN [36] have led to proposing a comprehensive set of guidelines to identify unreliable data [42]. These guidelines serve as the foundation for filtering solar irradiance data prior to analysis or model development, outlined as follows.

1. $GHI > 0, DHI > 0$ & $PAR > 0$
2. $GHI > PAR$
3. $GHI \leq 1.05\, E_a \cos^{1.2}(SZA) + 100$
4. $DHI \leq 0.95\, E_a \cos^{1.2}(SZA) + 50$
5. $GHI > DHI$
6. $SZA < 85°$

2. Present Models

Recent advancements in PAR estimation models have introduced a variety of approaches and parameters, as detailed in Table 2. A consistent feature across these models is their reliance on GHI, K_t, and, frequently, SZA. These elements are typically derived from SPA through calculations including GHI measurements; hence are easily attainable.

Among the 10 models evaluated, the Tan-Ismail model stands out for its simplicity [43]. It proposes a straightforward conversion ratio over a one-year period in Singapore. Escobedo et al. crafted a model that categorizes K_t into distinct segments [14,21]. This technique draws inspiration from the segmentation of DHI from GHI in numerous decomposition models, but it uniquely applies this strategy to extract PAR from GHI based on varied sky conditions over 4 years in Botucatu, Brazil. Akitsu et al. have contributed two models to the field [22]. The first, Akitsu1, integrates only GHI and vapor pressure, while the second, Akitsu2, also incorporates K_t. Similarly, Peng and Wang have developed models that utilize a piecewise mathematical function, with coefficients as functions of K_t [15,16]. These models also account for SZA, observing an exponential increase in hourly PAR for specific K_t intervals. Despite sharing a common equation (Equation (2)), the models diverge in their coefficient weighting systems (labeled as a - e), showcasing a tailored

Table 1

Details on the 15 sites selected for assessing PAR, with the first 10 sites are selected for the training and validation of models, whereas the remaining 5 sites serve the purpose of model testing. The provided average values of PAR and GHI pertain solely to data points deemed valid.

Station	Latitude	Longitude	Elevation	Period	Climate	Mean PAR	Mean GHI
	(°)	(°)	(m)		Classification	(W/m^2)	(W/m^2)
Brasschaat	51.307	4.520	16.0	2019–2023	Cfb	139.83	324.29
Gebesee	51.100	10.915	161.5	2020–2023	Cfb	134.24	312.41
Hainich	51.079	10.452	438.7	2019–2023	Cfb	137.75	317.46
Hesse	48.674	7.065	310.0	2021–2023	Cfb	190.16	410.28
Lamasquere	43.496	1.238	181.0	2020–2023	Cfb	213.13	449.92
Lanzhot	48.682	16.946	150.0	2022–2023	Cfb	157.19	362.31
Voulundgaard	56.038	9.161	67.7	2020–2023	Cfb	117.80	286.99
Borgo Cioffi	40.524	14.957	10.0	2023	Csa	227.76	535.14
Tõravere	58.264	26.462	70.0	2016–2019	Csa	115.92	279.47
Aurade	43.550	1.106	250.0	2019–2023	Cfb	197.16	427.11
Fontainebleau-Barbeau	48.476	2.781	103.0	2023	Cfb	178.77	368.97
Hohes Holz	52.086	11.222	193.0	2019–2023	Cfb	123.19	299.17
Lochristi	51.112	3.850	6.30	2019–2022	Cfb	152.92	340.86
Maasmechelen	50.980	5.631	87.0	2020–2023	Cfb	166.61	352.02
Castelporziano	41.704	12.357	19.0	2021–2023	Csa	244.28	481.57

Table 2

Details on the parameters used across the 10 PAR models under evaluation. Alados refers to Model 1 in Ref. [24].

Model	GHI	K_t	Pressure	SZA	Δ	ε	T_d
Tan-Ismail	✓						
Escobedo	✓	✓					
Akitsu 1	✓	✓					
Akitsu 2	✓	✓	✓				
Peng	✓	✓		✓			
Wang	✓	✓		✓			
Hu	✓			✓	✓		
Jacovides		✓			✓	✓	
Garcia-Rodriguez	✓	✓		✓	✓	✓	
Alados	✓			✓	✓	✓	✓

approach to PAR estimation based on varying K_t values.

$$PAR = (a + bK_t + cK_t^2 + Kd_t^3)\cdot\cos(SZA)^e \qquad (2)$$

Table 2 reveals that the latter five models utilize DHI or Perez equations, with Equations (3) and (4) mathematically defining clearness and brightness coefficients, respectively. Direct Normal Irradiance (DNI), a crucial component in these models, can be directly measured with a Pyrheliometer or estimated through Equation (5), which employs a closure equation [44]. Additionally, the optical air mass is denoted as m.

$$\varepsilon = \frac{\frac{DHI+DNI}{DHI} + 1.041(SZA)^3}{1 + 1.041(SZA)^3} \qquad (3)$$

$$\Delta = \frac{m\cdot DHI}{E_a} \qquad (4)$$

$$GHI = DHI + DNIcos(SZA) \qquad (5)$$

The estimated dew point temperature (T_d) was calculated utilizing Equation (6), which integrates relative humidity as a percentage (RH) and air temperature in ∘C (T). This method was adopted based on recommendations identified through a comprehensive review of the literature [45].

$$T_d = \frac{243.04\left(\ln\left(\frac{RH}{100}\right) + \frac{17.625T}{T+243.04}\right)}{17.625 - \ln\left(\frac{RH}{100}\right) - \frac{17.625T}{T+243.04}} \qquad (6)$$

3. New parameters

A clear objective of this research is to identify new parameters that directly influence PAR, with a focus on variables that are readily accessible and not derived empirically through decomposition models, such is the case if DHI measurements are missing. An integral part of this research involves leveraging clear sky irradiances, specifically GHI, DNI, and DHI under clear conditions. A key resource utilized in this investigation is McClear [46,47], which is favored for its open accessibility, requiring only the input of latitude, longitude, and elevation. It provides irradiance that would be observed at any given location worldwide under cloud-free conditions, with data granularity ranging from 1 min to one month. This service offers comprehensive data from 2004 to two days before the current date. The Clear DNI was used as that will include both clear GHI, and clear DHI as per equation (5).

Additionally, to explore further variables potentially affecting PAR estimations, this work incorporates satellite data from the Modern-Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2), courtesy of NASA's Global Modeling and Assimilation Office [48,49] as outlined in Table 3. This data source is pivotal in maintaining the work's standards for integrity and analytical performance. For intricate details on computational approaches and methodologies,

readers are encouraged to refer to the specified literature [50]. The data utilized, with an hourly temporal resolution and up to recent date with a lag time of <2 months, is carefully selected to align with the temporal resolution and period of measured PAR data. Notably, this work employs Aerosol Optical Depth at 550 nm (AOD550) data, chosen for its measurement within the spectral range of PAR. Additionally, total ozone column measurements, referred to as Ozone, are analyzed to assess their impact on the Fraunhofer lines and early wavelengths that coincide with the PAR spectral range.

4. Machine Learning Models

The master dataset incorporates the identical quantity of inputs for the MR model to ensure an equitable comparison among various ML models. These models utilize readily accessible input parameters, free from constraints or measurement difficulties associated with external parameters. The process involves training the model with a designated dataset, followed by testing with validated data as outlined earlier. The selection of the model entails evaluating parameters, choosing an optimal hyperparameter tuning method, applying cross-validation, and reviewing validation outcomes. This research analyzes 4 ML algorithms, utilizing their default hyperparameters except as specified in Table 4, which presents the selected optimized hyperparameters.

Utilizing the k-Nearest Neighbors (kNN) algorithm, useful in varied data landscapes, XGB and LGBM, leveraging their strengths in optimization and efficiency were used. As seen in Table 4, with 257 trees for LGBM and 163 for XGB, the models are finely tuned to avoid overfitting while capturing essential data nuances. Deep learning via ANN, with a 0.0068 learning rate and two hidden layers containing 100 and 50 neurons respectively, tackles complex nonlinear relationships effectively. The model's design, including a minimal alpha value of 0.0002, ensures a strategic balance between model simplicity and the ability to capture detailed patterns in data, particularly in the context of parameter-PAR dynamics.

5. Model Evaluation Metrics

In reviewing the literature, it becomes evident that a wide array of evaluation techniques exists, underscoring the importance of choosing an apt performance indicator. This work, drawing from prior research, adopts the Mean Bias Deviation (MBD) as outlined in equation (7) as its primary metric for assessment [51,52]. The MBD serves as a critical tool for quantifying the variance between observed data and predictions made by PAR models, thereby providing an index of the model's accuracy in mirroring real-world observations. This metric is particularly valuable for identifying the degree to which the model's estimations diverge from actual measurements, either through underestimations, which impacts the direct and diffuse components of PAR, or overestimations, which affects these components inversely. By favoring metrics expressed in percentages over those in W/m², the work simplifies interpretation and enhances comparability, as these

Table 3

The summary of Variables with Sources and Methods to be used in PAR prediction.

Variable	Source	Method
Clear DNI (W/m^2)	McClear	–
AOD 550 nm	Collection "M21TNXAER" with Parameter Code "TOTEXTTAU"	Data for the half-hour was computed by averaging the lead and lag hour.
Total Ozone Amount (atm-cm)	Collection "M2T1NXSLV" with Parameter Code "TO3"	Data for the half-hour was computed by averaging the lead and lag hour. Raw data divided by 1000 to get in atm-cm

Table 4

Optimized hyperparameters for the 4 machine learning models underway, the rest of the hyperparameters are kept at default values unless stated.

Model	n_neighbors	weights		
kNN	11	distance		
Model	**learning_rate**	**max_depth**	**n_estimators**	**num_leaves**
LGBM	0.2040	10	257	25
Model	**learning_rate**	**max_depth**	**n_estimators**	**colsample_bytree**	**subsample**
XGB	0.0499	9	163	0.8926	0.7334
Model	**learning_rate_init**	**hidden_layer_sizes**	**alpha**	
ANN	0.0068	(100, 50)	0.0002	

percentage-based measures remain unaffected by changes in solar elevation that could skew the data.

$$MBD = \frac{1}{x} \sum_{i=1}^{n} \left(\frac{PAR_{Modelled} - PAR_{Measured}}{PAR_{Measured}} \right) \times 100 \qquad (7)$$

$PAR_{Modelled}$ represents the PAR value from the specific model for the i-th data point and $PAR_{Measured}$ conveys the measured PAR value from quantum sensor for the i-th data point. The number of data points is depicted as n. The $PAR'_{Measured}$ represents the mean of the measured PAR values.

$$R^2 = 1 - \frac{\sum (PAR_{Measured} - PAR_{Modelled})}{\sum (PAR_{Measured} - PAR'_{Measured})} \qquad (8)$$

Furthermore, this analysis incorporates the coefficient of determination, denoted as R^2 as per equation (8), to evaluate the predictive accuracy of various PAR models. This coefficient compares the variance in measured PAR values with the model predictions, with values nearing 1 indicating a high level of predictive accuracy.

In order to evaluate the performance of the different models, it is crucial to assess the average discrepancy between the model's predictions and the observed PAR value. This is achieved using the Root Mean Square Error (RMSE), outlined in equation (9), where the metric is presented in W/m2. The statistical analysis (as per Fig. 2), combined with MBD and R2, facilitates a comprehensive diagnosis of model accuracy as class A indicators of dispersion [51]. Specifically, MBD identifies bias in a particular direction, while RMSE quantifies the general
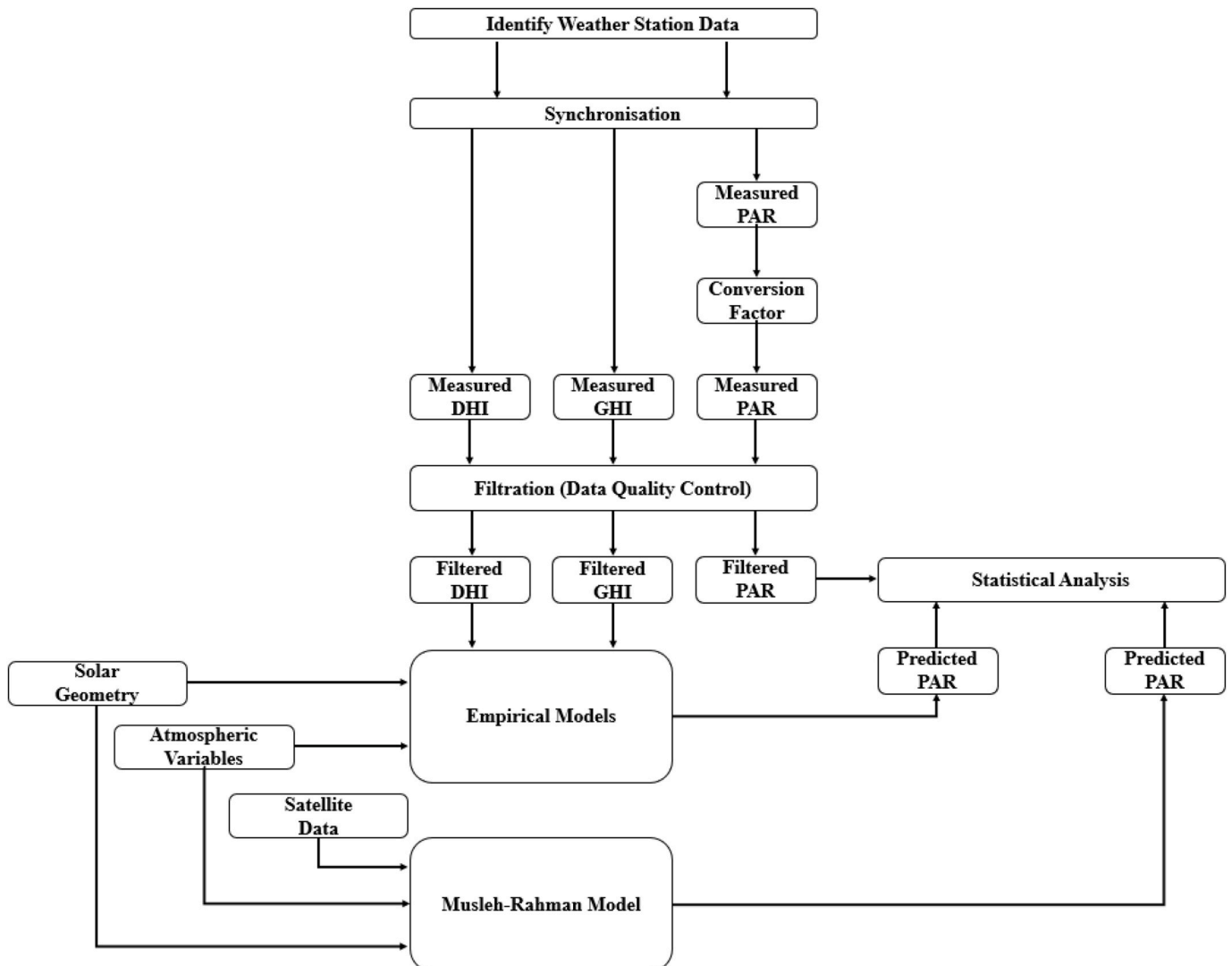

Fig. 2. The evaluation approach involves comparing empirical models with the newly developed Musleh-Rahman model, using open-source PAR measurement data.

magnitude of the errors.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(PAR_{Modelled} - PAR_{Measured})^2}{n}} \qquad (9)$$

6. Limitations of the Present Work

This research primarily assesses the empirical foundation of mathematical equations for estimating PAR using accessible data sources. However, focusing on machine learning approaches may yield richer insights. This sector is rapidly expanding, and not all model types, including stacked models, have been fully explored. While the LGBM model outperformed the Musleh-Rahman model, more nuanced hyperparameter tuning could enhance model performance further.

Additionally, the dataset used was limited to 30-min intervals. Given the volatility of sky conditions in temperate European climates, exploring higher temporal resolutions could be beneficial, particularly to understand the impact of cloud dynamics, such as accumulation and cloud enhancement ($K_t > 1$), on model performance. However, the primary constraint remains the limited dataset size and its 30-min temporal resolution.

3. Results and discussion

1. Interplay of Different Parameters and PAR

Previous studies have highlighted the multifaceted influences on PAR, suggesting no single factor can solely affect it. However, Figs. 3 and 4 demonstrate how individual parameter values can be linked to a range of PAR outcomes. Both figures present a detailed visualization of how various parameters interact with measured PAR, offering a nuanced understanding of their relationships and dependencies. Fig. 3a and b

shows a logistic growth concerning the cosine of SZA and K_t relative to measured PAR, illustrating an initial swift escalation that tapers off due to a limiting factor, embodying a sigmoidal curve that nears an asymptote. Conversely, Fig. 3c introduces a weighted function depicting dew point temperatures starting from 0 to 35 °C, where the data points exhibit a positive linear growth. Additionally, Fig. 3d suggests that lower Δ may inversely correlates with higher PAR values, evidenced by a denser concentration of points, indicating a potential inverse relationship between these variables.

Incorporating new variables into a new PAR model, alongside variables outlined in Fig. 3a-c, becomes essential due to the identified robust correlation among Clear DNI, AOD550, and Ozone levels as per Fig. 4. These parameters are crucial for PAR modeling because they influence the sunlight's path through the Earth's atmosphere, affecting its absorption and scattering potential. This, in turn, is influenced by molecular and aerosol interactions, as captured by AOD and ozone concentration metrics.

Analysis beginning with Fig. 4a reveals densification of data points at elevated DNI values, suggesting a potential exponential relationship where PAR escalates with rising DNI. This correlation is logical, given that increased solar irradiance directly translates to enhanced irradiance within PAR wavelengths. Furthermore, DNI is influenced by solar geometry, encompassing SZA and the Earth-sun distance. Higher solar positions correlate with increased DNI due to the reduced atmospheric path, minimizing scattering and absorption phenomena. Contrastingly, Fig. 4b illustrates a dense vertical aggregation, indicating a significant concentration of data points around a specific AOD 550 value. AOD 550, quantifying solar radiation extinction by atmospheric aerosols, is a unitless measure indicating the extent to which aerosols hinder sunlight from reaching the ground. Elevated AOD values suggest increased aerosol presence, which can scatter and absorb sunlight, diminishing the PAR reaching the Earth's surface. Since AOD measurements at 550 nm align with the PAR spectrum, it directly affects the light available for
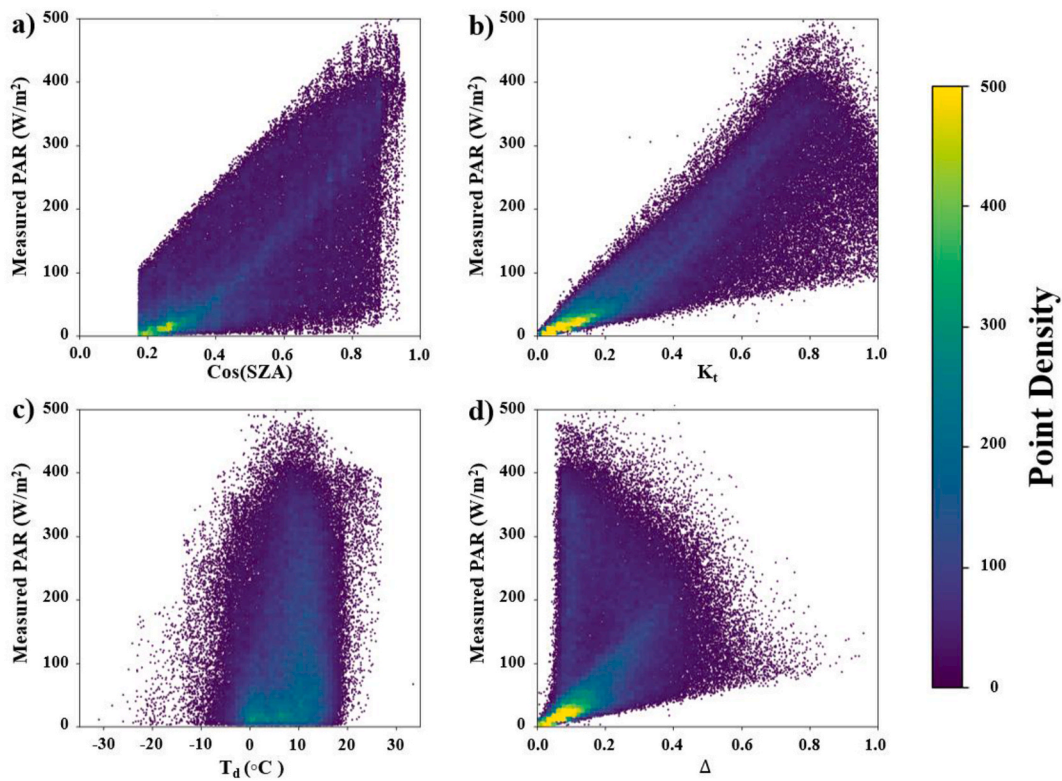

Fig. 3. The relationship of different parameters with regards to the measured PAR using data from both the training and testing segments across the 9 aforementioned sites. The concept of point density is employed to gauge the frequency of data point convergence within a scatter plot, specifically concerning the parameter and PAR, across a grid formation in increments of 0.01.
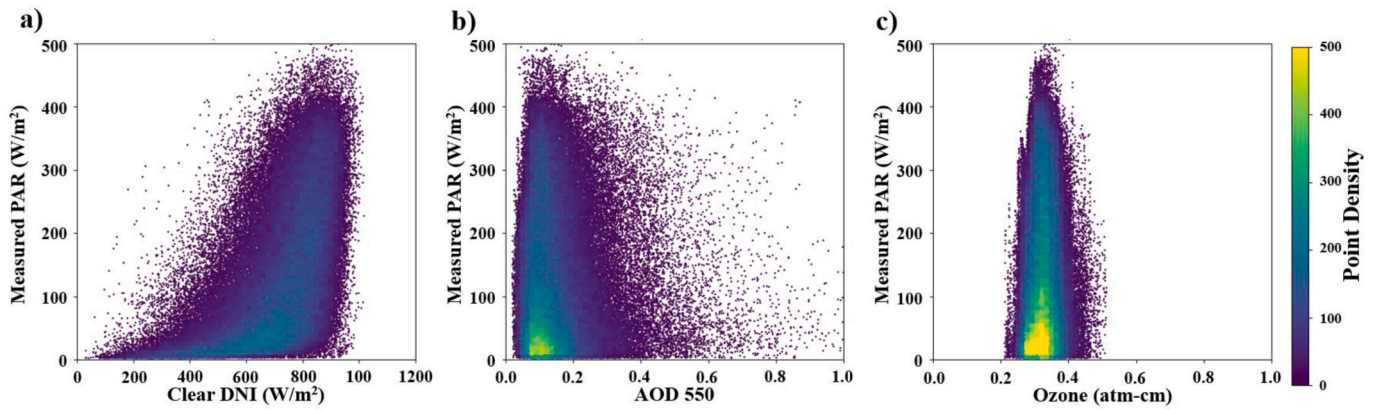
Fig. 4. Data from both the training and testing segments of the 9 aforementioned sites were used to investigate the relationship between the new parameters and the measured PAR. Data points in a scatter plot can be gauged by their density in increments of 0.01 using the point density concept, specifically concerning the parameter and PAR.

photosynthesis. The impact of AOD on PAR is further modulated by solar geometry; for instance, lower solar positions result in sunlight traversing more of the atmosphere, thereby encountering more aerosols and amplifying AOD's effect on PAR. Moreover, Fig. 4c displays a concentrated vertical distribution of PAR measurements around specific ozone concentrations, ranging between 0.2 and 0.3. Ozone predominantly absorbs sunlight in the UV spectrum, having a less immediate impact on PAR. Nonetheless, ozone can indirectly influence PAR through its effects on atmospheric thermal structure and composition, which can alter cloud formation and other atmospheric conditions affecting PAR levels at the surface. In this context, the relevance of Fraunhofer lines emerges, given the potential overlap of ozone absorption with these lines. However, the primary consideration for PAR remains the broader absorption characteristics of ozone, rather than the specific Fraunhofer lines.

3. Performance of recent Models and the Development of the new Regression Model

It can be highlighted that the performance of 10 models for estimating PAR, had MBD values in a wide exceeding range, ranging from 3.27 % to 33.51 % in Fig. 5. The Akitus1 and Tan-Ismail models, which utilize fewer parameters, demonstrate superior accuracy, indicating the significance of GHI and K_t in PAR estimation. This is further corroborated by Fig. 3b. However, the addition of atmospheric pressure in the Akitsu2 model may lead to overfitting, as suggested by its high MBD. The Escobedo models, which attempt to differentiate strategies based on K_t, do not align well nor conform strictly to the Cfb or Cfc Köppen-Geiger classifications. Models that incorporate DHI or Perez coefficients, such as those by Jacovides and Garcia-Rodriguez, show promise in more closely matching European temperate climates. Interestingly, the Alados model, validated in Spain (a location similar to Garcia-Rodriguez's), exhibits a significant increase in MBD (16.89 %) when substituting K_t with T_d, highlighting the critical role of K_t in PAR estimation and the unsuitability of the weighted T_d coefficient it employs in diverse European climates.

In response to these findings, this work proposes shifting away from DHI towards a new set of variables outlined in Fig. 4. These are integrated with K_t, GHI, SZA, and T_d based on their demonstrated correlation with PAR in literature and observations. The resulting Musleh-Rahman (MR) model incorporates the new variables in a weighted exponential formula. The model's coefficients, as per equation (10), are determined through nonlinear regression using least squares fit [53], a method chosen for its statistical consistency and relevance to reducing MBD. This iterative fitting process starts with initial values derived from the Alados and Garcia-Rodriguez models for K_t, GHI, and SZA while for T_d and the new variables are initially set to 0.1, ensuring a robust foundation for accurate PAR estimation.

$$PAR = GHI \cdot \left(\frac{1.386 + \ln K_t^{-0.059} + 1.06 \times 10^{-3} \cdot T_d + 0.185 \cdot \cos(SZA)}{e^{\left(6.60 \times 10^{-5} \cdot ClearDNI + 2.384 \cdot Ozone + 0.135 \cdot AOD550\right)}} \right) \quad (10)$$

The MR model demonstrates enhanced performance in predicting PAR values for European temperate climates, as evidenced in Fig. 6. It achieves a notable decrease in MBD by 3.08 % when compared to the best performing PAR estimations (i.e., the Garcia-Rodriguez model). This improvement is expected, given that the MR model is specifically calibrated using data from European climates, tailoring its coefficients to this particular environmental context. Fig. 6 reveals that, unlike its counterpart, the MR model exhibits a more diverse point density across the spectrum of irradiance levels, noted by its improved R^2 approaching one. It not only maintains accuracy at the initial irradiance levels but also extends this accuracy to higher levels of irradiance. This characteristic is crucial for enabling more precise estimations of PAR during critical periods, such as the harvesting season for certain crops, by providing insights into the required irradiance levels under different sky conditions. Consequently, the MR model's applicability extends beyond the limitations of data availability from stations measuring DHI, promoting its broader use in various applications. Additionally, Fig. 6 highlights a denser overlap between modelled and measured PAR across the entire range of PAR values, indicating a more robust and versatile
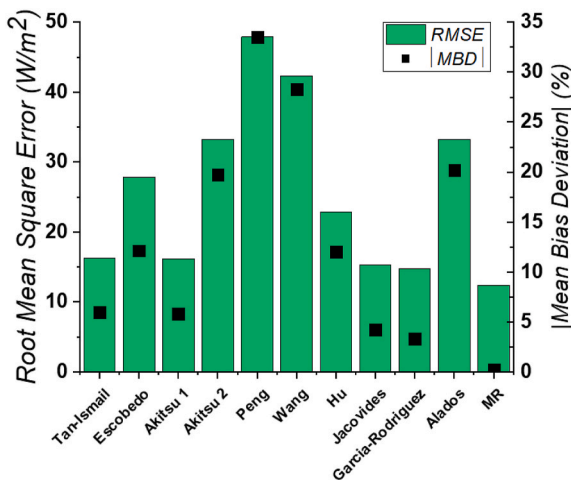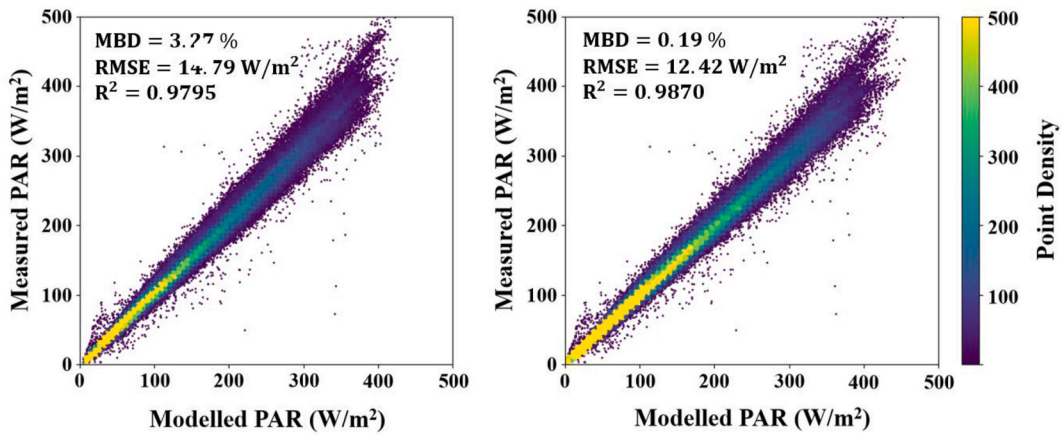

Fig. 5. In evaluating the efficacy of the 10 models and the newly introduced MR model across 9 distinct locations, the analysis specifically focuses on the RMSE and absolute MBD values. Among these, only the Escobedo model exhibits a negative MBD.

Fig. 6. A side-by-side comparison of measured versus modelled PAR using two distinct models: the Garcia-Rodriguez model is depicted on the left, while the MR model is featured on the right. The comparison aims to illustrate the accuracy of each model in replicating observed PAR data across 9 validation sites.

performance compared to the DHI-dependent Garcia-Rodriguez model.

4. Machine Learning Models

Fig. 7 demonstrates the superiority of the 4 examined ML algorithms over the 10 PAR empirical models mentioned in the earlier section in Fig. 5. Among these, ANN shows the most significant MBD at −2.32 % and RMSE of 22.91 W/m^2, with the kNN following at MBD of −0.25 %. The XGB and LGBM models exhibit near identical performance levels, each achieving MBD of −0.16 % and −0.15 %, respectively. Moreover, when considering the RMSE, slight differences emerge. The XGB model records a lower RMSE at 20.10 W/m^2, compared to the LGBM model at 20.55 W/m^2. This similarity in performance can be attributed to their gradient boosting methodology and the incorporation of regularization techniques, which mitigate overfitting by imposing penalties on coefficient magnitude. Such techniques are especially advantageous when dealing with training data that is susceptible to noise, a common occurrence in the variable weather conditions of European temperate climates. LGBM, kNN, and XGB exhibit MBD values below 1 % and RMSE values differing by a maximum of 3.50 W/m^2, indicating their high accuracy in predicting PAR. Both LGBM and XGB ensure effective handling of complex, nonlinear relationships, a feature bolstered by their sophisticated implementations and optimizations for large-scale, high-dimensional data analysis.

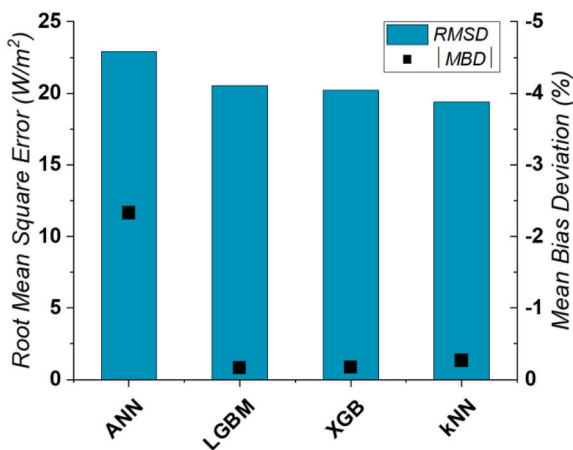Although the ANN surpasses the conventional PAR models in

performance, its interpretability remains a challenge. In contrast, XGB and LGBM not only deliver powerful predictive capabilities but also enhance interpretability through features like importance scores. Notably, the MR model introduced in this work closely approaches the performance of the XGB and LGBM models, with only a slight MBD difference of 0.03 % and 0.04 %, respectively. It is important to highlight that both the MR and ML models were trained and validated on data from 9 European climate sites, emphasizing the need to assess them in unseen dataset to determine their applicability, scalability and robustness in these settings.

5. Evaluation under Unseen Sites

Table 5 reveals that the performance of the 10 empirical models under has a significant MBD range from −8.41 % to 39.47 % and Table 6 shows RMSE ranging between 7.66 W/m^2 and 97.24 W/m^2. Among these, the Peng and Wang models remain notably inferior outcomes due to their empirical development within China, failing to account for the significant SZA encountered at higher latitudes; explaining their similar performance in MBD and RMSE values. Furthermore, employing a simple ratio conversion, exemplified by the Tan-Ismail method, results in an RMSE spread from 20.31W/m^2 to 51.96 W/m^2, with its MBD values being positive indicating overestimations throughout. The higher numerical value occurring under a Csa climate close to the sea, which implies elevated humidity, thus, varying precipitation and temperature conditions. This suggests that a straightforward arithmetic transformation may not be sufficiently accurate.

Segmenting the analysis based on sky clarity yielded a modest improvement of 1.94 % in absolute MBD terms, yet the overall performance remains unsatisfactory when using the Escobedo model. However, incorporating DHI through Perez coefficients offers some enhancement in model accuracy. Among the 10 models, Garcia-Rodriguez still remaining the better model, standing out with a comparatively better MBD of 7.71 % and the lowest RMSE of 24.11 W/m^2, with the lowest drop to −1.99 % observed in the Hohes-Holz scenario, likely due to the alignment of GHI and PAR within the validated climate conditions it has.

Fig. 8 highlights for Hohes Holz, the Garcia-Rodriguez's model fluctuating monthly, influenced by seasonal PAR and irradiance changes. Its efficacy is high during June to September, with MBD values under 1 %, but drops markedly in winter, notably in January and December, due to unsuitable coefficients for high-latitude climates and inability to handle winter's steep solar zenith angles. This points to the need for models like MR, which sidesteps DHI derivatives, showing improved accuracy across six locations, with reasonable achievements in Aurade and Lochristi. The MR model benefits from incorporating new

Fig. 7. Based on the same data input as the MR model, the RMSE and MBD values are specifically examined in evaluating the performance of the 4 ML models using the validation dataset.

Table 5

An analysis of the MBD percentage was conducted to assess the estimation accuracy of PAR at new testing sites. This comparison involved both empirical and machine learning models. The model achieving the lowest (thus, most accurate) MBD value is distinguished with a bold highlight.

Model	Aurade	Fontainebleau-Barbeau	Hohes Holz	Lochristi	Maasmechelen	Castelporziano	Overall
Tan-Ismail	12.99	18.60	8.00	9.82	15.85	24.17	10.35
Escobedo	−6.07	10.05	−16.42	−8.88	10.45	15.88	−8.41
Akitsu 1	12.84	18.44	−4.49	9.68	15.71	24.00	10.20
Akitsu 2	28.39	33.95	13.68	24.01	30.78	40.57	24.69
Peng	44.61	49.39	26.74	38.97	46.08	57.21	39.47
Wang	38.72	43.28	21.87	33.43	40.17	50.31	33.84
Hu	14.41	20.42	7.56	15.93	21.53	21.76	14.42
Jacovides	11.34	16.28	−0.89	7.99	13.89	22.53	8.60
Garcia-Rodriguez	10.74	15.92	−1.99	7.21	13.10	21.71	7.71
Alados	20.01	25.73	16.08	24.21	29.92	25.51	21.44
MR	6.04	**−1.54**	**0.67**	4.08	**−3.88**	**3.22**	4.03
ANN	−5.82	−9.66	4.41	−4.30	−9.76	−14.11	−6.00
LGBM	**−5.35**	−9.24	4.79	**−3.90**	−9.34	−13.87	**−3.85**
XGB	−5.52	−9.34	4.79	−3.94	−9.40	−14.07	−3.93
kNN	−6.07	−9.49	4.56	−4.42	−9.75	−13.32	−4.02

Table 6

An evaluation of the RMSE (in W/m^2) to assess the PAR of the MBD percentage was conducted to assess the accuracy of PAR at the unseen test sites. This comparison involved both empirical and machine learning models. Models that recorded the lowest RMSE values in the given location are marked in bold.

Model	Aurade	Fontainebleau-Barbeau	Hohes Holz	Lochristi	Maasmechelen	Castelporziano	Overall
Tan-Ismail	29.24	31.61	23.09	20.31	27.08	51.96	26.21
Escobedo	17.67	18.94	33.32	22.19	18.78	36.70	25.92
Akitsu 1	28.98	31.39	11.37	20.14	26.87	51.67	26.05
Akitsu 2	54.46	52.42	20.65	37.22	47.65	77.59	43.98
Peng	73.62	68.60	34.74	53.79	63.39	97.24	59.17
Wang	66.72	62.36	29.39	47.99	57.35	89.27	53.22
Hu	29.72	34.32	12.59	26.53	34.74	47.22	27.42
Jacovides	26.91	28.37	8.88	18.41	24.64	49.28	24.76
Garcia-Rodriguez	26.42	28.56	7.79	17.93	24.15	48.20	24.11
Alados	39.33	41.88	23.09	38.73	46.11	54.68	36.27
MR	18.81	**8.25**	**7.66**	13.75	**12.29**	**12.72**	19.83
ANN	19.79	19.52	12.00	13.67	18.93	40.58	22.01
LGBM	**18.47**	19.74	11.48	**13.12**	19.46	38.33	**18.67**
XGB	18.54	19.69	11.29	13.61	19.47	38.69	18.69
kNN	21.02	21.85	11.81	15.36	21.43	38.05	20.27

variables, enhancing precision and suggesting effectiveness in ML contexts. In comparison, ML models trained in temperate climates surpass all models, with LGBM leading and MR closely following in performance.

While ML models, as depicted in Fig. 8, exhibit better stability than Garcia-Rodriguez, they nonetheless show less variability than the MR model. This observation suggests that future research could explore the classification of ML models by climate or seasonal attributes to further refine their performance. The location of Castelporziano was identified as the least accurately modelled by ML methods, likely due to the predominance of training and validation data from higher latitudes within Cfb climate classifications, indicating a potential area for adjustment. Nonetheless, ML models still outshine their counterparts across the board.

Exploring the performance nuances of the Garcia-Rodriguez, MR, and LGBM models against different PAR irradiance levels, Fig. 9 shows kernel density estimations shed light on their efficacy. These estimations reveal a bell-shaped distribution for PAR values, highlighting a peak around 50 W/m^2. The Garcia-Rodriguez model aligns well with observed PAR levels but tends to slightly overestimate at the peak as per Fig. 9a, predominantly covering lower irradiance values below 200 W/m^2, reflective of most dataset observations. From Fig. 9e, the MR model, through its integration of McClear DNI, avoids the peak overestimation seen in Garcia-Rodriguez, providing a more consistent density across the 0–100 W/m^2 range. It exhibits a broader and more even distribution of predicted PAR values, particularly reducing bias towards higher or lower extremes as seen in Fig. 9b. This balanced approach aligns the MR model more closely with the actual distribution of PAR values within the

dataset, improving its accuracy for a wider range of conditions. Conversely, the LGBM, while similar to MR in avoiding peak biases, demonstrates a slight increase in density within the moderate PAR range (201–300 W/m^2), suggesting it is more attuned to capturing these mid-range values noted in Fig. 9f. This slight shift in prediction density may give the LGBM a marginal advantage in accuracy for datasets with a moderate PAR value prevalence. Moreover, its performance across the entire PAR spectrum indicates a versatile capability to accurately forecast across diverse irradiance levels. The LGBM model's nuanced distribution suggests it might offer the most comprehensive and balanced performance among the three, especially in contexts requiring accurate predictions across a broad spectrum of PAR irradiance.

Through examining Fig. 10, it becomes evident that the Garcia-Rodriguez model's point density is predominantly concentrated within the initial 100 PAR irradiance values. This observation aligns with previous discussions, highlighting a decrease in modeling performance, in terms of point density, as PAR values escalate. Specifically, the density of grid interactions diminishes, elucidated by an MBD of 7.71 % demonstrates the model's inclination to overestimate actual measured PAR values by an RMSE of 24.11 W/m^2. Despite this tendency for overestimation, the model significantly elucidates the variance observed in measured PAR data, as reflected by an R^2 value of 0.9463. This indicates that although the model exhibits a consistent bias towards higher estimates, its predictions align closely with the trends observed in actual measurements, reinforcing insights presented in Fig. 10. In contrast, the MR model exhibits a reduction in performance at unseen sites by 4.28 W/m^2 in terms of RMSE and MBD of 3.68 % relative to the Garcia-Rodriguez model, yielding an MBD of 4.03 % and an RMSE of 19.83
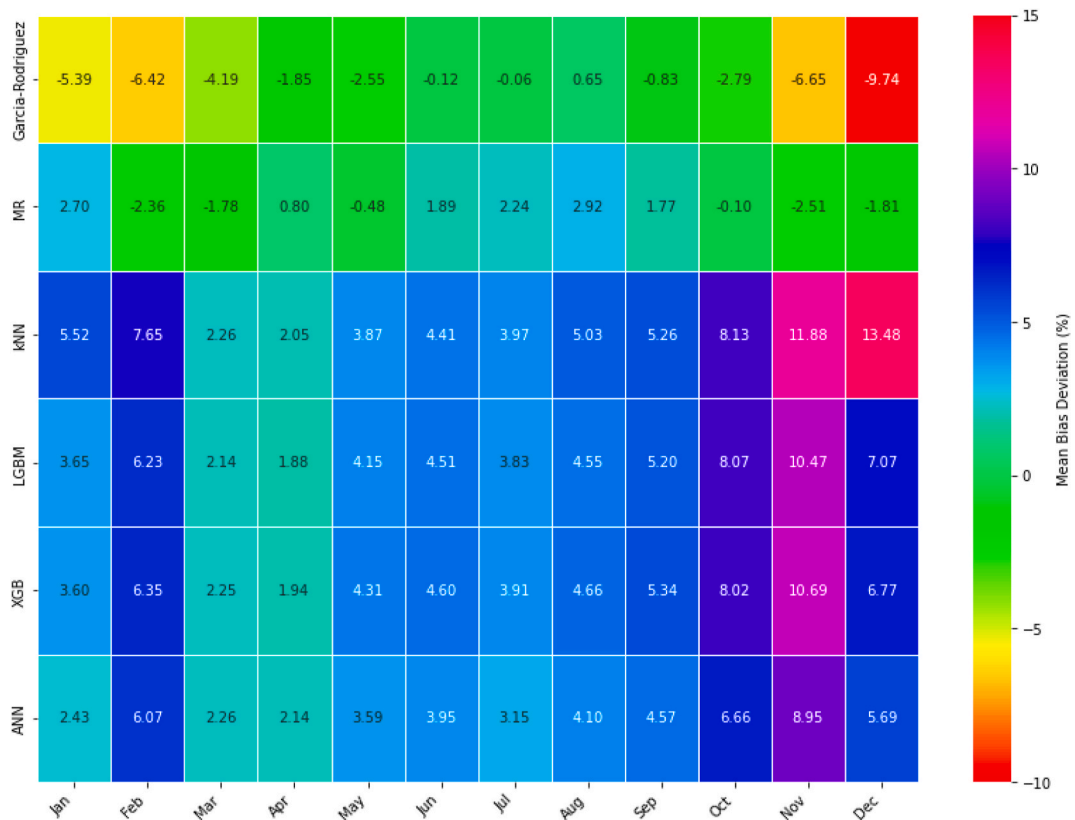
Fig. 8. The heatmap illustrates the MBD expressed as a percentage, comparing the optimal empirical model, the MR model, and 4 machine learning models, utilizing the hsv colour scheme, for the Hohez Holz dataset in the year 2022. A value closer to zero denotes superior accuracy.

W/m^2. This reduced deviation suggests a narrower average error margin compared to the findings in Fig. 9a. Furthermore, an enhanced R^2 value signifies a superior model fit to the measured data, implying not only an accurate trend prediction but also greater precision on average. The density of points near the 1:1 line is notably high for lower PAR values, with no distinct peak value as indicated in Fig. 9b, pointing towards numerous accurate predictions. The LGBM boasts the highest R^2 value among the models, signifying the most accurate overall predictive capability. The negative MBD highlights the model's reliability, albeit with a tendency towards conservative estimates.

It becomes apparent that eliminating DHI and incorporating McClear DNI, Ozone, and AOD550 not only reduces MBD and improves the R^2 value but also enhances the model's robustness, making it applicable beyond sites equipped to measure DHI. Specifically, the inclusion of McClear DNI addresses the issue of early PAR values, while Ozone and AOD550 adjustments rectify overestimations beyond 200 W/m^2, facilitating more accurate representation of fluctuations. However, further exploration is necessary to refine estimates of clear irradiance and explore the potential for sky condition classification through parameters or monthly data analysis. This approach could mitigate seasonal variances and, by integrating ML with the MR model, tailor coefficients to suit various Köppen-Geiger climate classifications, thereby broadening the model's applicability.

4. Conclusion

In conclusion, this work embarked on an evaluation of 10 empirical models designed for PAR prediction, with a subset of 4 models incorporating DHI across 15 temperate European environments, highlighting the need for diverse solar, atmospheric, and cloud cover variations at higher latitudes. This process culminated in the development of the Musleh-Rahman (MR) model. This approach eliminates the need for

DHI, instead clear-sky DNI, as well as Ozone and AOD550nm data sourced from satellite observations. The MR model's performance was then tested and confirmed across 9 European climates in conjunction with 4 ML models: kNN, ANN, LGBM, and XGB. The MR model demonstrated superior performance, particularly in a validation study across nine European climates where it reduced the Mean Bias Deviation (MBD) to 0.19 % from the 3.08 % observed in the Garcia-Rodriguez model and the Root Mean Square Error (RMSE) reduced by 4.28 W/m^2. After tuning the hyperparameters, 3 out of the 4 machine learning models demonstrated superior performance compared to the MR model. In further testing across 6 unseen European locations, the MR model showed MBD values ranging from −3.88 % to 6.04 %. The MR model and the LGBM were particularly noted for their precision in lower PAR values, as evidenced by kernel density and scatter plot analyses, achieving R2 values of 0.9709 and 0.9712, respectively, with having RMSE values of The RMSE values of 19.83 W/m2 and 18.67 W/m2, though the LGBM had an absolute MBD improvement of 0.18 %. Future research should focus on evaluating additional weather variables within optimally hypertuned machine learning models, including the use of more diverse and stacked models. There is also interest in refining clear irradiance estimations for temperate climates, particularly at higher temporal resolutions such as minute–minute.

Such advancements could mitigate seasonal variability and facilitate a deeper understanding of the interplay of PAR and cloud effects for various Köppen-Geiger climate classifications, further refining PAR prediction accuracy.

CRediT authorship contribution statement

Yazan J.K. Musleh: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Tasmiat Rahman:** Methodology, Validation, Resources,
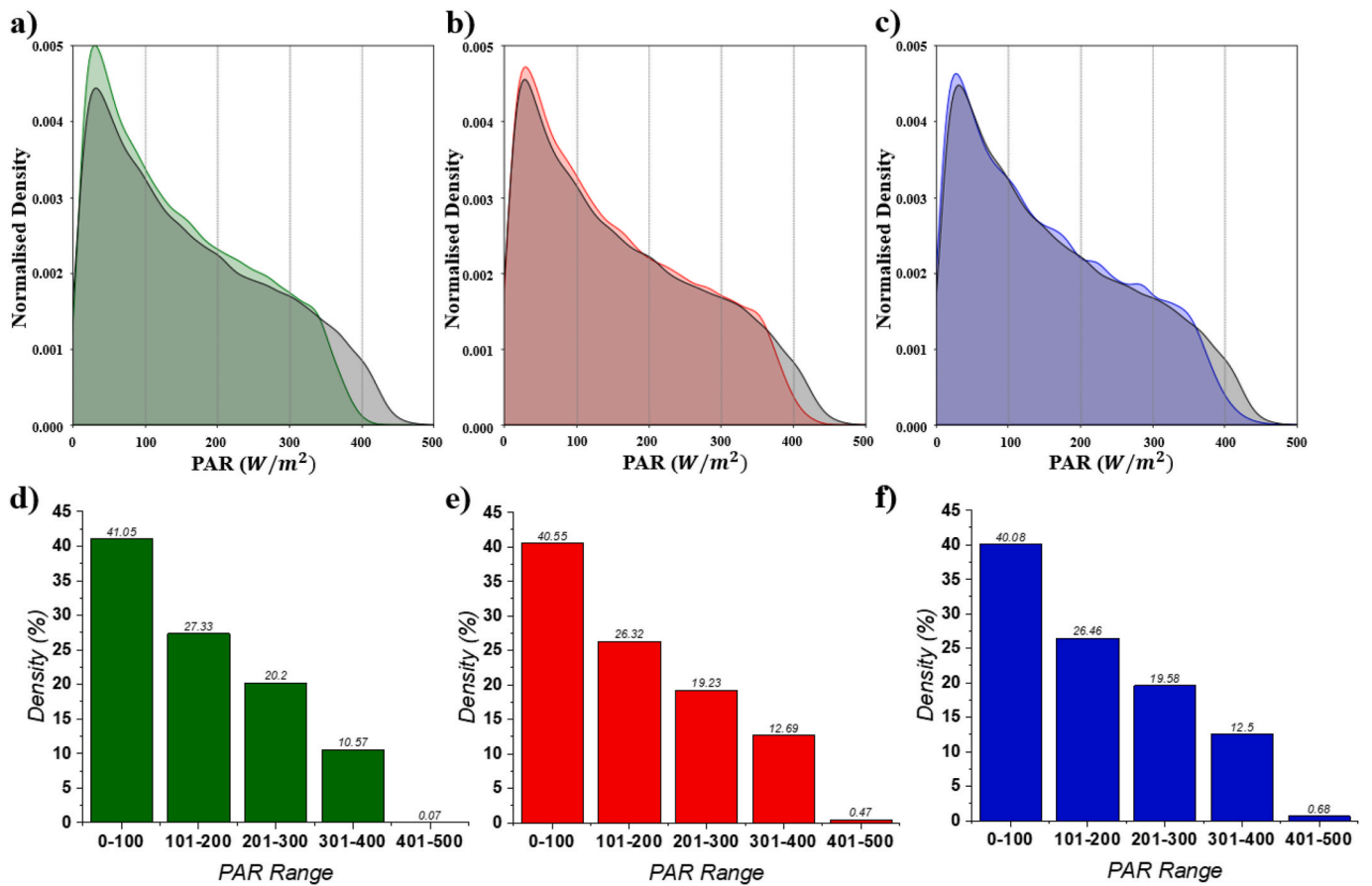
Fig. 9. Comparative performance of the best-performing models with the normalised density distributions (**a-c**) and the associated data density percentages within specific PAR intervals (**d-f**) for three key models: the Garcia-Rodriguez model (illustrated in green), the MR model (red), and the LGBM (blue).
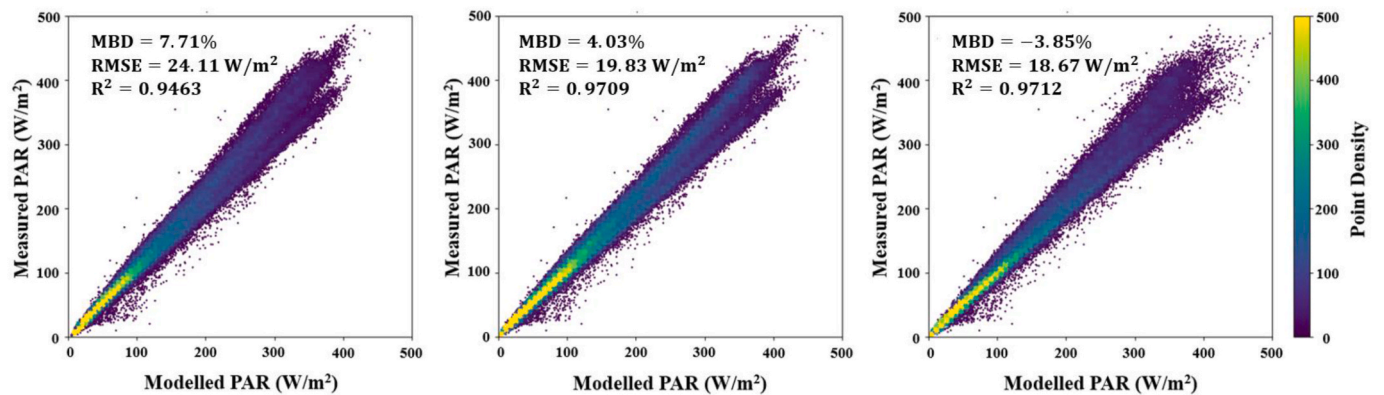

Fig. 10. Comparative scatter plots of measured PAR against the best-performing models with Garcia-Rodriguez (Left), MR (Middle), and LGBM (Right) at 6 unseen European Temperate Locations.

Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

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.

Data availability

Data will be made available on request.

Acknowledgment

Additionally, they express appreciation to ICOS for supplying the photosynthetic photon flux density alongside the total and diffuse shortwave radiation dataset utilized across the remaining stations discussed in this article.

References

[1] Akbari Variani H, Afshar A, Vahabzadeh M, Molajou A, Akbari Varyani MM. Development of a novel framework for agriculture simulation model for food-energy-water nexus analysis in watershed-scale. J Clean Prod Dec. 2023;429: 139492. https://doi.org/10.1016/j.jclepro.2023.139492. Elsevier BV.

[2] Peng Y, et al. Energy performance assessment of photovoltaic greenhouses in summer based on coupled optical-electrical-thermal models and plant growth requirements. Energy Convers Manag Jul. 2023;287:117086. https://doi.org/10.1016/j.enconman.2023.117086. Elsevier BV.

[3] Ferreira RF, Marques Lameirinhas RA, Bernardo CP Correia V, Torres JPN, Santos M. Agri-PV in Portugal: how to combine agriculture and photovoltaic production. Energy for Sustainable Development Apr. 2024;79:101408. https://doi.org/10.1016/j.esd.2024.101408. Elsevier BV.

[4] Willockx B, Lavaert C, Cappelle J. Geospatial assessment of elevated agrivoltaics on arable land in Europe to highlight the implications on design, land use and economic level. Energy Rep Nov. 2022;8:8736–51. https://doi.org/10.1016/j.egyr.2022.06.076. Elsevier BV.

[5] Willockx B, Lavaert C, Cappelle J. Performance evaluation of vertical bifacial and single-axis tracked agrivoltaic systems on arable land. Renew Energy Nov. 2023; 217:119181. https://doi.org/10.1016/j.renene.2023.119181. Elsevier BV.

[6] Campana PE, Stridh B, Amaducci S, Colauzzi M. Optimisation of vertically mounted agrivoltaic systems. J Clean Prod Nov. 2021;325:129091. https://doi.org/10.1016/j.jclepro.2021.129091. Elsevier BV.

[7] PVPS Task 1. TRENDS IN PHOTOVOLTAIC applications 2022 task 1 strategic PV analysis and Outreach PVPS. 2022. www.iea-pvps.org.

[8] International Roadmap for Photovoltaic (ITRPV). Thirteenth Edition 2022. p. 2022.

[9] Mouhib E, et al. Enhancing land use: integrating bifacial PV and olive trees in agrivoltaic systems. Appl Energy Apr. 2024;359:122660. https://doi.org/10.1016/j.apenergy.2024.122660. Elsevier BV.

[10] Ghosh A. Nexus between agriculture and photovoltaics (agrivoltaics, agriphotovoltaics) for sustainable development goal: a review. Sol Energy Dec. 2023;266:112146. https://doi.org/10.1016/j.solener.2023.112146. Elsevier BV.

[11] Ma Lu S, et al. Photosynthetically active radiation separation model for high-latitude regions in agrivoltaic systems modeling. J Renew Sustain Energy Jan. 01, 2024;16(1). https://doi.org/10.1063/5.0181311. AIP Publishing.

[12] Ma Lu S, et al. Photosynthetically active radiation decomposition models for agrivoltaic systems applications. Sol Energy Sep. 2022;244:536–49. https://doi.org/10.1016/j.solener.2022.05.046. Elsevier BV.

[13] Lozano IL, Sánchez-Hernández G, Guerrero-Rascado JL, Alados I, Foyo-Moreno I. Analysis of cloud effects on long-term global and diffuse photosynthetically active radiation at a Mediterranean site. Atmos Res Apr. 2022;268:106010. https://doi.org/10.1016/j.atmosres.2021.106010. Elsevier BV.

[14] Nwokolo S. A global review of empirical models for estimating photosynthetically active radiation. Trends in Renewable Energy Oct. 2018;4(2). https://doi.org/10.17737/tre.2018.4.2.0079. Future Energy Service and Publishing.

[15] Wang L, Gong W, Li C, Lin A, Hu B, Ma Y. Measurement and estimation of photosynthetically active radiation from 1961 to 2011 in Central China. Appl Energy Nov. 2013;111:1010–7. https://doi.org/10.1016/j.apenergy.2013.07.001. Elsevier BV.

[16] Peng S, Du Q, Lin A, Hu B, Xiao K, Xi Y. Observation and estimation of photosynthetically active radiation in Lhasa (Tibetan Plateau). Adv Space Res Mar. 2015;55(6):1604–12. https://doi.org/10.1016/j.asr.2015.01.002. Elsevier BV.

[17] Wang L, Gong W, Hu B, Lin A, Li H, Zou L. Modeling and analysis of the spatiotemporal variations of photosynthetically active radiation in China during 1961–2012. Renew Sustain Energy Rev Sep. 2015;49:1019–32. https://doi.org/10.1016/j.rser.2015.04.174. Elsevier BV.

[18] Aguiar LJG, et al. Modeling the photosynthetically active radiation in South West Amazonia under all sky conditions. Theor Appl Climatol Dec. 02, 2011;108(3–4): 631–40. https://doi.org/10.1007/s00704-011-0556-z. Springer Science and Business Media LLC.

[19] Ferrera-Cobos F, Vindel JM, Valenzuela RX, González JA. Models for estimating daily photosynthetically active radiation in oceanic and mediterranean climates and their improvement by site adaptation techniques. Adv Space Res Apr. 2020;65 (8):1894–909. https://doi.org/10.1016/j.asr.2020.01.018. Elsevier BV.

[20] Proutsos N, Alexandris S, Liakatas A, Nastos P, Tsiros IX. PAR and UVA composition of global solar radiation at a high altitude Mediterranean forest site. Atmos Res May 2022;269:106039. https://doi.org/10.1016/j.atmosres.2022.106039. Elsevier BV.

[21] Escobedo JF, Gomes EN, Oliveira AP, Soares J. Ratios of UV, PAR and NIR components to global solar radiation measured at Botucatu site in Brazil. Renew Energy Jan. 2011;36(1):169–78. https://doi.org/10.1016/j.renene.2010.06.018. Elsevier BV.

[22] Akitsu TK, et al. The variability and seasonality in the ratio of photosynthetically active radiation to solar radiation: a simple empirical model of the ratio. Int J Appl Earth Obs Geoinf Apr. 2022;108:102724. https://doi.org/10.1016/j.jag.2022.102724. Elsevier BV.

[23] Perez R, Ineichen P, Seals R, Michalsky J, Stewart R. Modeling daylight availability and irradiance components from direct and global irradiance. Sol Energy 1990;44 (5):271–89. https://doi.org/10.1016/0038-092x(90)90055-h. Elsevier BV.

[24] Alados I, Foyo-Moreno I, Alados-Arboledas L. Photosynthetically active radiation: measurements and modelling. Agric For Meteorol Jan. 1996;78(1–2):121–31. https://doi.org/10.1016/0168-1923(95)02245-7. Elsevier BV.

[25] Jacovides CP, Timvios FS, Papaioannou G, Asimakopoulos DN, Theofilou CM. Ratio of PAR to broadband solar radiation measured in Cyprus. Agric For Meteorol Feb. 2004;121(3–4):135–40. https://doi.org/10.1016/j.agrformet.2003.10.001. Elsevier BV.

[26] Hu B, Wang Y, Liu G. Measurements and estimations of photosynthetically active radiation in Beijing. Atmos Res Sep. 2007;85(3–4):361–71. https://doi.org/10.1016/j.atmosres.2007.02.005. Elsevier BV.

[27] de Blas M, García-Rodríguez A, García I, Torres JL. Validation and calibration of models to estimate photosynthetically active radiation considering different time scales and sky conditions. Adv Space Res Oct. 2022;70(7):1737–60. https://doi.org/10.1016/j.asr.2022.07.005. Elsevier BV.

[28] CIE, Standard. Spitial distribution of Daylight-CIE Stndard general sky. CIE S 011.1/E; 2003.

[29] García-Rodríguez A, Granados-López D, García-Rodríguez S, Díez-Mediavilla M, Alonso-Tristán C. Modelling Photosynthetic Active Radiation (PAR) through meteorological indices under all sky conditions. Agric For Meteorol Nov. 2021;310: 108627. https://doi.org/10.1016/j.agrformet.2021.108627. Elsevier BV.

[30] Zainali S, et al. Direct and diffuse shading factors modelling for the most representative agrivoltaic system layouts. Appl Energy Jun. 2023;339:120981. https://doi.org/10.1016/j.apenergy.2023.120981. Elsevier BV.

[31] Yang D. Estimating 1-min beam and diffuse irradiance from the global irradiance: a review and an extensive worldwide comparison of latest separation models at 126 stations. Renew Sustain Energy Rev May 2022;159:112195. https://doi.org/10.1016/j.rser.2022.112195. Elsevier BV.

[32] Yang D, et al. Regime-dependent 1-min irradiance separation model with climatology clustering. Renew Sustain Energy Rev Jan. 2024;189:113992. https://doi.org/10.1016/j.rser.2023.113992. Elsevier BV.

[33] Widmer J, Christ B, Grenz J, Norgrove L. Agrivoltaics, a promising new tool for electricity and food production: a systematic review. Renew Sustain Energy Rev Mar. 2024;192:114277. https://doi.org/10.1016/j.rser.2023.114277. Elsevier BV.

[34] ICOS Carbon Portal." https://www.icos-cp.eu/.

[35] ISO, "Solar energy – Specification and classification of instruments for measuring hemispherical solar and direct solar radiation," ISO 9060:1990..

[36] Carrara A, et al. Radiation measurements at ICOS ecosystem stations. Int Agrophys Dec. 01, 2018;32(4):589–605. https://doi.org/10.1515/intag-2017-0049. Walter de Gruyter GmbH.

[37] Ohmura A, et al. Baseline Surface Radiation Network (BSRN/WCRP): new precision radiometry for climate research. Bull Am Meteorol Soc 1998;79:2115–36.

[38] Oh M, et al. Analysis of minute-scale variability for enhanced separation of direct and diffuse solar irradiance components using machine learning algorithms. Energy Feb. 2022;241:122921. https://doi.org/10.1016/j.energy.2021.122921. Elsevier BV.

[39] Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci Data Oct. 30, 2018;5(1). https://doi.org/10.1038/sdata.2018.214. Springer Science and Business Media LLC.

[40] Gueymard CA. Revised composite extraterrestrial spectrum based on recent solar irradiance observations. Sol Energy Jul. 2018;169:434–40. https://doi.org/10.1016/j.solener.2018.04.067. Elsevier BV.

[41] Reda I, Andreas A. Solar position algorithm for solar radiation applications. Sol Energy 2004;76(5):577–89. https://doi.org/10.1016/j.solener.2003.12.003. Elsevier BV.

[42] Forstinger A, et al. Expert quality control of solar radiation ground data sets. In: Proceedings of the ISES solar world Congress 2021. International Solar Energy Society; 2021. https://doi.org/10.18086/swc.2021.38.02.

[43] Tan PY, Ismail MRB. Photosynthetically active radiation and comparison of methods for its estimation in equatorial Singapore. Theor Appl Climatol Feb. 13, 2015;123(3–4):873–83. https://doi.org/10.1007/s00704-015-1399-9. Springer Science and Business Media LLC.

[44] Gueymard CA, Ruiz-Arias JA. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. Sol Energy Apr. 2016;128:1–30. https://doi.org/10.1016/j.solener.2015.10.010. Elsevier BV.

[45] de Blas M, García-Rodríguez A, García I, Torres JL. Validation and calibration of models to estimate photosynthetically active radiation considering different time scales and sky conditions. Adv Space Res Oct. 2022;70(7):1737–60. https://doi.org/10.1016/j.asr.2022.07.005. Elsevier BV.

[46] Lefèvre M, et al. McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmos Meas Tech Sep. 17, 2013;6(9): 2403–18. https://doi.org/10.5194/amt-6-2403-2013. Copernicus GmbH.

[47] CAMS McClear. CAMS McClear radiation service. SODA-Pro Web Services, Copernicus Atmosphere Monitoring Service; 2023. https://www.soda-pro.com/web-services/radiation/cams-mcclear.

[48] Gelaro R, McCarty W, Suárez MJ, Todling R, Molod A, Takacs L, Randles CA, Darmenov A, Bosilovich MG, Reichle R, et al. Modern-Era retrospective analysis for research and applications, Version2 (MERRA-2). J Clim.

[49] NASA Goddard Earth Sciences Data and Information Services Center, "NASA Goddard Earth Sciences Data and Information Services Center Datasets." Available: https://disc.gsfc.nasa.gov/datasets/..

[50] Sun X, Bright JM, Gueymard CA, Acord B, Wang P, Engerer NA. Worldwide performance assessment of 75 global clear-sky irradiance models using Principal Component Analysis. Renew Sustain Energy Rev Sep. 2019;111:550–70. https://doi.org/10.1016/j.rser.2019.04.006. Elsevier BV.

[51] Gueymard CA. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: towards a better bankability of solar projects. Renew Sustain Energy Rev Nov. 2014;39:1024–34. https://doi.org/10.1016/j.rser.2014.07.117. Elsevier BV.

[52] Liu Y, Liu Y, Cai H, Zhang J. An innovative short-term multihorizon photovoltaic power output forecasting method based on variational mode decomposition and a capsule convolutional neural network. Appl Energy Aug. 2023;343:121139. https://doi.org/10.1016/j.apenergy.2023.121139. Elsevier BV.

[53] Rogers C, et al. Daily leaf area index from photosynthetically active radiation for long term records of canopy structure and leaf phenology. Agric For Meteorol Jul. 2021;304–305:108407. https://doi.org/10.1016/j.agrformet.2021.108407. Elsevier BV.

[54] Yildiz B, Bilbao JI, Sproul AB. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renew Sustain Energy Rev Jun. 2017;73:1104–22. https://doi.org/10.1016/j.rser.2017.02.023. Elsevier BV.

[55] Yamasaki Junior M, Freire RZ, Seman LO, Stefenon SF, Mariani VC, dos Santos Coelho L. Optimized hybrid ensemble learning approaches applied to very short-term load forecasting. Int J Electr Power Energy Syst Jan. 2024;155:109579. https://doi.org/10.1016/j.ijepes.2023.109579. Elsevier BV.

[56] Murphy J, Chanat J. Leveraging machine learning to automate regression model evaluations for large multi-site water-quality trend studies. Environ Model Software Dec. 2023;170:105864. https://doi.org/10.1016/j.envsoft.2023.105864. Elsevier BV.

[57] Elsheikh AH, Sharshir SW, Abd Elaziz M, Kabeel AE, Guilan W, Haiou Z. Modeling of solar energy systems using artificial neural network: a comprehensive review. Sol Energy Mar. 2019;180:622–39. https://doi.org/10.1016/j.solener.2019.01.037. Elsevier BV.

[58] Elsheikh AH, Katekar VP, Muskens OL, Deshmukh SS, Elaziz MA, Dabour SM. Utilization of LSTM neural network for water production forecasting of a stepped solar still with a corrugated absorber plate. Process Saf Environ Protect Apr. 2021;148:273–82. https://doi.org/10.1016/j.psep.2020.09.068. Elsevier BV.

[59] Ghandourah E, Prasanna YS, Elsheikh AH, Moustafa EB, Fujii M, Deshmukh SS. Performance prediction of aluminum and polycarbonate solar stills with air cavity using an optimized neural network model by golden jackal optimizer. Case Stud Therm Eng Jul. 2023;47:103055. https://doi.org/10.1016/j.csite.2023.103055. Elsevier BV.

[60] Essa FA, Elaziz MA, Al-Betar MA, Elsheikh AH. Performance prediction of a reverse osmosis unit using an optimized Long Short-term Memory model by hummingbird optimizer. Process Saf Environ Protect Jan. 2023;169:93–106. https://doi.org/10.1016/j.psep.2022.10.071. Elsevier BV.

[61] Mercier TM, Sabet A, Rahman T. Vision transformer models to measure solar irradiance using sky images in temperate climates. Applied Energy 2024;362:122967. https://doi.org/10.1016/j.apenergy.2024.122967. Elsevier BV.

[62] T.M. Mercier, A. Sabet, T. Rahman, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops (2023).