

Predictive Models for Photosynthetic Active Radiation Irradiance in Temperate Climates

Yazan J.K. Musleh^{a,*} and Tasmia Rahman^a

^aUniversity 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 550nm (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/m². 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/m², 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 R² 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 R² 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 prediction models.

Keywords — Solar irradiance, Temperate Climates, Photosynthetic Active Radiation, Empirical Modelling, Regression Modelling, Machine Learning

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 - 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., $\mu\text{mol photons/m}^2/\text{s}$) or in terms of flux density (i.e., irradiance W/m²) [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][25][26][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 $4.57 \mu\text{mol}/\text{m}^2/\text{s}$ 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 Töravere is included [37], resulting in 15 locations altogether as per Figure 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-minute 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, 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].

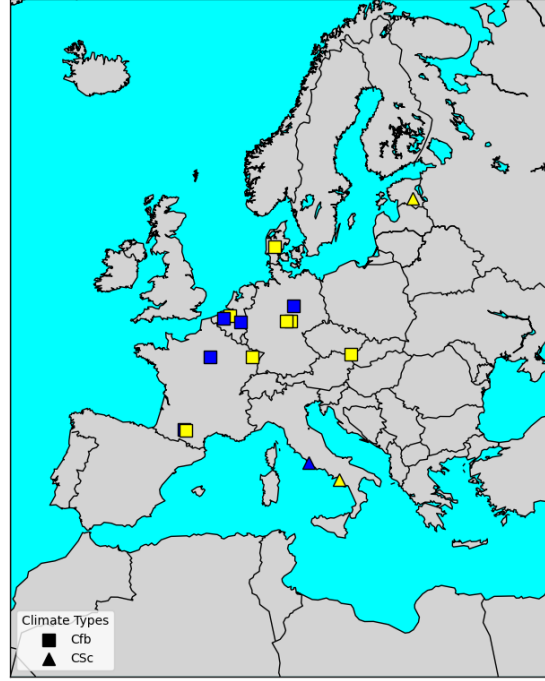


Figure 1: The spatial layout of the 15 chosen locations includes 9 sites marked in yellow for the purpose of model training and validation, and 6 sites indicated in blue dedicated to evaluating the models against new data. Among these, 12 sites are classified under Temperate Oceanic Climates (Cfb), while 3 sites fall within the Cool Summer Mediterranean Climate category (Csc) as per the Köppen–Geiger climate classification [39].

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 [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 to 6000. Consequently, this allows for the quantification of sky clarity through the sky clearness index, K_t , as outlined in equation 1.

$$K_t = \frac{GHI}{E_a} \quad (1)$$

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. $SZA < 85^\circ$

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 (m)	Period	Climate Classification	Mean PAR (W/m ²)	Mean GHI (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
Töravere	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

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 approach to PAR estimation based on varying K_t values.

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

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 .

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

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

$$GHI = DHI + DNI \cos(SZA) \quad (5)$$

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].

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

Table 2: Details on the parameters used across the 10 PAR models under evaluation. Alados refers to Model 1 in [24].

Model	GHI	K _t	Pressure	SZA	Δ	ε	T _d
Tan-Ismael	✓						
Escobedo	✓	✓					
Akitsu 1	✓	✓					
Akitsu 2	✓	✓	✓				
Peng	✓	✓		✓			
Wang	✓	✓		✓			
Hu	✓			✓	✓		
Jacovides		✓			✓	✓	
Garcia-Rodriguez	✓	✓		✓	✓	✓	
Alados	✓			✓	✓	✓	✓

3. New parameters

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 550nm	Collection "M21TNXAER" with Parameter Code "TOTEXTAU"	Data for the half-hour was computed by averaging the lead and lag hour.
Total Ozone Amount (atm-cm)	Collection "M21NXSLV" 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

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 one minute 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 550nm (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

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		

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/m^2 , the work simplifies interpretation and enhances comparability, as these percentage-based measures remain unaffected by changes in solar elevation that could skew the data.

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

$PAR_{Modelled}$ 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.

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

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/m^2 . The statistical analysis (as per Figure 2), combined with MBD and R^2 , 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 magnitude of the errors.

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

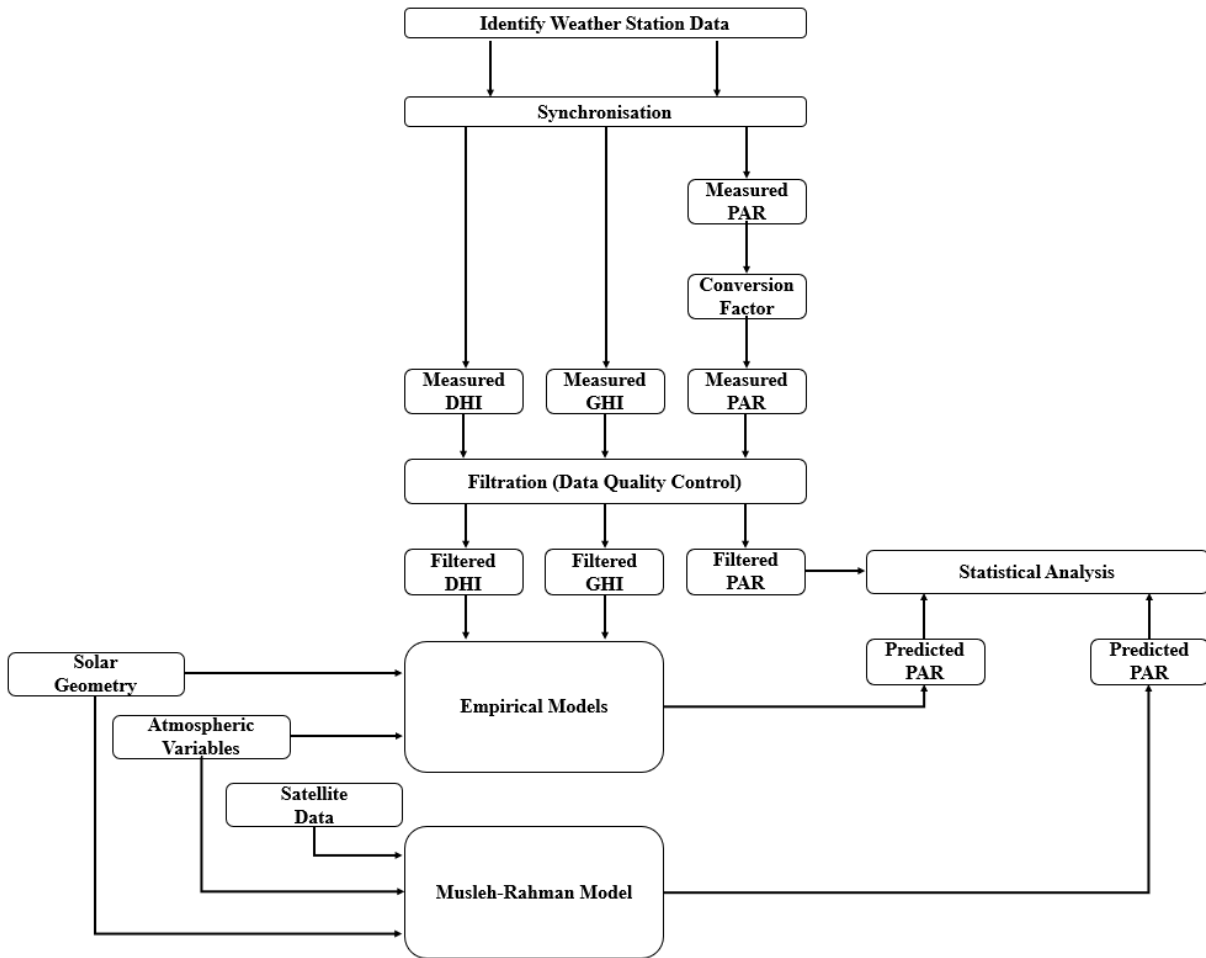


Figure 2: The evaluation approach involves comparing empirical models with the newly developed Musleh-Rahman model, using open-source PAR measurement data.

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-minute 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-minute temporal resolution.

3. RESULTS AND DISCUSSION

1. Interplay of Different Parameters and PAR

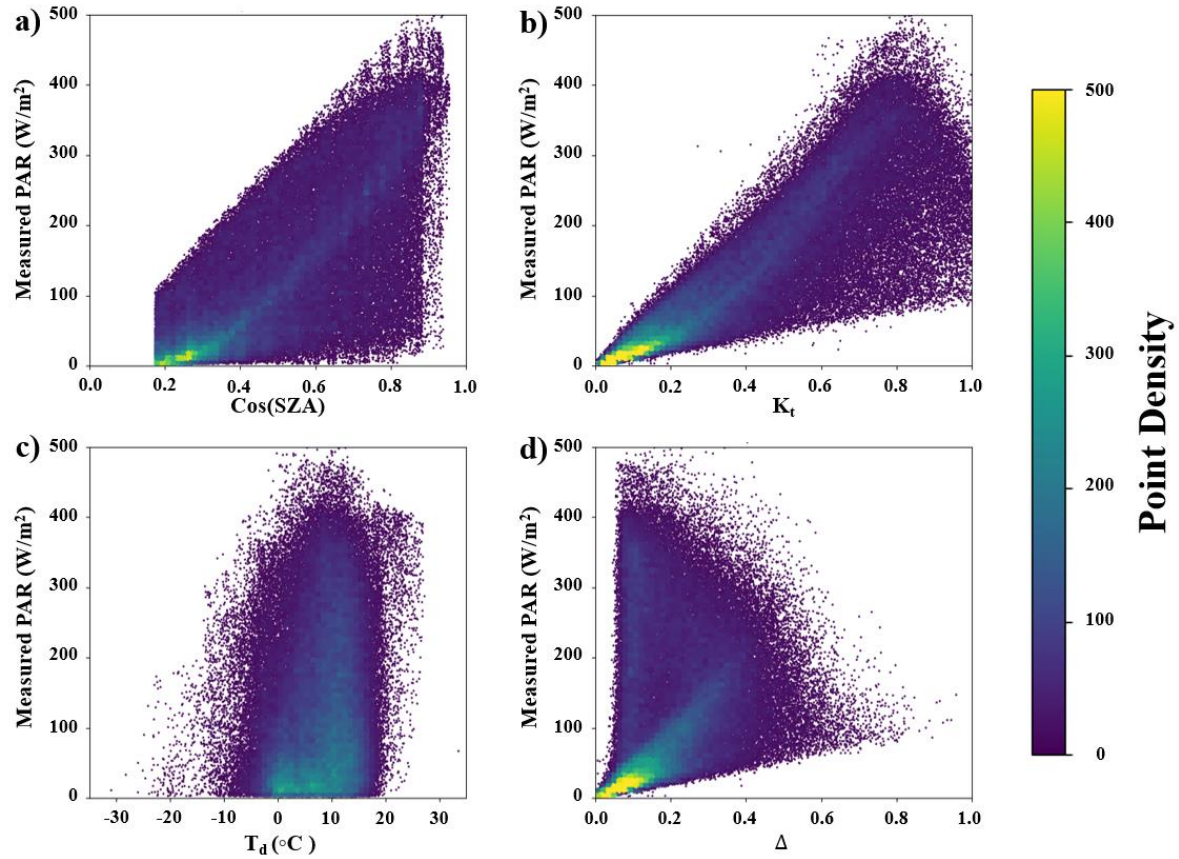


Figure 3: The relationship of different parameters with regards to the measured PAR using data from both the training and testing segments across the 9 aforementioned sites. The concept of point density is employed to gauge the frequency of data point convergence within a scatter plot, specifically concerning the parameter and PAR, across a grid formation in increments of 0.01.

Previous studies have highlighted the multifaceted influences on PAR, suggesting no single factor can solely affect it. However, Figures 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. Figures 3a and 3b show 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, Figure 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, Figure 3d suggests that lower Δ may inversely correlate with higher PAR values, evidenced by a denser concentration of points, indicating a potential inverse relationship between these variables.

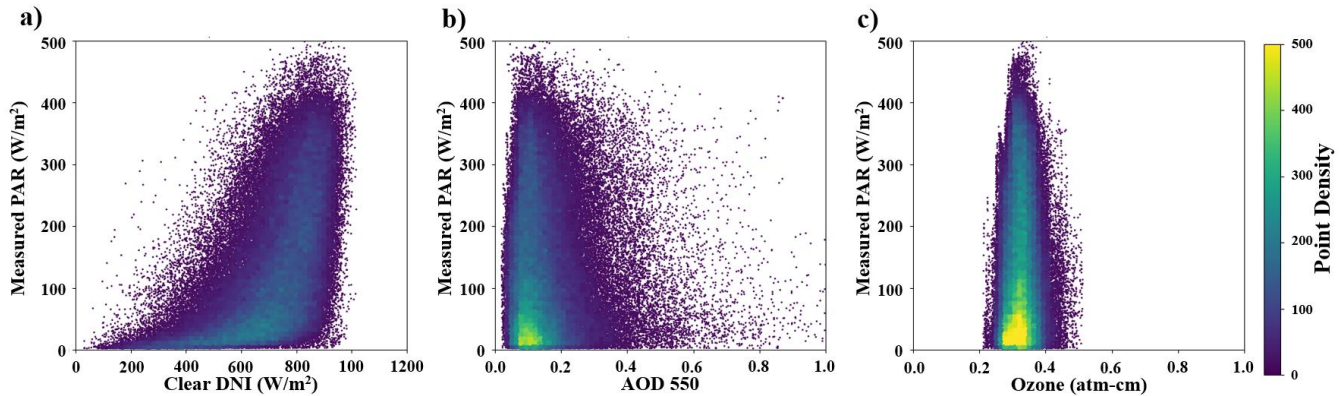


Figure 4: Data from both the training and testing segments of the 9 aforementioned sites were used to investigate the relationship between the new parameters and the measured PAR. Data points in a scatter plot can be gauged by their density in increments of 0.01 using the point density concept, specifically concerning the parameter and PAR.

Incorporating new variables into a new PAR model, alongside variables outlined in Figures 3a-3c, becomes essential due to the identified robust correlation among Clear DNI, AOD550, and Ozone levels as per Figure 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 Figure 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, Figure 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 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, Figure 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 Figure 5. The Akitsu1 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 Figure 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.

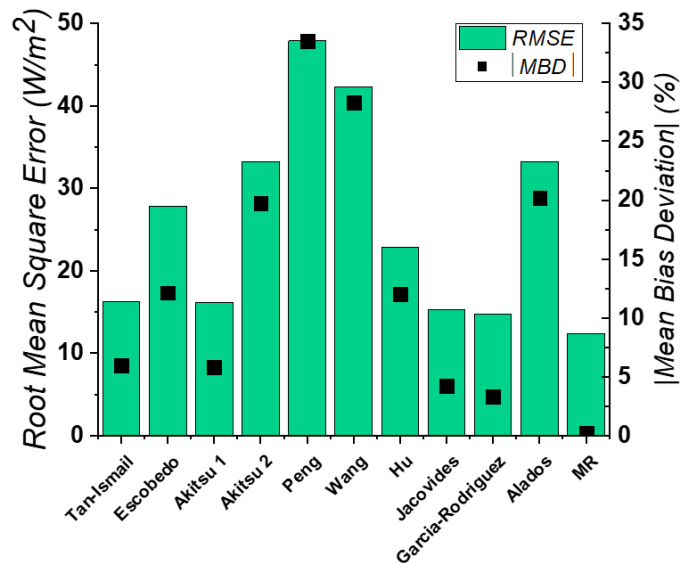


Figure 5: In evaluating the efficacy of the 10 models and the newly introduced MR model across 9 distinct locations, the analysis specifically focuses on the RMSE and absolute MBD values. Among these, only the Escobedo model exhibits a negative MBD.

In response to these findings, this work proposes shifting away from DHI towards a new set of variables outlined in Figure 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.

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

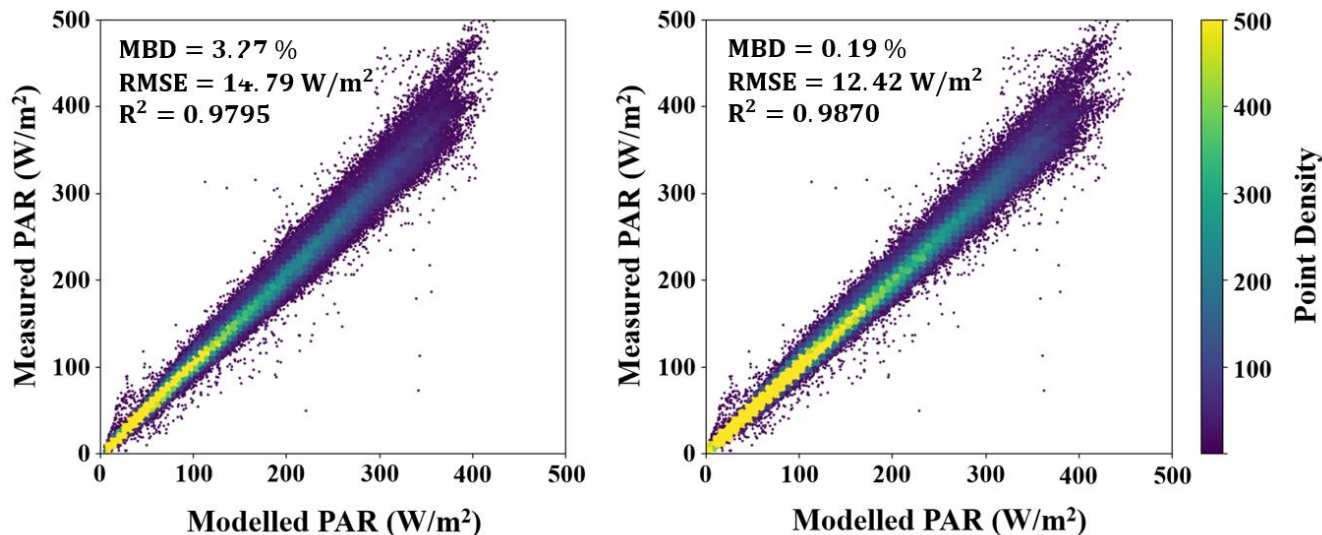


Figure 6: A side-by-side comparison of measured versus modelled PAR using two distinct models: the Garcia-Rodriguez model is depicted on the left, while the MR model is featured on the right. The comparison aims to illustrate the accuracy of each model in replicating observed PAR data across 9 validation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

The MR model demonstrates enhanced performance in predicting PAR values for European temperate climates, as evidenced in Figure 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. Figure 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, Figure 6 highlights a denser overlap between modelled and measured PAR across the entire range of PAR values, indicating a more robust and versatile performance compared to the DHI-dependent Garcia-Rodriguez model.

4. Machine Learning Models

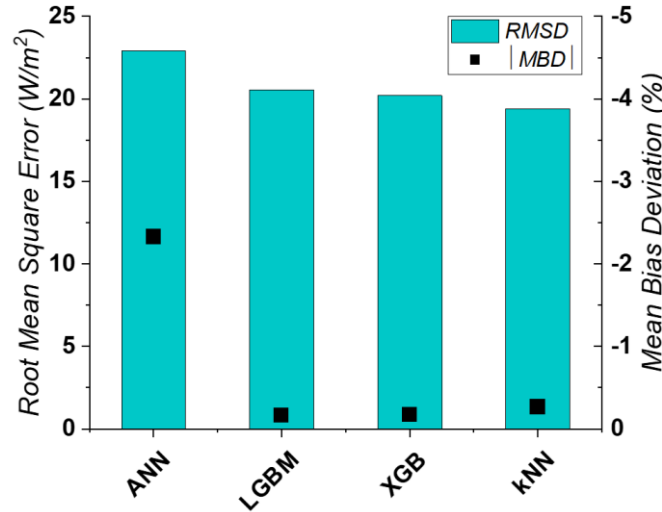


Figure 7: Based on the same data input as the MR model, the RMSE and MBD values are specifically examined in evaluating the performance of the 4 ML models using the validation dataset.

Figure 7 demonstrates the superiority of the 4 examined ML algorithms over the 10 PAR empirical models mentioned in the earlier section in Figure 5. Among these, ANN shows the most significant MBD at -2.32% and RMSE of 22.91 W/m², 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/m², compared to the LGBM model at 20.55 W/m². 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/m², 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/m² and 97.24 W/m². 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-Ismael method, results in an RMSE spread from 20.31 W/m² to 51.96 W/m², 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/m², 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.

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

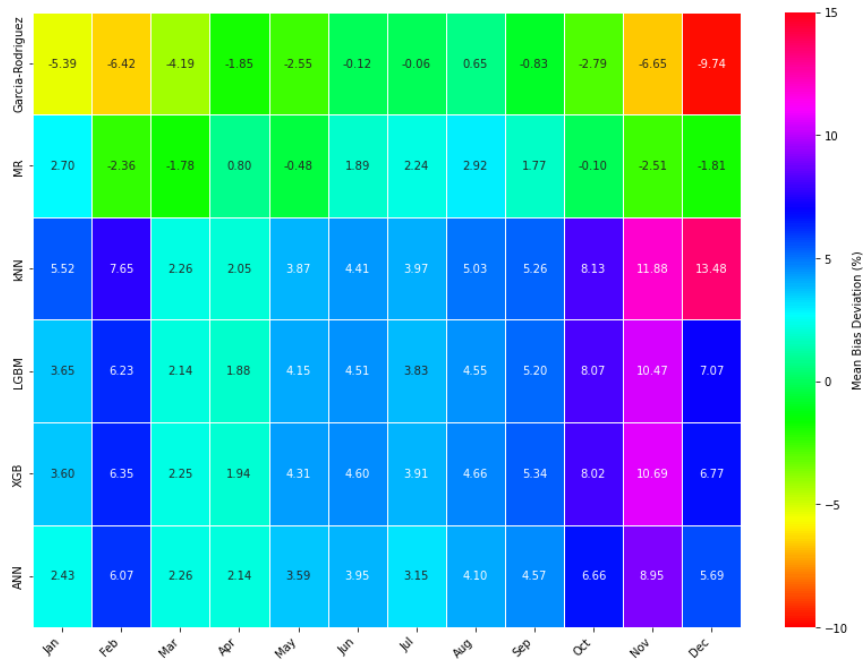


Figure 8: The heatmap illustrates the MBD expressed as a percentage, comparing the optimal empirical model, the MR model, and 4 machine learning models, utilizing the hsv colour scheme, for the Hohez Holz dataset in the year 2022. A value closer to zero denotes superior accuracy.

Figure 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 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 Figure 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 modeled 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.

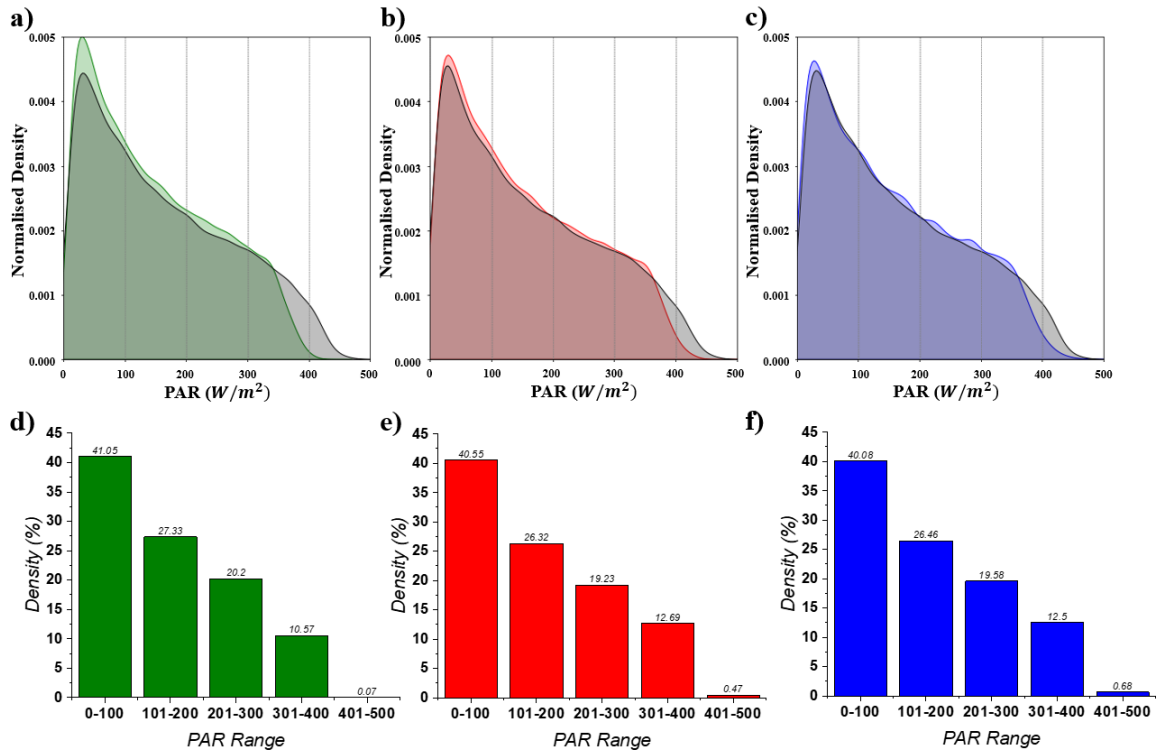


Figure 9: Comparative performance of the best-performing models with the normalised density distributions (a-c) and the associated data density percentages within specific PAR intervals (d-f) for three key models: the Garcia-Rodriguez model (illustrated in green), the MR model (red), and the LGBM (blue).

Exploring the performance nuances of the Garcia-Rodriguez, MR, and LGBM models against different PAR irradiance levels, Figure 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 Figure 9a, predominantly covering lower irradiance values below $200 W/m^2$, reflective of most dataset observations. From Figure 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 to $100 W/m^2$ range. It exhibits a broader and more even distribution of predicted PAR values, particularly reducing bias towards higher or lower extremes as seen in Figure 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 Figure 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 Figure 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 Figure 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 W/m^2 . This reduced deviation suggests a narrower average error margin compared to the findings in Figure 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 Figure 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.

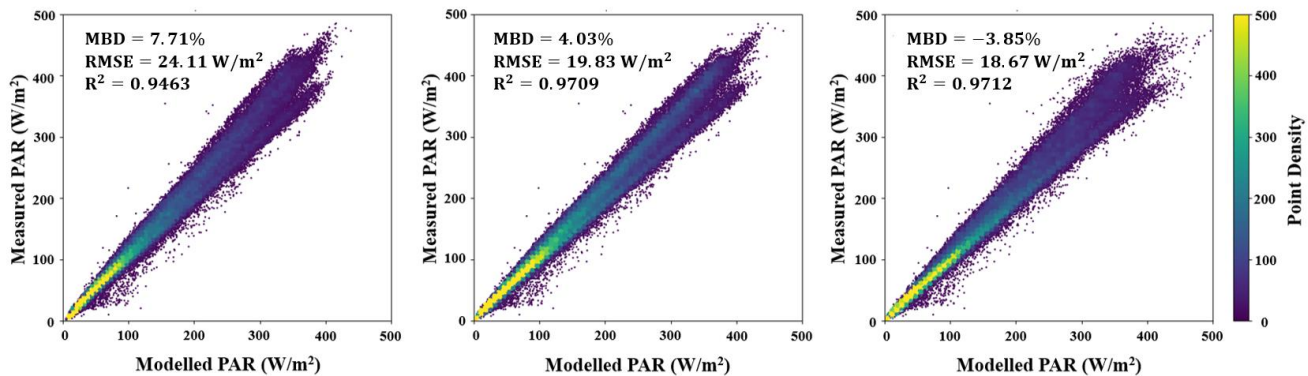


Figure 10: Comparative scatter plots of measured PAR against the best-performing models with Garcia-Rodriguez (Left), MR (Middle), and LGBM (Right) at 6 unseen European Temperate Locations.

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 R^2 values of 0.9709 and 0.9712, respectively, with having RMSE values of 19.83 W/m^2 and 18.67 W/m^2 , 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.

ACKNOWLEDGMENT

Y. J. K. Musleh acknowledges the support of the Engineering and Physical Sciences Research Council (EPSRC) Doctoral Training Partnership (DTP) Funding. T. Rahman acknowledges support from EPSRC grant EP/X033333/1. The authors extend their gratitude to Epp Anderson for generously providing the radiation data for Tõravere. 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] H. Akbari Varyani, A. Afshar, M. Vahabzadeh, A. Molajou, and M. M. Akbari Varyani, “Development of a novel framework for agriculture simulation model for food-energy-water nexus analysis in watershed-scale,” *Journal of Cleaner Production*, vol. 429. Elsevier BV, p. 139492, Dec. 2023. doi: 10.1016/j.jclepro.2023.139492.
- [2] Y. Peng et al., “Energy performance assessment of photovoltaic greenhouses in summer based on coupled optical-electrical-thermal models and plant growth requirements,” *Energy Conversion and Management*, vol. 287. Elsevier BV, p. 117086, Jul. 2023. doi: 10.1016/j.enconman.2023.117086.
- [3] R. F. Ferreira, R. A. Marques Lameirinhas, C. P. Correia V. Bernardo, J. P. N. Torres, and M. Santos, “Agri-PV in Portugal: How to combine agriculture and photovoltaic production,” *Energy for Sustainable Development*, vol. 79. Elsevier BV, p. 101408, Apr. 2024. doi: 10.1016/j.esd.2024.101408.
- [4] B. Willockx, C. Lavaert, and J. Cappelle, “Geospatial assessment of elevated agrivoltaics on arable land in Europe to highlight the implications on design, land use and economic level,” *Energy Reports*, vol. 8. Elsevier BV, pp. 8736–8751, Nov. 2022. doi: 10.1016/j.egyr.2022.06.076.
- [5] B. Willockx, C. Lavaert, and J. Cappelle, “Performance evaluation of vertical bifacial and single-axis tracked agrivoltaic systems on arable land,” *Renewable Energy*, vol. 217. Elsevier BV, p. 119181, Nov. 2023. doi: 10.1016/j.renene.2023.119181.
- [6] P. E. Campana, B. Stridh, S. Amaducci, and M. Colauzzi, “Optimisation of vertically mounted agrivoltaic systems,” *Journal of Cleaner Production*, vol. 325. Elsevier BV, p. 129091, Nov. 2021. doi: 10.1016/j.jclepro.2021.129091.
- [7] PVPS Task 1. (2022). TRENDS IN PHOTOVOLTAIC APPLICATIONS 2022 Task 1 Strategic PV Analysis and Outreach PVPS. Retrieved from www.ica-pvps.org
- [8] International Roadmap for Photovoltaic (ITRPV). (2022). 13th edition
- [9] E. Mouhib et al., “Enhancing land use: Integrating bifacial PV and olive trees in agrivoltaic systems,” *Applied Energy*, vol. 359. Elsevier BV, p. 122660, Apr. 2024. doi: 10.1016/j.apenergy.2024.122660.
- [10] A. Ghosh, “Nexus between agriculture and photovoltaics (agrivoltaics, agriphotovoltaics) for sustainable development goal: A review,” *Solar Energy*, vol. 266. Elsevier BV, p. 112146, Dec. 2023. doi: 10.1016/j.solener.2023.112146.
- [11] S. Ma Lu et al., “Photosynthetically active radiation separation model for high-latitude regions in agrivoltaic systems modeling,” *Journal of Renewable and Sustainable Energy*, vol. 16, no. 1. AIP Publishing, Jan. 01, 2024. doi: 10.1063/5.0181311.
- [12] S. Ma Lu et al., “Photosynthetically active radiation decomposition models for agrivoltaic systems applications,” *Solar Energy*, vol. 244. Elsevier BV, pp. 536–549, Sep. 2022. doi: 10.1016/j.solener.2022.05.046.
- [13] I. L. Lozano, G. Sánchez-Hernández, J. L. Guerrero-Rascado, I. Alados, and I. Foyo-Moreno, “Analysis of cloud effects on long-term global and diffuse photosynthetically active radiation at a Mediterranean site,” *Atmospheric Research*, vol. 268. Elsevier BV, p. 106010, Apr. 2022. doi: 10.1016/j.atmosres.2021.106010.
- [14] S. Nwokolo, “A Global Review of Empirical Models for Estimating Photosynthetically Active Radiation,” *Trends in Renewable Energy*, vol. 4, no. 2. Future Energy Service and Publishing, Oct. 2018. doi: 10.17737/tre.2018.4.2.0079.
- [15] L. Wang, W. Gong, C. Li, A. Lin, B. Hu, and Y. Ma, “Measurement and estimation of photosynthetically active radiation from 1961 to 2011 in Central China,” *Applied Energy*, vol. 111. Elsevier BV, pp. 1010–1017, Nov. 2013. doi: 10.1016/j.apenergy.2013.07.001.

- [16] S. Peng, Q. Du, A. Lin, B. Hu, K. Xiao, and Y. Xi, "Observation and estimation of photosynthetically active radiation in Lhasa (Tibetan Plateau)," *Advances in Space Research*, vol. 55, no. 6. Elsevier BV, pp. 1604–1612, Mar. 2015. doi: 10.1016/j.asr.2015.01.002.
- [17] L. Wang, W. Gong, B. Hu, A. Lin, H. Li, and L. Zou, "Modeling and analysis of the spatiotemporal variations of photosynthetically active radiation in China during 1961–2012," *Renewable and Sustainable Energy Reviews*, vol. 49. Elsevier BV, pp. 1019–1032, Sep. 2015. doi: 10.1016/j.rser.2015.04.174.
- [18] L. J. G. Aguiar et al., "Modeling the photosynthetically active radiation in South West Amazonia under all sky conditions," *Theoretical and Applied Climatology*, vol. 108, no. 3–4. Springer Science and Business Media LLC, pp. 631–640, Dec. 02, 2011. doi: 10.1007/s00704-011-0556-z.
- [19] F. Ferrera-Cobos, J. M. Vindel, R. X. Valenzuela, and J. A. González, "Models for estimating daily photosynthetically active radiation in oceanic and mediterranean climates and their improvement by site adaptation techniques," *Advances in Space Research*, vol. 65, no. 8. Elsevier BV, pp. 1894–1909, Apr. 2020. doi: 10.1016/j.asr.2020.01.018.
- [20] N. Proutsos, S. Alexandris, A. Liakatas, P. Nastos, and I. X. Tsiros, "PAR and UVA composition of global solar radiation at a high altitude Mediterranean forest site," *Atmospheric Research*, vol. 269. Elsevier BV, p. 106039, May 2022. doi: 10.1016/j.atmosres.2022.106039.
- [21] J. F. Escobedo, E. N. Gomes, A. P. Oliveira, and J. Soares, "Ratios of UV, PAR and NIR components to global solar radiation measured at Botucatu site in Brazil," *Renewable Energy*, vol. 36, no. 1. Elsevier BV, pp. 169–178, Jan. 2011. doi: 10.1016/j.renene.2010.06.018.
- [22] T. K. Akitsu et al., "The variability and seasonality in the ratio of photosynthetically active radiation to solar radiation: A simple empirical model of the ratio," *International Journal of Applied Earth Observation and Geoinformation*, vol. 108. Elsevier BV, p. 102724, Apr. 2022. doi: 10.1016/j.jag.2022.102724.
- [23] R. Perez, P. Ineichen, R. Seals, J. Michalsky, and R. Stewart, "Modeling daylight availability and irradiance components from direct and global irradiance," *Solar Energy*, vol. 44, no. 5. Elsevier BV, pp. 271–289, 1990. doi: 10.1016/0038-092x(90)90055-h.
- [24] I. Alados, I. Foyo-Moreno, and L. Alados-Arboledas, "Photosynthetically active radiation: measurements and modelling," *Agricultural and Forest Meteorology*, vol. 78, no. 1–2. Elsevier BV, pp. 121–131, Jan. 1996. doi: 10.1016/0168-1923(95)02245-7.
- [25] C. P. Jacovides, F. S. Timvios, G. Papaioannou, D. N. Asimakopoulos, and C. M. Theofilou, "Ratio of PAR to broadband solar radiation measured in Cyprus," *Agricultural and Forest Meteorology*, vol. 121, no. 3–4. Elsevier BV, pp. 135–140, Feb. 2004. doi: 10.1016/j.agrformet.2003.10.001.
- [26] B. Hu, Y. Wang, and G. Liu, "Measurements and estimations of photosynthetically active radiation in Beijing," *Atmospheric Research*, vol. 85, no. 3–4. Elsevier BV, pp. 361–371, Sep. 2007. doi: 10.1016/j.atmosres.2007.02.005.
- [27] M. de Blas, A. García-Rodríguez, I. García, and J. L. Torres, "Validation and calibration of models to estimate photosynthetically active radiation considering different time scales and sky conditions," *Advances in Space Research*, vol. 70, no. 7. Elsevier BV, pp. 1737–1760, Oct. 2022. doi: 10.1016/j.asr.2022.07.005.
- [28] CIE, Standard. "Spatial Distribution of Daylight-CIE Standard General Sky." CIE S 011.1/E (2003).
- [29] A. García-Rodríguez, D. Granados-López, S. García-Rodríguez, M. Díez-Mediavilla, and C. Alonso-Tristán, "Modelling Photosynthetic Active Radiation (PAR) through meteorological indices under all sky conditions," *Agricultural and Forest Meteorology*, vol. 310. Elsevier BV, p. 108627, Nov. 2021. doi: 10.1016/j.agrformet.2021.108627.
- [30] S. Zainali et al., "Direct and diffuse shading factors modelling for the most representative agrivoltaic system layouts," *Applied Energy*, vol. 339. Elsevier BV, p. 120981, Jun. 2023. doi: 10.1016/j.apenergy.2023.120981.
- [31] D. Yang, "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," *Renewable and Sustainable Energy Reviews*, vol. 159. Elsevier BV, p. 112195, May 2022. doi: 10.1016/j.rser.2022.112195.

- [32] D. Yang et al., “Regime-dependent 1-min irradiance separation model with climatology clustering,” *Renewable and Sustainable Energy Reviews*, vol. 189. Elsevier BV, p. 113992, Jan. 2024. doi: 10.1016/j.rser.2023.113992.
- [33] J. Widmer, B. Christ, J. Grenz, and L. Norgrove, “Agrivoltaics, a promising new tool for electricity and food production: A systematic review,” *Renewable and Sustainable Energy Reviews*, vol. 192. Elsevier BV, p. 114277, Mar. 2024. doi: 10.1016/j.rser.2023.114277.
- [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] A. Carrara et al., “Radiation measurements at ICOS ecosystem stations,” *International Agrophysics*, vol. 32, no. 4. Walter de Gruyter GmbH, pp. 589–605, Dec. 01, 2018. doi: 10.1515/intag-2017-0049.
- [37] A. Ohmura et al., "Baseline Surface Radiation Network (BSRN/WCRP): new precision radiometry for climate research," *Bull. Am. Meteorol. Soc.*, vol. 79, pp. 2115–2136, 1998.
- [38] M. Oh et al., “Analysis of minute-scale variability for enhanced separation of direct and diffuse solar irradiance components using machine learning algorithms,” *Energy*, vol. 241. Elsevier BV, p. 122921, Feb. 2022. doi: 10.1016/j.energy.2021.122921.
- [39] H. E. Beck, N. E. Zimmermann, T. R. McVicar, N. Vergopolan, A. Berg, and E. F. Wood, “Present and future Köppen-Geiger climate classification maps at 1-km resolution,” *Scientific Data*, vol. 5, no. 1. Springer Science and Business Media LLC, Oct. 30, 2018. doi: 10.1038/sdata.2018.214.
- [40] C. A. Gueymard, “Revised composite extraterrestrial spectrum based on recent solar irradiance observations,” *Solar Energy*, vol. 169. Elsevier BV, pp. 434–440, Jul. 2018. doi: 10.1016/j.solener.2018.04.067.
- [41] I. Reda and A. Andreas, “Solar position algorithm for solar radiation applications,” *Solar Energy*, vol. 76, no. 5. Elsevier BV, pp. 577–589, 2004. doi: 10.1016/j.solener.2003.12.003.
- [42] A. Forstinger et al., “Expert Quality Control of Solar Radiation Ground Data Sets,” *Proceedings of the ISES Solar World Congress 2021*. International Solar Energy Society, 2021. doi: 10.18086/swc.2021.38.02.
- [43] P. Y. Tan and M. R. B. Ismail, “Photosynthetically active radiation and comparison of methods for its estimation in equatorial Singapore,” *Theoretical and Applied Climatology*, vol. 123, no. 3–4. Springer Science and Business Media LLC, pp. 873–883, Feb. 13, 2015. doi: 10.1007/s00704-015-1399-9.
- [44] C. A. Gueymard and J. A. Ruiz-Arias, “Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance,” *Solar Energy*, vol. 128. Elsevier BV, pp. 1–30, Apr. 2016. doi: 10.1016/j.solener.2015.10.010.
- [45] M. de Blas, A. García-Rodríguez, I. García, and J. L. Torres, “Validation and calibration of models to estimate photosynthetically active radiation considering different time scales and sky conditions,” *Advances in Space Research*, vol. 70, no. 7. Elsevier BV, pp. 1737–1760, Oct. 2022. doi: 10.1016/j.asr.2022.07.005.
- [46] M. Lefèvre et al., “McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions,” *Atmospheric Measurement Techniques*, vol. 6, no. 9. Copernicus GmbH, pp. 2403–2418, Sep. 17, 2013. doi: 10.5194/amt-6-2403-2013.
- [47] CAMS McClear. "CAMS McClear Radiation Service," SODA-Pro Web Services, Copernicus Atmosphere Monitoring Service, 2023. Available: <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] X. Sun, J. M. Bright, C. A. Gueymard, B. Acord, P. Wang, and N. A. Engerer, “Worldwide performance assessment of 75 global clear-sky irradiance models using Principal Component Analysis,” *Renewable and Sustainable Energy Reviews*, vol. 111. Elsevier BV, pp. 550–570, Sep. 2019. doi: 10.1016/j.rser.2019.04.006.

- [51] C. A. Gueymard, “A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects,” *Renewable and Sustainable Energy Reviews*, vol. 39. Elsevier BV, pp. 1024–1034, Nov. 2014. doi: 10.1016/j.rser.2014.07.117.
- [52] Y. Liu, Y. Liu, H. Cai, and J. Zhang, “An innovative short-term multihorizon photovoltaic power output forecasting method based on variational mode decomposition and a capsule convolutional neural network,” *Applied Energy*, vol. 343. Elsevier BV, p. 121139, Aug. 2023. doi: 10.1016/j.apenergy.2023.121139.
- [53] C. Rogers et al., “Daily leaf area index from photosynthetically active radiation for long term records of canopy structure and leaf phenology,” *Agricultural and Forest Meteorology*, vol. 304–305. Elsevier BV, p. 108407, Jul. 2021. doi: 10.1016/j.agrformet.2021.108407.
- [54] B. Yildiz, J. I. Bilbao, and A. B. Sproul, “A review and analysis of regression and machine learning models on commercial building electricity load forecasting,” *Renewable and Sustainable Energy Reviews*, vol. 73. Elsevier BV, pp. 1104–1122, Jun. 2017. doi: 10.1016/j.rser.2017.02.023.
- [55] M. Yamasaki Junior, R. Z. Freire, L. O. Seman, S. F. Stefenon, V. C. Mariani, and L. dos Santos Coelho, “Optimized hybrid ensemble learning approaches applied to very short-term load forecasting,” *International Journal of Electrical Power & Energy Systems*, vol. 155. Elsevier BV, p. 109579, Jan. 2024. doi: 10.1016/j.ijepes.2023.109579.
- [56] J. Murphy and J. Chanut, “Leveraging machine learning to automate regression model evaluations for large multi-site water-quality trend studies,” *Environmental Modelling & Software*, vol. 170. Elsevier BV, p. 105864, Dec. 2023. doi: 10.1016/j.envsoft.2023.105864.
- [57] A. H. Elsheikh, S. W. Sharshir, M. Abd Elaziz, A. E. Kabeel, W. Guilan, and Z. Haiou, “Modeling of solar energy systems using artificial neural network: A comprehensive review,” *Solar Energy*, vol. 180. Elsevier BV, pp. 622–639, Mar. 2019. doi: 10.1016/j.solener.2019.01.037.
- [58] A. H. Elsheikh, V. P. Katekar, O. L. Muskens, S. S. Deshmukh, M. A. Elaziz, and S. M. Dabour, “Utilization of LSTM neural network for water production forecasting of a stepped solar still with a corrugated absorber plate,” *Process Safety and Environmental Protection*, vol. 148. Elsevier BV, pp. 273–282, Apr. 2021. doi: 10.1016/j.psep.2020.09.068.
- [59] E. Ghandourah, Y. S. Prasanna, A. H. Elsheikh, E. B. Moustafa, M. Fujii, and S. S. Deshmukh, “Performance prediction of aluminum and polycarbonate solar stills with air cavity using an optimized neural network model by golden jackal optimizer,” *Case Studies in Thermal Engineering*, vol. 47. Elsevier BV, p. 103055, Jul. 2023. Doi: 10.1016/j.csite.2023.103055.
- [60] F. A. Essa, M. A. Elaziz, M. A. Al-Betar, and A. H. Elsheikh, “Performance prediction of a reverse osmosis unit using an optimized Long Short-term Memory model by hummingbird optimizer,” *Process Safety and Environmental Protection*, vol. 169. Elsevier BV, pp. 93–106, Jan. 2023. doi: 10.1016/j.psep.2022.10.071
- [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).