Frontiers in Life Sciences and Related Technologies 6(1) (2025) 45-52

Contents lists available at Dergipark

Frontiers in Life Sciences and Related Technologies

Journal homepage: http://dergipark.org.tr/en/pub/flsrt

Research article

# Assessing the performance of multivariate data analysis for predicting solar radiation using alternative meteorological variables

Mustafa Ozbuldu*[1], Yunus Emre Sekerli[1]

[1] *Hatay Mustafa Kemal University, Faculty of Agriculture, Department of Biosystems Engineering, 31060, Hatay, Türkiye*

**Abstract**

This research analyses how well the Partial Least Squares Regression models could predict the monthly average daily global solar radiation for seven stations in the Mediterranean region of Türkiye. Five model scenarios were created with the SARAH-3 satellite dataset from 2005 to 2023 and using ERA5-AG meteorological variables. These included maximum and minimum temperature configurations, dew point temperature, precipitation, wind speed, and vapor pressure. Different models were examined for their prediction success by using different criteria and assessing the models with varying performance evaluation benchmarks. Based on the results, the models were accurate, mainly when all the predictor variables were used. The highest predictive performance was observed at Burdur station with KGE=0.937, NSE=0.901, and RSR=0.322. The greater regional variations showcased the specific meteorological parameters' relevancy. The results also support the adequacy of the ERA5-AG dataset for climate modelling and resource evaluation purposes. Unlike traditional regression approaches, this study demonstrates the efficiency of PLSR in handling high-dimensional climatic datasets for solar radiation prediction. These findings support the reanalysis of data in renewable energy and agricultural applications, particularly in data-limited regions.

***Keywords:*** *ERA5-AG; Mediterranean region; partial least squares regression; solar radiation prediction*

## 1. Introduction

Solar radiation and climate data are crucial to understanding agricultural productivity, renewable energy, and environmental systems. Accurate modeling and forecasting of these factors are critical as climate change and its repercussions threaten food security and energy sustainability (Bai et al., 2024). The evolution of remote sensing with reanalysis datasets and compiled statistics gives rise to numerous opportunities that make predicting and analyzing solar radiation and climate elements possible (Farbo et al., 2024). Other recent studies also highlight the limitless future of Artificial Intelligence (AI) technology in calculating solar radiation across various locations (Rabault et al., 2025).

Solar radiation is the primary driver of photosynthesis, affecting plant growth, yield, and biomass accumulation (Fraga et al., 2024). According to studies, more than 50% of agricultural yield variation is directly due to climatic factors such as solar radiation and precipitation (Munnoli et al., 2023). Furthermore, the utilization efficiency of solar radiation is critical for maximizing crop biomass production. The capture and conversion of solar radiation into biomass are significantly influenced by factors such as leaf area index (LAI) and spatial distribution of plants (Koester et al., 2014; Sgarbossa et al., 2018; Kaur et al., 2024). For example, studies on maize showed that optimizing plant arrangement increased the efficiency of solar radiation use, leading to higher yields (Sgarbossa et al., 2018). Similarly, sugarcane and rice demonstrated a positive relation between higher solar radiation and yield values (Marin and Carvalho, 2012; Wang et al., 2016). These findings highlight the need for precise solar radiation data that can be used in crop modeling and management practices to successfully

* Corresponding author.
E-mail address: mustafaozbuldu@mku.edu.tr (M. Ozbuldu).
https://doi.org/10.51753/flsrt.1590684 Author contributions
Received 24 November 2024; Accepted 19 April 2025
Available online 30 April 2025

optimize agricultural outputs (Castro et al., 2018; Perdinan et al., 2021). In addition, understanding the sensitivity of vegetation to radiation and soil moisture changes is another vital issue in regions facing water scarcity (Liu et al., 2025).

Remote sensing is essential in solar radiation analysis because it provides comprehensive spatial and temporal coverage. High-resolution solar radiation data, which is critical for various applications ranging from crop modeling to renewable energy system optimization, can be obtained with satellite-derived datasets, such as SARAH and Himawari-8 (Kaskaoutis and Polo, 2019; Hama et al., 2020; Ghazouani et al., 2021; Kong et al., 2024; Pfeifroth et al., 2024).

Reanalysis data like ERA5 provide comprehensive, high-quality datasets by integrating ground-based observations with model outputs. These datasets help agricultural decision-making, hydrological modeling, and climate research by addressing data gaps in remote and under-monitored regions (Katsekpor et al., 2024; Soci et al., 2024). For instance, since reanalysis data can effectively represent weather conditions, it can help in runoff simulations and irrigation planning in mountainous regions where data is scarce (Wang et al., 2024). In addition, it can be used in agricultural systems ranging from predicting evapotranspiration to optimizing growing seasons and evaluating the effects of climate on crop yield (Ishak et al., 2010; Hama et al., 2020; Pelosi et al., 2020; Araújo et al., 2022).

In cases where data sets have high dimensionality and multicollinearity, Partial Least Square Regression (PLSR) is a versatile technique that can predict complex interactions between variables. The PLSR technique differs from traditional regression techniques by allowing efficient dimensionality reduction without losing predictive accuracy by removing latent variables that maximize the covariance between predictors and response variables (Wangeci et al., 2024). This aspect makes PLSR ideal for climatic and agricultural studies where large and highly correlated datasets must be analyzed (Wu et al., 2022; Li et al., 2023; Dai et al., 2024). For instance, PLSR was successfully used in a study on rice genotypes to examine how different sowing dates influenced spikelet formation about solar radiation and temperature (Wang et al., 2019). In another study for the assessment of biophysical parameters of grassland, PLSR was used alongside reflectance data to improve the prediction of solar radiation effects on vegetation (Sakowska et al., 2016).

Moreover, these findings prove further use in remote sensing and reanalysis, thus enhancing agricultural resource management and decision-making. The distinctions become clear when assessing the performance and applicability of traditional station-based techniques to remote sensing methods for solar radiation and climate data. Classical methods using ground measurements, such as pyranometers and solarimeters, can determine solar radiation values for a given location with high precision and accuracy. These methods capture acceptable temporal variations and changes like direct and scattered components (Teke et al., 2015). However, these methods are geographically limited to a specific location. They require networks with multiple stations to transmit data to large areas, which can be logistically challenging and require high investment costs. Moreover, local conditions may affect ground measurement, such as the shading of close buildings or greenery, which are not typical of the regional environment (Harmsen et al., 2014; Olpenda et al., 2018).

In contrast, remote sensing techniques provide extensive spatial and temporal access, unlike traditional station-based methods, allowing continuous large-scale solar radiation monitoring. These techniques use satellite imagery and complex models of the atmosphere to predict radiation over many terrains, thus filling essential gaps in regions not covered by ground-based systems (Irvem and Ozbuldu, 2018; Polo and Kaskaoutis, 2023). More excellent remote sensing coverage benefits agriculture and renewable energy in areas with limited ground-based infrastructure (Kosmopoulos et al., 2018; Hama et al., 2020).

Remote sensing techniques are associated with disadvantages such as atmospheric conditions (cloud cover, haze, and aerosols) that create uncertainty in the radiation prediction, reducing accuracy. Such problems often encourage combining remote sensing data with classical ground-based measurements to increase the precision of radiation measurements. Studies emphasize that the combination of these techniques provides an ideal solution due to the adequacy and unlimited coverage of satellite data and the sensitivity of ground-based methods for agricultural and renewable energy systems (Zhou et al., 2017; Wang et al., 2021). In addition to these accuracy-enhancement methods, the use of remote sensing along with advanced machine learning (ML) post-processing procedures can also help solve these issues (Rabault et al., 2025).

The applications of solar radiation data in agriculture extend to renewable energy systems. In this research, the combining of remote sensing reanalysis datasets and PLSR was examined to improve solar radiation prediction from climatic variables. Using these tools, it was aimed to enhance the agricultural productivity, optimize energy systems, and deepen the understanding of climatic processes. Although the application of machine learning and statistical techniques to predict solar radiation is on the rise, there is limited research on evaluating PLSR using reanalysis datasets. In order to fill this gap, the present work tested the prediction performance of PLSR against ERA5-AG and SARAH-3 datasets, assessing its effectiveness with high-dimensional climate variables under varying meteorological conditions in the Mediterranean region.

## 2. Materials and methods

### *2.1. Study area and dataset*

This work was carried out using the average monthly solar radiation data of the provincial centers of Adana, Hatay, Osmaniye, Antalya, Mersin, Isparta, and Burdur in the Mediterranean region of Türkiye. The data was retrieved from the European Exploitation of Meteorological Satellites Organization's (EUMETSAT) SARAH-3 (Surface Radiation Data Set-Heliosat 3) dataset for the years 2005-2023. The study area is shown in Fig. 1. The SARAH data set is critical in understanding the solar radiation dynamics of Europe, and therefore, is an integral part of the solar radiation European system (CM SAF).

Compared to its predecessors, SARAH-3 is enhanced by lower-resolution solar radiation data from the METEOSAT satellite series. This data is supplied with a spatial resolution of 0.05°. The surface radiation value is calculated using the Heliosat technique (Thomas et al., 2023).

Data has undergone processing, cross-referencing satellite images and using ground truthing algorithms to ascertain precision on the data being analyzed (Mikelsons et al., 2022). With the satellite's ground observation, it was possible to create more reliable, optimized datasets critical for climate monitoring and modelling (Pfeifroth et al., 2018; Manara et al., 2020).

Inclusion for its systematic nature makes the dataset's accessibility and comprehensiveness unparalleled for most researchers and analysts, including policymakers and the industry (Kothe et al., 2017). The selected SARAH-3 solar radiation data, with its corresponding metadata, can be found at the JRC website (JRC, 2024). The solar radiation data from SARAH3 was analysed and tested for quality using the Kolmogorov-Smirnov data normalization tests individually per station in R Studio. Statistical granularity of the presented test results and data sets is highlighted in Table 1. Normality tests reported that all stations met the requirements, $p<0.05$ reported (Azad et al., 2024). ERA5-AG and SARAH-3 were chosen due to their long-term availability, high geographical accuracy, and spatial detail. Their reliability has been previously proven by comparing them with *in situ* measurements.

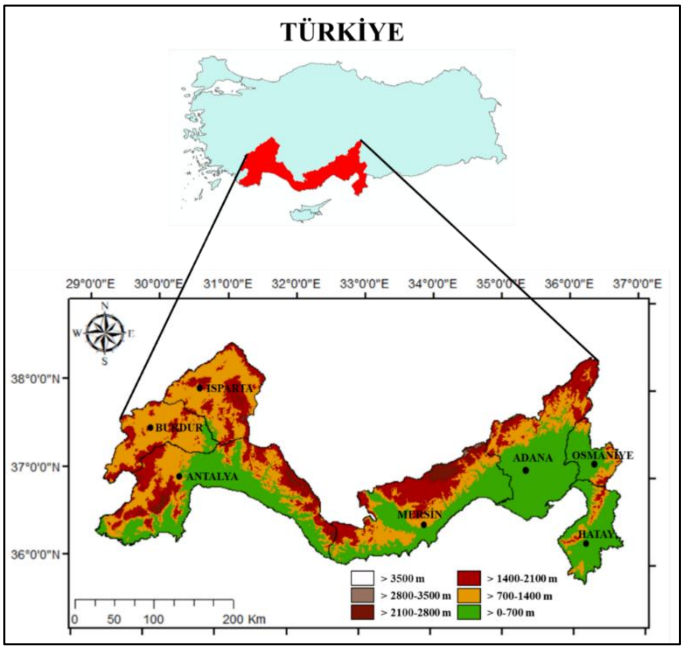


**Fig. 1.** Location of the study areas.

**Table 1**
Descriptive statistical values of monthly average solar radiation (kWh $m^{-2}$ $day^{-1}$) data.

| Station | Average | Standard Deviation | Max. | Min. | Kolmogoro-Smirnov p value |
|---|---|---|---|---|---|
| Adana | 5.03 | 1.97 | 8.09 | 1.84 | 0.002 |
| Hatay | 5.18 | 2.24 | 8.46 | 1.36 | 0.001 |
| Osmaniye | 4.86 | 1.93 | 8.11 | 1.81 | 0.003 |
| Mersin | 5.07 | 1.98 | 8.19 | 1.76 | 0.001 |
| Antalya | 5.21 | 2.07 | 8.44 | 1.97 | 0.001 |
| Burdur | 5.10 | 2.10 | 8.60 | 1.85 | 0.003 |
| Isparta | 4.84 | 2.03 | 8.45 | 1.57 | 0.006 |

### *2.2. Model inputs and creation of scenarios*

In the study, maximum temperature ($T_{max}$), minimum temperature ($T_{min}$), average dew point temperature ($T_{dew}$), precipitation (P), wind speed (WS) and vapor pressure (VP) were used as predictor variables to be used in the models that will predict the monthly average solar radiation. The ERA5-AG reanalysis dataset is a notable improvement in climate data as it contains high-resolution meteorological data from ECMWF's 5th generation of atmospheric reanalysis (ERA5)-a new and improved version. It benefits agricultural, hydrological, and environmental studies because it covers a wide area and has a high spatial resolution of 0.1° x 0.1° (about 10 km) (Zhou and Ismaeel, 2020). Simanjuntak et al. (2022) claimed that the dataset is essential for agriculture because it contains wind, solar, and other parameters for studying and modeling the environment or weather-changing factors. This research acquired the required ERA5-AG data from Google Earth Engine for the corresponding study areas. Five scenarios concerning model inputs were structured with the data discussed above to predict the impact of climate parameters. The model scenarios are presented in Table 2. Initial tests showed that wind speed had little impact in the stable climate regions, while precipitation had more significant impacts in the coastal areas. Atmospheric conditions (with accompanying changes in wind speeds) could considerably affect energy production, which indicates that under more stable conditions (low variability in wind), the wind speed does not have much impact. Furthermore, in areas with changing precipitation patterns, precipitation can diminish solar radiation by obstructing light in the region (Vizzo et al., 2021; Pérez et al., 2023).

**Table 2**
Model input scenarios.

| Model Scenarios | Input variables |
|---|---|
| M1 | $T_{max} + T_{min} + T_{dew} + P + WS + VP$ |
| M2 | $T_{max} + T_{min} + T_{dew}$ |
| M3 | $T_{max} + T_{min} + T_{dew} + P$ |
| M4 | $T_{max} + T_{min} + T_{dew} + VP$ |
| M5 | $T_{max} + T_{min} + T_{dew} + WS$ |

### *2.3. Multivariate data analysis*

The data were analyzed to assess the potential for predicting solar radiation based on climatic variables, including $T_{min}$, $T_{max}$, $T_{dew}$, WS, and VP. Regression analysis were conducted using the Partial Least Squares Regression method, implemented in the multivariate statistical software UnScrambler (version 9.7, Camo, Oslo, Norway). The PLS method was selected as more suitable than other classical techniques (such as Multiple Linear Regression and Principal Component Regression) for datasets with highly correlated variables (Esbensen, 2009). In the PLS analysis, 70% of the data (from 2005 to 2017) were used for model calibration, while the remaining 30% (from 2018 to 2023) served as the validation set.

### *2.4. Criteria for evaluating model performance*

In this study, the results received from the PLS regression model were evaluated using five separate performance assessment metrics. The coefficient of determination ($R^2$) is one of the most important criteria used to evaluate the regression model's goodness of fit. As stated in (Kasuya, 2018), "A higher $R^2$ value indicates a better fit." Root Mean Square Error (RMSE) has emerged as one of the most common metrics used to evaluate the accuracy of predictive models. It captures the discrepancies between predicted values ($M_i$) and observed values ($O_i$) and assesses the overall effectiveness of the model (Chai and Draxler, 2014). It is widely accepted that lower RMSE values suggest improved model performance. However, the actual value of this parameter depends on data dataset size. On the contrary, the Root Mean Standard Deviation Ratio (RSR) was introduced by (Singh et al., 2005) as a model comparison statistic that improves the interpretability of these values. RSR adjusts the RMSE values by the standard deviation of the observations to produce a constant value.

As for the calculation of the NSE, its coefficient is a metric to measure the predictive power of a model. The coefficient can take any value lower than 1, preferably zero or higher. Closer to 1 suggests that the model prediction result is satisfactory (Moriasi et al., 2007). The KGE is one of the most common metrics used for model evaluation, especially for hydrological models. KGE is composed of three main components: the correlation coefficient (r), the variability ratio (α) and the mean bias (β), making the model evaluation much more informative (Smit and Van Tol, 2022). These model performance metrics are derived from the results calculated in Equations 1-5. The quantitative results from these equations were benchmarked against the qualitative evaluations listed in Table 3 and Table 4.

$$R^2 = \left( \frac{\sum_{i=1}^{n}(O_i - \bar{O}) \times (M_i - \bar{M})}{\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2} \times \sqrt{\sum_{i=1}^{n}(M_i - \bar{M})^2}} \right)^2 \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n}(O_i - M_i)^2} \quad (2)$$

$$RSR = \frac{RMSE}{\sigma_{obs}} \quad (3)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - M_i)^2}{\sum_{i=1}^{n}(O_i - \bar{O})^2} \quad (4)$$

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (5)$$

**Table 3**
NSE and RSR performance evaluation table (Singh et al., 2005; Moriasi et al., 2007).

| Performance | NSE | RSR | $R^2$ |
|---|---|---|---|
| Very good | 0.75<NSE≤1.00 | 0.00≤RSR≤0.50 | $0.90 \le R^2 \le 1.00$ |
| Good | 0.65<NSE≤0.75 | 0.50<RSR≤0.60 | $0.75 \le R^2 < 0.90$ |
| Satisfactory | 0.50<NSE≤0.65 | 0.60<RSR≤0.70 | $0.50 \le R^2 < 0.75$ |
| Unsatisfactory | NSE≤0.50 | RSR>0.70 | $0.50 > R^2$ |

**Table 4**
KGE performance evaluation table (Towner et al., 2019).

| Performance | KGE |
|---|---|
| Good | KGE>0.75 |
| Intermediate | 0.5<KGE<0.75 |
| Poor | 0<KGE< 0.5 |
| Very poor | KGE<0 |

## 3. Result and discussion

This study evaluated PLS regression prediction models for their ability to predict average monthly solar radiation at seven different stations in the Mediterranean region of Türkiye under five different input scenarios. In the models created using variables obtained from the ERA5-AG reanalysis dataset, 70% of the dataset (2005-2017) was used as training and 30% (2018-2023) as a test dataset. The performance results of the test period, monthly average daily global solar radiation predictions obtained from the models for all stations, are given in Table 5.

Table 5 shows that the models generally have high values for KGE and NSE and low values for RSR. This indicates that the error rate of the models is low, and they have a good prediction performance. The M1 model showed the best performance in most of the stations. This is since model includes all variables ($T_{max}$, $T_{min}$, $T_{dew}$, P, WS, VP) and therefore represents the atmospheric conditions most comprehensively. The fact that M1 has higher KGE and NSE values emphasizes the importance of considering meteorological components together. Among the other models, performance decreases were generally observed when the number of variables was reduced. The model M2 with only temperature variables typically produced lower KGE and NSE values.

Considering the results obtained on a station basis, the M4 model stands out with the highest KGE (0.907) and NSE (0.854) values at Adana station. Moreover, the lowest RSR value (0.356) shows that the model minimizes the error rate. This result can be attributed to Adana’s high humidity and low wind speed variability. Therefore, VP is considered to be more critical in solar radiation prediction. On the other hand, the results show that the exclusion of WS and P variables from the prediction model may be less effective for Adana.

**Table 5**
Prediction performances of the models based on different input scenarios.

| Stations | MODELS | KGE | NSE | RSR |
|---|---|---|---|---|
| Adana | M1 | 0.911 | 0.851 | 0.363 |
| | M2 | 0.892 | 0.842 | 0.365 |
| | M3 | 0.906 | 0.843 | 0.372 |
| | **M4** | **0.907** | **0.854** | **0.356** |
| | M5 | 0.878 | 0.833 | 0.369 |
| Antalya | **M1** | **0.873** | **0.760** | **0.459** |
| | M2 | 0.851 | 0.725 | 0.480 |
| | M3 | 0.875 | 0.758 | 0.465 |
| | M4 | 0.863 | 0.743 | 0.471 |
| | M5 | 0.861 | 0.748 | 0.462 |
| Burdur | **M1** | **0.937** | **0.901** | **0.322** |
| | M2 | 0.914 | 0.832 | 0.398 |
| | M3 | 0.929 | 0.859 | 0.367 |
| | M4 | 0.921 | 0.843 | 0.389 |
| | M5 | 0.933 | 0.879 | 0.349 |
| Hatay | **M1** | **0.927** | **0.928** | **0.249** |
| | M2 | 0.818 | 0.685 | 0.496 |
| | M3 | 0.808 | 0.660 | 0.514 |
| | M4 | 0.792 | 0.665 | 0.495 |
| | M5 | 0.935 | 0.902 | 0.296 |
| Isparta | M1 | 0.875 | 0.816 | 0.390 |
| | M2 | 0.873 | 0.778 | 0.434 |
| | M3 | 0.871 | 0.776 | 0.435 |
| | M4 | 0.879 | 0.786 | 0.429 |
| | **M5** | **0.917** | **0.861** | **0.353** |
| Mersin | M1 | 0.901 | 0.862 | 0.346 |
| | M2 | 0.872 | 0.852 | 0.346 |
| | M3 | 0.890 | 0.844 | 0.366 |
| | **M4** | **0.910** | **0.861** | **0.357** |
| | M5 | 0.844 | 0.854 | 0.331 |
| Osmaniye | M1 | 0.907 | 0.881 | 0.318 |
| | M2 | 0.898 | 0.848 | 0.360 |
| | M3 | 0.899 | 0.846 | 0.363 |
| | M4 | 0.899 | 0.848 | 0.361 |
| | **M5** | **0.912** | **0.884** | **0.316** |

The most successful model at Antalya station was M1 (KGE=0.873, NSE=0.760, RSR=0.459). This model, in which all inputs were included, was able to reflect the complex meteorological structure of Antalya in the best way. Because Antalya is in a geography, where coastal and mountainous areas

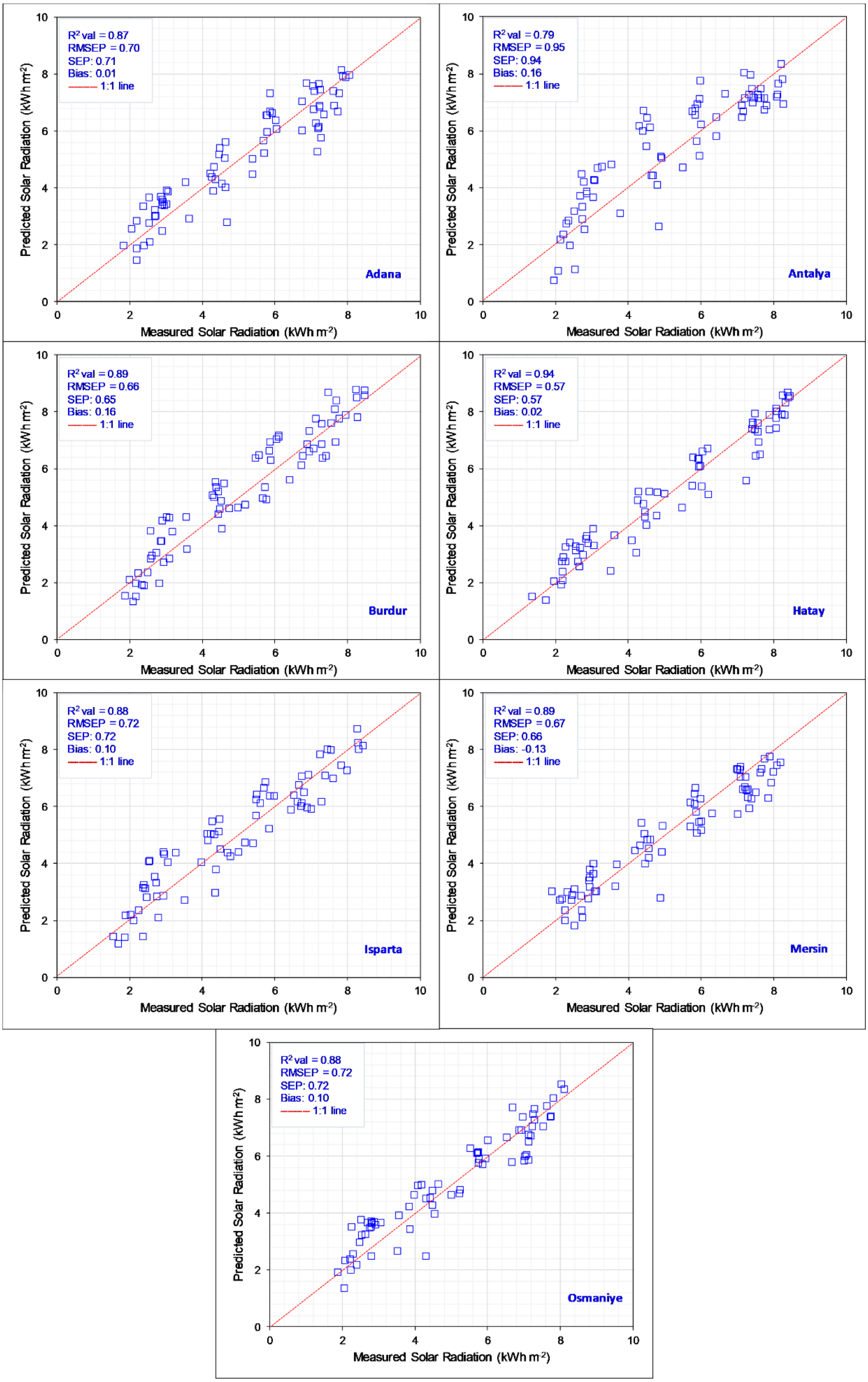


**Fig. 2.** Scatter plots of monthly average daily global solar radiation predictions obtained for each station for the test period.

merge, parameters such as wind speed and precipitation are important factors affecting solar radiation. Therefore, the inclusion of all parameters in M1 provided better performance of the model.

In Burdur station, the M1 model showed the best performance (KGE=0.937, NSE=0.901, RSR=0.322). It is known that Burdur is under the influence of a terrestrial climate and has low humidity levels. Therefore, the effect of parameters such as $T_{dew}$, VP, and WS on solar radiation becomes more noticeable. Because M1 includes all parameters, it provided high accuracy at this station. The most successful model at Hatay station was M1 (KGE=0.927, NSE=0.928, RSR=0.249). Although Hatay has a Mediterranean climate similar to Adana, the inclusion of all parameters produced a more successful result due to geographical differences. Including all inputs in M1 improved the performance of the model, particularly if factors such as precipitation and wind speed affect solar radiation.

The M5 model provided the highest performance at Isparta station (KGE=0.917, NSE=0.861, RSR=0.353). The M5 model considered WS instead of VP. Although Isparta is under the influence of a continental climate, it is seen that the land structure makes the changes in wind speed more important for solar radiation prediction. This shows that WS is a more effective parameter for predicting solar radiation in this region.

The M4 model provided the best performance at Mersin station (KGE=0.910, NSE=0.861, RSR=0.357). The inclusion of vapor pressure in M4 indicates that humidity levels in Mersin are a determining factor in solar radiation prediction. The most successful model at Osmaniye station was M5 (KGE=0.912, NSE=0.884, RSR=0.316). The inclusion of wind speed in the M5 model indicates that the meteorological conditions in Osmaniye play a determining role in solar radiation. The main reason for the differences in the performance of the models is the variations in each station's meteorological, geographical, and environmental conditions. In humid and hot regions such as Adana, vapor pressure is critical for prediction performance. In contrast, the effect of wind speed may be more pronounced in areas with terrestrial climate characteristics, such as Burdur. Similarly, in coastal regions such as Antalya and Hatay, the effect of high humidity and wind speed should be considered together. Scatter plots generated according to the predictions obtained from the most successful models for each station are given in Fig. 2. The performance differences among stations may be linked to their geographical and meteorological characteristics. For instance, Burdur's stable conditions enhanced model accuracy, whereas Antalya's coastal variability led to slightly lower performance.

In general, the predicted values are in good agreement with the measured data. In particular, the predictions obtained for the models at Hatay, Mersin, and Burdur stations are very close to the 1:1 line, indicating that over- or under-predicted values are limited. According to the scatter plots, it is concluded that the predicted values of monthly average global solar radiation have a high level of accuracy.

The $R^2$ values obtained from the models (0.79-0.94) show that the model predictions are highly accurate. According to the RMSE results (0.57-0.95), it was determined that the models predicted the monthly average daily global solar radiation with a very low error amount. In similar modelling studies where solar radiation prediction models were evaluated, Mohammadi et al. (2015) calculated the performance of the regression prediction model as $R^2$=0.84 in their research in Iran. Shamshirband et al. (2016) calculated $R^2$=97.37 and RMSE=0.18 in their study using machine learning in Iran. Karaman et al. (2021) calculated the RMSE value as 0.0297 and the $R^2$ value as 0.99 in their prediction model using ANN in Karaman province in Türkiye. Soria et al. (2022) used a multiple linear regression model in their study in Peru and found the $R^2$ value as 0.556. These results show that the PLS regression model method is much more successful than the models created with traditional regression approaches, although not as successful as ANN and machine learning based approaches. In addition, the findings of this study show that the ERA5-AG reanalysis dataset can be used for solar radiation prediction in developing countries such as Türkiye, where the meteorological observation gauge network is sparse.

## 4. Conclusion

This study created models for solar radiation prediction using PLSR from climate variables of ERA5-AG reanalysis datasets in different scenarios. The prediction success of these models was evaluated based on various criteria. The study conducted in the seven Türkiye's Mediterranean Region stations found that PLSR models effectively processed high-dimensional and multicollinear climate data. Among all the evaluated models, those that included all features (temperature, dew point, wind speed, and vapor pressure) provided the best predicted solar radiation accuracy as indicated by KGE, NSE, $R^2$, and low RSR values. The findings also highlight the role of local climate and topographic features in the prediction model developing process. Indicating this, including wind speed and precipitation variables, is essential in coastal regions such as Antalya and Hatay. In contrast, wind speed data has become particularly important in Isparta, which has a more continental climate. This study underscores PLSR models' versatility in fitting regional datasets, which provide a reasonable balance between precision and model complexity when combined with classical regression and sophisticated machine learning techniques. According to study results, the ERA5-AG dataset can be a valid resource for predicting solar radiation in regions with limited data availability. This research helps to reduce the gaps created by insufficient meteorological observation networks, while also helping to maximize solar energy use and improve agricultural productivity. These results are crucial in optimizing photovoltaic energy use in areas where data is scarce. Accurate solar radiation prediction also helps support climate resilience strategies and enhances agricultural planning by predicting the potential crop yields.

**Conflict of interest:** The authors declare that they have no conflict of interests.

**Informed consent:** The authors declare that this manuscript did not involve human or animal participants and informed consent was not collected.

## References

Araújo, C. S .P., Silva, I. A. C., Ippolito, M., & Almeida, C. D. G. C. (2022). Evaluation of air temperature estimated by ERA5-Land reanalysis using surface data in Pernambuco, Brazil. *Environmental Monitoring and Assessment, 194*(5), 381.

Azad, M. A. K., Mallick, J., Islam, A. R. M. T., Ayen, K., & Hasanuzzaman, M. D. (2024). Estimation of solar radiation in data-scarce subtropical region using ensemble learning models based on a novel CART-based feature selection. *Theoretical and Applied Climatology, 155*, 349-369.

Bai, J. M., Xiao-Wei, W., Arslan, E., & Zong, X. (2024). Global solar radiation and its interactions with atmospheric substances and their effects on air temperature change in Ankara province. *Climate, 12(3)*, 35.

Castro, J. R., Cuadra, S. V., Pinto, L. B., Souza, J. M. H., Santos, M. P., & Heinemann, A. B. (2018). Parametrization of models and use of estimated global solar radiation data in the irrigated rice yield simulation. *Revista Brasileira de Meteorologia, 33*(2), 238-246.

Chai, T. & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development, 7*(3), 1247-1250.

Dai, Y., Yu, S., Ma, T., Ding, J., Chen, K., Zeng, G., & Zhang, M. (2024). Improving the estimation of rice above-ground biomass based on spatio-temporal uav imagery and phenological stages. *Frontiers in Plant Science, 15*.

Esbensen, K. H. (2009). *Multivariate data analysis—In practice* (5th ed.). Camo Software AS.

Farbo, A., Trombetta, N. G., Palma, L. d., & Mondino, E. B. (2024). Estimation of intercepted solar radiation and stem water potential in a table grape vineyard covered by plastic film using sentinel-2 data: a comparison of ols-, mlr-, and ml-based methods. *Plants, 13(9)*, 1203.

Fraga, H., Freitas, T. R., Moriondo, M., Molitor, D., & Santos, J. A. (2024). Determining the climatic drivers for wine production in the côa region (portugal) using a machine learning approach. *Land, 13(6)*, 749.

Ghazouani, N., Bawadekji, A., El-Bary, A. A., Elewa, M. M., Becheikh, N., Alassaf, Y., & Hassan, G. (2021). Performance evaluation of temperature-based global solar radiation models—case study: Arar City, KSA. *Sustainability, 14*, 35.

Hama, A., Tanaka, K., Mochizuki, A., Tsuruoka, Y., & Kondoh, A. (2020). Improving the UAV-based yield estimation of paddy rice by using the solar radiation of geostationary satellite Himawari-8. *Hydrological Research Letters, 14*(1), 56-61.

Harmsen, E. W., Cruz, P. T., & Mecikalski, J. R. (2014). Calibration of selected pyranometers and satellite-derived solar radiation in Puerto Rico. *International Journal of Renewable Energy Technology, 5*(1), 43-54.

Irvem, A., & Ozbuldu, M. (2018). Accuracy of satellite-based solar data to estimate solar energy potential for Hatay Province, Turkey. *Bitlis Eren University Journal of Science, 7*(2), 361-369.

Ishak, A., Bray, M., Remesan, R., & Han, D. (2010). Estimating reference evapotranspiration using numerical weather modelling. *Hydrological Processes, 24*(24), 3490-3509.

JRC, (2024). Official Website of The Joint Research Centre: EU Science Hub, https://joint-research-centre.ec.europa.eu/, Last Accessed on March, 2025.

Karaman, Ö. A., Tanyıldızı Agir, T., & Arsel, İ. (2021). Estimation of solar radiation using modern methods. *Alexandria Engineering Journal, 60*, 2447-2455.

Kaskaoutis, D., & Polo, J. (2019). Editorial for the special issue "solar radiation, modeling, and remote sensing". *Remote Sensing, 11*(10), 1198.

Kasuya, E. (2018). On the use of r and r squared in correlation and regression. *Ecological Research, 34*(1), 235-236.

Katsekpor, J. T., Greve, K., Yamba, E. I., & Amoah, E. G. (2024). Comparative analysis of satellite and reanalysis data with ground-based observations in northern ghana. *Meteorological Applications, 31(4)*, e2226.

Kaur, N., Snider, J. L., Paterson, A. H., Virk, G., Parkash, V., Roberts, P. M., & Li, C. (2024). Genotypic variation in functional contributors to yield for a diverse collection of field-grown cotton. *Crop Science, 64(3)*, 1846-1861.

Koester, R. P., Skoneczka, J. A., Cary, T. R., Diers, B. W., & Ainsworth, E. A. (2014). Historical gains in soybean (Glycine max Merr.) seed yield are driven by linear increases in light interception, energy conversion, and partitioning efficiencies. *Journal of Experimental Botany, 65*(12), 3311-3321.

Kong, H., Wang, J., Cai, L., Cao, J., Zhou, M., & Fan, Y. (2024). Surface solar radiation resource evaluation of Xizang region based on station observation and high-resolution satellite dataset. *Remote Sensing, 16(8)*, 1405.

Kosmopoulos, P. G., Kazadzis, S., El-Askary, H. M., Taylor, M., Gkikas, A., Proestakis, E., Kontoes, C. H., & El-Khayat, M. M. (2018). Earth-observation-based estimation and forecasting of particulate matter impact on solar energy in Egypt. *Remote Sensing, 10*(12), 1870.

Kothe, S., Pfeifroth, U., Cremer, R., Trentmann, J., & Hollmann, R. (2017). A satellite-based sunshine duration climate data record for Europe and Africa. *Remote Sensing, 9*(5), 429.

Li, Q., Bessafi, M., & Li, P. (2023). Mapping prediction of surface solar radiation with linear regression models: case study over Reunion Island. *Atmosphere, 14*(9), 1331.

Liu, Y., Xiao, J., Li, X., & Li, Y. (2025). Critical soil moisture detection and water–energy limit shift attribution using satellite-based water and carbon fluxes over China. *Hydrology and Earth System Sciences, 29*, 1241-1258.

Manara, V., Stocco, E., Brunetti, M., Diolaiuti, G., Fugazza, D., Pfeifroth, U., & Maugeri, M. (2020). Comparison of surface solar irradiance from ground observations and satellite data (1990–2016) over a complex orography region (Piedmont—northwest Italy). *Remote Sensing, 12*(23), 3882.

Marin, F. R., & Carvalho, G. L. (2012). Spatio-temporal variability of sugarcane yield efficiency in the state of São Paulo, Brazil. *Pesquisa Agropecuária Brasileira, 47*(2), 149-156.

Mikelsons, K., Wang, M., Kwiatkowska, E., Jiang, L., Dessailly, D., & Gossn, J. I. (2022). Statistical evaluation of sentinel-3 OLCI ocean color data retrievals. *IEEE Transactions on Geoscience and Remote Sensing, 60*, 1-19.

Mohammadi, K., Shamshirband, S., Anisi, M. H., Alam, K. A., & Petković, D. (2015). Support vector regression based prediction of global solar radiation on a horizontal surface. *Energy Conversion and Management, 91*, 433-441.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the American Society of Agricultural and Biological Engineers, 50*(3), 885-900.

Munnoli, S., Shivashankar, K., & Golshetti, A. (2023). Influences of changes in rainfall and solar radiation on performance of kharif sorghum. *International Journal of Environment and Climate Change, 13*(6), 186-193.

Olpenda, A., Stereńczak, K., & Będkowski, K. (2018). Modeling solar radiation in the forest using remote sensing data: a review of approaches and opportunities. *Remote Sensing, 10*(5), 694.

Pelosi, A., Terribile, F., D'Urso, G., & Chirico, G. B. (2020). Comparison of ERA5-Land and UERRA MESCAN-SURFEX reanalysis data with spatially interpolated weather observations for the regional assessment of reference evapotranspiration. *Water, 12*(6), 1669.

Perdinan, Winkler, J. A., & Andresen, J. A. (2021). Evaluation of multiple approaches to estimate daily solar radiation for input to crop process models. *Atmosphere, 12*(1), 8.

Pérez, C., Rivero, M., Escalante, M., Ramirez, V., & Guilbert, D. (2023). Influence of atmospheric stability on wind turbine energy production: A case study of the coastal region of Yucatan. *Energies, 16*(10), 4134.

Pfeifroth, U., Sánchez-Lorenzo, A., Manara, V., Trentmann, J., & Hollmann, R. (2018). Trends and variability of surface solar radiation in Europe based on surface- and satellite-based data records. *Journal of Geophysical Research: Atmospheres, 123*(3), 1735-1754.

Pfeifroth, U., Drücke, J., Kothe, S., Trentmann, J., Schröder, M., & Hollmann, R. (2024). SARAH-3 – satellite-based climate data records of surface solar radiation. *Earth System Science Data, 16*, 5243-5265.

Polo, J. & Kaskaoutis, D. (2023). Editorial on new challenges in solar radiation, modeling and remote sensing. *Remote Sensing, 15*(10), 2633.

Rabault, J., Sætra, M. L., Dobler, A., Eastwood, S., & Berge, E. (2025). Data fusion of complementary data sources using Machine Learning enables higher accuracy Solar Resource Maps. *Solar Energy, 290*, 113337.

Sakowska, K., Juszczak, R., & Gianelle, D. (2016). Remote sensing of grassland biophysical parameters in the context of the Sentinel-2 satellite mission. *Journal of Sensors, 2016*, 1-16.

Sgarbossa, J., Schwerz, F., Elli, E., Tibolla, L., Schmidt, D., & Caron, B. (2018). Agroforestry systems and their effects on the dynamics of solar radiation and soybean yield. *Comunicata Scientiae, 9*(3), 492-502.

Shamshirband, S., Mohammadi, K., Tong, C. W., Zamani, M., Motamedi, S., & Ch, S. (2016). A hybrid SVM-FFA method for prediction of monthly mean global solar radiation. *Theoretical and Applied Climatology, 125*(1-2), 53-65.

Simanjuntak, C., Gaiser, T., Ahrends, H. E., & Srivastava, A. K. (2022). Spatial and temporal patterns of agrometeorological indicators in maize producing provinces of South Africa. *Scientific Reports, 12*(1), 12072.

Singh, J., Knapp, H. V., Arnold, J. G., & Demissie, M. (2005). Hydrologic modeling of the Iroquois River watershed using HSPF and SWAT. *Journal of the American Water Resources Association, 41*(2), 361-375.

Smit, E., & Van Tol, J. (2022). Impacts of soil information on process-based hydrological modelling in the Upper Goukou Catchment, South Africa. *Water, 14*(3), 407.

Soci, C., Hersbach, H., Simmons, A. J., Poli, P., Bell, B., Berrisford, P., & Thépaut, J. (2024). The era5 global reanalysis from 1940 to 2022. *Quarterly Journal of the Royal Meteorological Society, 150(764)*, 4014-4048.

Soria, J. J., Poma, O., Sumire, D. A., Rojas, J. H. F., & Chipa, S. M. R. (2022). Multiple linear regression model of environmental variables, predictors of global solar radiation in the area of East Lima, Peru. *IOP Conference Series: Earth and Environmental Science, 1006*(1), 012009.

Teke, A., Yildirim, H. B., & Celik, O. (2015). Evaluation and performance comparison of different models for the estimation of solar radiation. *Renewable and Sustainable Energy Reviews, 50*, 1097-1107.

Thomas, C., Nyamsi, W. W., Arola, A., Pfeifroth, U., Trentmann, J., Dorling, S., & Alexandr, A. (2023). Smart approaches for evaluating photosynthetically active radiation at various stations based on MSG Prime satellite imagery. *Atmosphere, 14*(8), 1259.

Towner, J., Cloke, H. L., Zsoter, E., Flamig, Z., Hoch, J. M., Bazo, J., ... & Stephens, E. M. (2019). Assessing the performance of global hydrological models for capturing peak river flows in the Amazon basin. *Hydrology and Earth System Sciences, 23*(7), 3057-3080.

Vizzo, J. I., Cabrerizo, M. J., Helbling, E. W., & Villafañe, V. E. (2021). Extreme and gradual rainfall effects on winter and summer estuarine phytoplankton communities from Patagonia (Argentina). *Marine Environmental Research*, *163*, 105235.

Wangeci, A., Adén, D., Nikolajsen, T., Greve, M. H., & Knadel, M. (2024). Combining laser-induced breakdown spectroscopy and visible near-infrared spectroscopy for predicting soil organic carbon and texture: a Danish national-scale study. *Sensors*, *24(14)*, 4464.

Wang, D., Laza, M. R. C., Cassman, K. G., Huang, J., Nie, L., Ling, X., Centeno, G. S., Cui, K., Wang, F., Li, Y., & Peng, S. (2016). Temperature explains the yield difference of double-season rice between tropical and subtropical environments. *Field Crops Research, 198*, 303-311.

Wang, Y., Wang, L., Chen, H., Xiang, J., Zhang, Y., Shi, Q., Zhu, D., & Zhang, Y. (2019). Sowing dates have different effects on spikelet formation among different photoperiod-sensitive rice genotypes. *Agronomy Journal, 111*(5), 2263-2275.

Wang, Y., Zhou, T., & Zhou, W. (2021). Solar radiation distribution method for a photovoltaic greenhouse based on the maximization of annual economic benefits. *Research Square*.

Wang, X., Zhou, J., Ma, J., Luo, P., Fu, X., Feng, X., Zhang, X., Jia, Z., Wang, X., & Huang, X. (2024). Evaluation and comparison of reanalysis data for runoff simulation in the data-scarce watersheds of alpine regions. *Remote Sensing, 16*(5), 751.

Wu, X., Liu, H., Hartmann, H., Ciais, P., Kimball, J. S., Schwalm, C. R., Camarero, J. J., Chen, A., Gentine, P., Yang, Y., Zhang, S., Li, X., Xu, C., Zhang, W., Li, Z., & Chen, D. (2022). Timing and order of extreme drought and wetness determine bioclimatic sensitivity of tree growth. *Earth's Future, 10*(7), e2021EF002530.

Zhou, Q., Flores, A., Glenn, N., Walters, R., & Han, B. (2017). A machine learning approach to estimation of downward solar radiation from satellite-derived data products: an application over a semi-arid ecosystem in the U.S. *PLoS One, 12*(8), e0180239.

Zhou, Q., & Ismaeel, A. (2020). Seasonal cropland trends and their nexus with agrometeorological parameters in the Indus River plain. *Remote Sensing, 13*(1), 41.

***Cite as:*** Ozbuldu, M., & Sekerli, Y. E. (2025). Assessing the performance of multivariate data analysis for predicting solar radiation using alternative meteorological variables. *Front Life Sci RT, 6*(1), 45-52.