UNIVERSIDAD NACIONAL DE SAN MARTIN

RCSI
https://revistas.unsm.edu.pe/index.php/rcsi
e-ISSN: 2709-992X

Original article
Artículo original
Jan-Jun, 2026

VEGA-RAD: Hybrid physical-statistical model for the daily prediction of solar radiation in the Amazon

VEGA-RAD: Modelo híbrido físico–estadístico para la predicción diaria de radiación solar en la Amazonía

Agreda-Vega, Juan Francisco[1*]
Sare-Lara, Evergisto[1]
Rosales-Huamani, Jimmy Aurelio[2,3]

[1]Statistics and Applied Machine Learning Group, Universidad Nacional de San Martín, Tarapoto, Perú
[2]Department of Computer Science, Higher Polytechnic School, Universidad de Alcalá, Madrid, España
[3]SMultidisciplinary Sensing, Universal Accessibility and Machine Learning Group, Universidad Nacional de Ingeniería, Lima, Perú

Received: 05 Oct. 2025 | **Accepted:** 27 Dec. 2025 | **Published:** 20 Jan. 2026

Corresponding author*: juan.agreda@unsm.edu.pe

How to cite this article: Agreda-Vega, J. F., Sare-Lara, E. & Rosales-Huamani, J. A. (2026). VEGA-RAD: Hybrid physical-statistical model for the daily prediction of solar radiation in the Amazon. *Revista Científica de Sistemas e Informática*, *6*(1), 1454. https://doi.org/10.51252/rcsi.v6i1.1454

ABSTRACT

Daily solar radiation forecasting in the Peruvian Amazon represents a relevant challenge due to the high atmospheric variability that characterizes the region. In this study, VEGA-RAD (Vega Radiative Adaptive Dynamics) is formulated and evaluated as a hybrid physical–statistical model for daily solar radiation prediction in tropical environments. The model integrates an interpretable physical–astronomical proxy, stochastic temporal memory, and an adaptive statistical correction based on machine learning to capture residual nonlinearities. The analysis is conducted using daily ERA5 reanalysis data for the period 2017–2025, obtained through the Open-Meteo API. The results show a reduction in mean absolute error (MAE) from 1.699 to 0.477 kWh/m^2/d and an increase in the coefficient of determination (R^2) from 0.635 to 0.854. These improvements are supported by paired inferential analysis (Wilcoxon) and non-parametric bootstrap resampling. In addition, conformal prediction intervals achieve coverage levels consistent with the nominal 90 % and 95 % levels, with a temporally stable average width, indicating a conservative and reliable quantification of predictive uncertainty. The proposed VEGA-RAD model is presented as a reproducible, interpretable, and robust tool for energy applications in Amazonian contexts.

Keywords: machine learning; climate uncertainty; hybrid model; daily solar radiation

RESUMEN

La predicción diaria de la radiación solar en la Amazonía peruana es un desafío relevante debido a su elevada variabilidad atmosférica. En este estudio se formula y evalúa VEGA-RAD (Vega Radiative Adaptive Dynamics), un modelo híbrido físico–estadístico para la predicción diaria de radiación solar en regiones tropicales. El modelo integra un proxy físico–astronómico, memoria temporal estocástica y una corrección estadística adaptativa basada en aprendizaje automático para capturar no linealidades residuales. El análisis se realizó con datos diarios ERA5 (2017–2025) obtenidos mediante la API de Open-Meteo. Los resultados muestran una reducción del MAE de 1.699 a 0.477 kWh/m^2/d y un aumento del R^2 de 0.635 a 0.854. Estas mejoras fueron confirmadas mediante análisis inferencial pareado (Wilcoxon) y remuestreo bootstrap. Además, los intervalos conformales alcanzan coberturas coherentes con los niveles nominales del 90 % y 95 %, con ancho medio estable en el tiempo, con ancho medio estable, evidenciando una cuantificación de la incertidumbre conservadora y confiable. El modelo híbrido "VEGA-RAD" se presenta como una herramienta reproducible, interpretable y robusta para aplicaciones energéticas en contextos amazónicos.

Palabras clave: aprendizaje automático; incertidumbre climática; modelo híbrido; radiación solar diaria

1. INTRODUCTION

Solar radiation prediction is a central issue for the efficient integration of renewable energy into modern electrical systems. The high spatiotemporal variability of solar radiation affects the sizing of photovoltaic plants, the stability of power grids, and the formulation of sustainable energy policies, particularly in regions with high climatic complexity (Bakır, 2024; Demir, 2025; Shringi et al., 2025; Tandon et al., 2025; Yadav et al., 2025; Zerouali et al., 2025).

In tropical regions such as the Peruvian Amazon, this challenge is intensified due to persistent cloud cover, pronounced seasonality, and the scarcity of reliable meteorological stations. In this context, global reanalysis products have emerged as a robust alternative for characterizing solar radiation and associated atmospheric variables (Hersbach et al., 2020; Huang et al., 2021).

Among these sources, ERA5 stands out for its temporal consistency, global spatial coverage, and extensive validation across multiple climatic zones. Its open access through programming interfaces facilitates reproducibility and methodological transparency in solar radiation prediction studies (Demir, 2025; Open-Meteo, 2025).

In parallel, advances in machine learning and deep learning have enabled models capable of capturing complex nonlinear relationships between solar radiation and atmospheric factors. In particular, hybrid approaches based on CNN–SVR, CNN–LSTM, and metaheuristic optimization show substantial improvements in daily solar radiation prediction across different climates (Ghimire et al., 2022; Hamdaouy et al., 2025; Y.H. et al., 2024; Raju et al., 2025; Şener & Tuğal, 2025).

Recent review studies confirm that hybrid models dominate the state of the art by integrating physical knowledge, algorithmic optimization, and adaptive learning, consistently outperforming purely statistical or physical approaches (Celik et al., 2025; Ghareeb et al., 2025; Rajput et al., 2025; Shringi et al., 2025).

For intra-hour and intraday horizons, multimodal architectures that combine sky images with meteorological variables allow for highly accurate forecasting of rapid cloud changes. Proposals such as SkyNet and other multimodal models demonstrate high robustness against atmospheric variability (Abad-Alcaraz et al., 2025; Hou et al., 2025; Ruan et al., 2026).

Additionally, multiscale decomposition techniques, wavelet transforms, and ensemble methods reinforce predictive stability. The use of photovoltaic power data, satellite products, and exogenous variables also improves the representation of radiative dynamics (Arseven & Çınar, 2025; Cui et al., 2025; Solano & Affonso, 2023; Wu et al., 2025; Yan et al., 2025).

Other environmental factors, such as air quality, have emerged as relevant predictors of solar radiation attenuation. Recent studies show that including atmospheric pollutants significantly enhances the accuracy of machine learning–based models (Aladwani et al., 2025; Hu et al., 2025).

In medium- and long-term energy planning, evaluations based on bias-corrected CMIP6 projections and downscaling provide regional evidence of future changes in solar radiation. In this context, regional numerical models such as WRF–Solar and UV solar radiation studies confirm the high climate sensitivity of tropical regions (Amorim et al., 2024; Gao et al., 2025; Jadhav & Bhawar, 2025; Krishnan & Ravi Kumar, 2025; Vignesh Kumar et al., 2025; Alves et al., 2025; Zhu et al., 2025).

Despite these advances, a scientific gap remains in the development of models specifically designed for the Peruvian Amazon that integrate interpretable physical foundations, stochastic temporal memory, adaptive statistical correction, and explicit uncertainty quantification. In response to this gap, the present work proposes and formulates an original hybrid model called VEGA-RAD (Vega Radiative Adaptive Dynamics), aimed at daily solar radiation prediction using ERA5 data and well-calibrated point and probabilistic estimates.

2. MATERIALS AND METHODS

2.1. Study Area

The study area corresponds to the city of Tarapoto, located in the Peruvian Amazon (latitude −6.5°, longitude −76.3°). The region has a humid tropical climate, characterized by persistent cloud cover, high relative humidity, and pronounced intra- and interannual variability in solar radiation. These conditions make Tarapoto a particularly challenging environment for solar radiation prediction and the planning of photovoltaic systems.

2.2. Data

Daily ERA5 reanalysis data were obtained via the Open-Meteo API, an open-access platform that provides climate reanalysis products through RESTful web services. Specifically, the programmatic access endpoint for ERA5 is available at: https://archive-api.open-meteo.com/v1/era5

Data access was carried out through parameterized HTTP requests, specifying geographic coordinates, the time period, and the meteorological variables of interest. Official API documentation and query examples are available on the Open-Meteo portal (https://open-meteo.com/). The data were downloaded on September 26, 2025.

Table 1. Sample of Daily ERA5 Reanalysis Records for Tarapoto (01.01.2017–26.09.2025). Units: GHI in kWh/m^2/d, cloud cover in %, temperature in °C, relative humidity in %, and wind speed in m/s

	Date	GHI	cloud	t2m	rh2m	wind
1	01/2017	12.56	99	22.9	95	6.2
2	01/2017	17.56	88	23.3	87	10.6
3	01/2017	23.65	73	24.7	77	11.1
4	01/2017	18.3	88	25.2	75	14.8
5	01/2017	10.48	98	23.3	87	8.8

Table 1 presents a representative sample of the daily ERA5 reanalysis records, highlighting the joint variability between global solar radiation and the associated meteorological variables.

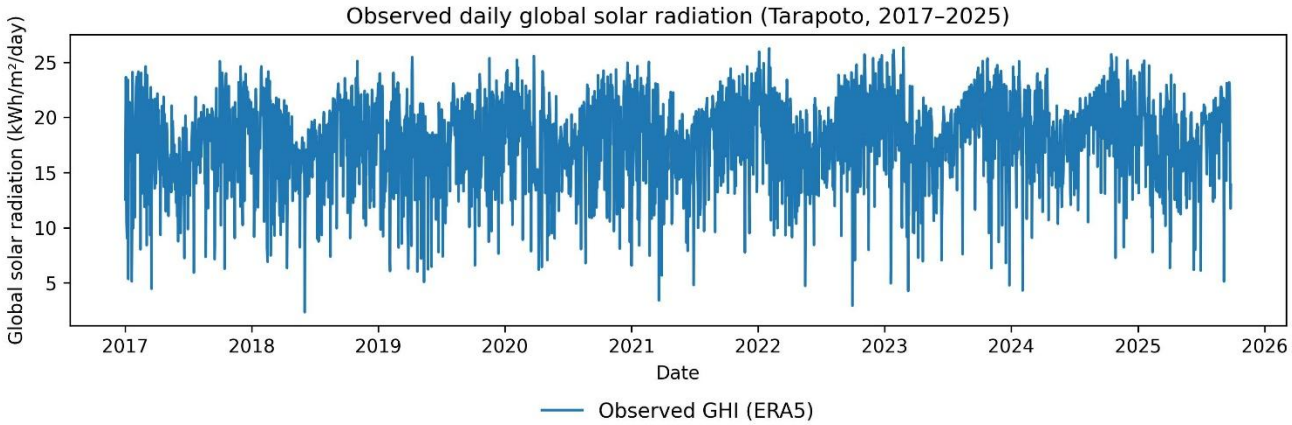

Figure 1. Time series of observed global horizontal solar radiation (ERA5) for Tarapoto (2017–2025). Annual seasonality and high daily variability associated with cloud cover and atmospheric conditions are evident

Figure 1 shows the continuous evolution of daily GHI, revealing seasonal and interannual patterns characteristic of the Amazonian climate, which justifies the use of a hybrid approach with temporal memory.

2.2.1. Data Preprocessing

Prior to the formulation and training of the VEGA-RAD model, daily ERA5 reanalysis data were subjected to a preprocessing procedure aimed at ensuring temporal coherence, numerical stability, and reproducibility

of the analysis. First, the downloaded records were organized chronologically, and days with missing values in any of the considered variables were discarded, resulting in a final dataset composed exclusively of complete observations.

Since ERA5 products incorporate global-scale quality control procedures, no additional outlier detection or removal methods were applied, in order to preserve physically plausible extremes associated with intense atmospheric events characteristic of the Amazonian climate. Likewise, no imputation of missing values was performed, as the analysis was restricted to days with complete information.

As part of temporal aggregation, daily series were used directly, consistent with the prediction horizon of the study. From these series, a physical–astronomical solar radiation proxy was constructed as a reference, serving as the deterministic component of the model, and a logarithmic residual between the observed radiation and this reference was defined, which constitutes the target variable for the machine learning block.

To capture temporal dependence and atmospheric persistence, lags of cloud cover and the residual were generated over short- and medium-term windows, along with moving averages summarizing the recent system dynamics. Additionally, Fourier harmonic terms based on the day of the year were incorporated to explicitly model the annual seasonality of solar radiation.

No dimensionality reduction techniques, such as principal component analysis (PCA), were applied, since one of the objectives of VEGA-RAD is to preserve the physical interpretability of atmospheric variables and their transformations. All variables were used on scales consistent with their physical meaning, ensuring transparent integration between the physical, stochastic, and statistical components of the model.

2.3. Methodology

This section presents the methodological formulation of the proposed hybrid model, VEGA-RAD (Vega Radiative Adaptive Dynamics), developed for daily solar radiation prediction in the Peruvian Amazon. The methodology is structured into five clearly defined sequential components: (i) the hybrid VEGA-RAD formulation, which integrates physical foundations and stochastic memory with an adaptive statistical correction stage; (ii) the physical–astronomical proxy block, responsible for modeling reference solar radiation based on celestial mechanics principles; (iii) the stochastic memory block, which captures the temporal dependence and annual seasonality of solar radiation; (iv) the adaptive machine learning–based statistical correction block, applied to the logarithmic residual of the physical–memory model, along with the model training, validation, and testing procedure, and the hyperparameter tuning strategy under a progressive temporal validation scheme; and (v) the conceptual model framework, which visually synthesizes the proposed architecture, the flow of information between components, and the generation of point predictions and conformal intervals.

As a preliminary step before modeling, the daily series were temporally aligned, cleaned of missing values, and transformed using lags, moving averages, and harmonic terms, with the aim of preserving temporal coherence and avoiding information leakage between training and testing sets.

For comparative purposes, two configurations of the proposed model were evaluated. The baseline configuration considers only the physical–astronomical proxy and simple temporal lags, without incorporating explicit harmonic terms or adaptive statistical correction. The optimized configuration corresponds to the full VEGA-RAD formulation, integrating stochastic memory, harmonic seasonality via Fourier series, and a machine learning–based statistical correction. Both configurations were evaluated under the same experimental protocol, enabling a direct and consistent comparison of their predictive performance in the Results section.

2.3.1. Hybrid VEGA-RAD Formulation

Solar radiation at point i and time t is defined as:

$$G_i(t) = \Phi_i(t) \exp\left(Z_i^C(t) + Z_i^X(t) + Z_i^H(t)\right) \quad (1)$$

Where $\Phi_i(t)$ is the reference astronomical flux, $Z_i^C(t)$ represents the cloud cover contribution, $Z_i^X(t)$ models the atmospheric composition (absorption, aerosols, and water vapor) and $Z_i^H(t)$ describes the local dynamics through stochastic memory.

2.3.2. Block 1: Physical–Astronomical Proxy

The reference component is calculated as:

$$\Phi_i(t) = I_0(t)\tau_a(t)\, cos\theta_z(t) \quad (2)$$

Where $\Phi_i(t) = I_{sc}\left(1 + 0.033cos\frac{2\pi d}{365}\right)$ represents the extraterrestrial radiation corrected for orbital eccentricity (with d as the Julian day), $\tau_a(t)$ is the large-scale atmospheric transmission coefficient and $\theta_z(t)$ is the solar zenith angle. This block constitutes the deterministic component of the model, governed by celestial mechanics and large-scale atmospheric attenuation.

2.3.3. Block 2: Stochastic Memory

To capture the temporal dependence and annual seasonality of solar radiation, it is defined as:

$$Z_i^H(t) = \sum_{k=1}^{p} \alpha_k G_i(t-k) + \sum_{m=1}^{M}\left[\beta_m sin\left(\frac{2\pi mt}{365}\right) + \gamma_m \cos\left(\frac{2\pi mt}{365}\right)\right] \quad (3)$$

Where p is the autoregressive order, α_k measures the influence of past solar radiation values, M is the number of Fourier harmonics used and capture intra- and (β_m, γ_m) interannual seasonal variation.

2.3.4. Block 3: Adaptive Statistical Correction

The logarithmic residual is defined as:

$$R_i(t) = lnG_i^{obs}(t) - ln\hat{G}_i^{proxy}(t) \quad (4)$$

Where $G_i^{obs}(t)$ is the observed solar radiation and $\hat{G}_i^{proxy}(t) = \Phi_i(t)\exp\left(Z_i^H(t)\right)$ corresponds to the estimate generated by the physical–memory block. This residual is modeled using a machine learning algorithm:

$$\hat{R}_i(t) = f_\theta(x_i(t)) \quad (5)$$

Where $x_i(t) = \{cloud(t), T_{2m}(t), RH(T), wind(t), R_i(t-k)\}$ is the vector of atmospheric predictors and temporal lags and f_θ corresponds to a *HistGradientBoosting regressor*.

2.3.5. Final Prediction

The estimated solar radiation is reconstructed as:

$$\hat{G}_i(t) = \hat{G}_i^{proxy}(t) \exp\left(\hat{R}_i(t)\right) \quad (6)$$

In general terms, the VEGA-RAD model combines physical foundations, stochastic memory, and adaptive statistical correction to produce robust solar radiation predictions in Amazonian contexts.

2.3.6. Model Training, Validation, and Testing

The experimental procedure for the VEGA-RAD model was designed according to the temporal nature of the data and to prevent information leakage. The complete set of daily observations was chronologically ordered and partitioned following a temporal validation scheme, in which training data always precedes validation and test data.

The machine learning–based statistical correction was applied to the logarithmic residual between the observed radiation and the physical–astronomical proxy. A Histogram-based Gradient Boosting Regressor

was used for model training, selected for its ability to capture complex nonlinear relationships while maintaining numerical stability and computational efficiency.

Model tuning was performed using time-series cross-validation with a forward-chaining partitioning scheme, ensuring a realistic evaluation of predictive performance.

The model's performance was assessed on an independent test set using standard regression metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). Additionally, for probabilistic evaluation, conformal intervals were constructed on the test set to quantify the uncertainty associated with predictions and assess model calibration from a frequentist perspective.

To prevent overfitting in the statistical correction stage, regularization mechanisms and early stopping were incorporated into the machine learning component.

2.3.6.1. Hyperparameter Tuning

Hyperparameter tuning was carried out systematically and in a controlled manner, taking into account the temporal nature of the data and the goal of avoiding overfitting. In the machine learning component, based on Histogram Gradient Boosting, hyperparameters related to model complexity and regularization were adjusted, including maximum tree depth, learning rate, number of boosting iterations, and regularization terms. Selection was performed using progressive temporal validation (forward chaining), seeking a balance between predictive capacity and model stability.

Additionally, the structural hyperparameters of the VEGA-RAD model, such as the autoregressive order of the stochastic memory, the number of Fourier harmonics, and the moving average window lengths, were determined through exploratory analysis and preliminary comparative evaluation on the training set. Once selected, all hyperparameters were kept fixed during the final evaluation on the independent test set, ensuring the validity of the experimental protocol and the reproducibility of results.

The final hyperparameter configuration used in the VEGA-RAD model is summarized in Table 2.

Table 2. Final Hyperparameter Configuration of the VEGA-RAD Model

Model Component	Hyperparameter	Value Used	Description	Selection Criterion
Stochastic Memory	Autoregressive order p (residual)	p = 4 (lags: 1, 2, 7, 14 days)	Temporal dependence of the logarithmic residual	Exploratory Evaluation during Training
Stochastic Memory	Cloud lags	1, 2, 7, 14 days	Cloud persistence	Residual Stability
Clarity Index (CMF)	CMF lags	1, 2, 7 days	Atmospheric state smoothing	Predictive Robustness
Seasonality	Fourier harmonics (M)	3	Capture of annual and intra-annual seasonality	Spectral Analysis
Temporal Smoothing	Moving average window	7 days	High-frequency noise reduction	Temporal Stability
HGBR	max_depth	8	Maximum tree depth	Complexity Control
HGBR	max_iter	2200	Number of boosting iterations	Error Convergence
HGBR	learning_rate	0.045	Learning rate	Bias–Variance Trade-off
HGBR	min_samples_leaf	15	Minimum samples per leaf	Structural Regularization
HGBR	L2 regularization	5×10^{-3}	L2 regularization	Overfitting Prevention
HGBR	early_stopping	Activated	Automatic early stopping	Numerical Stability
HGBR	validation_fraction	0.1	Internal validation fraction	Training Oversight
HGBR	n_iter_no_change	60	Early stopping patience	Robust Convergence

2.4. Model Diagram

This section presents the conceptual diagram of the VEGA-RAD model, aiming to visually summarize the model architecture, the flow of information between its components, and the processing sequence followed from the input meteorological data to the generation of point predictions and conformal intervals.

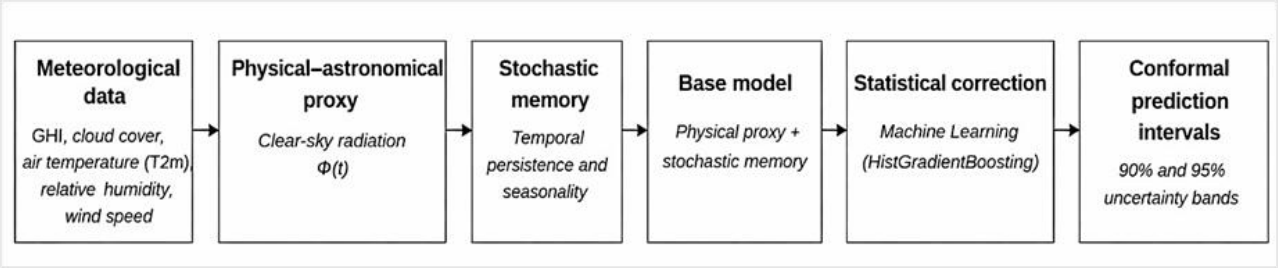

Figure 2. Conceptual diagram of the VEGA-RAD model, illustrating the interaction between the physical–astronomical proxy, stochastic memory, and statistical correction via machine learning, as well as the inclusion of conformal intervals for uncertainty quantification

The full details of data preprocessing, the training scheme, hyperparameter selection, and the construction of conformal intervals are described in Sections 2.3.6, 2.3.6.1, and 3.

3. RESULTS AND DISCUSSION

This section presents and analyzes the results obtained with the VEGA-RAD model, applied to the daily prediction of solar radiation in the city of Tarapoto for the period from January 1, 2017, to September 26, 2025. The analysis was conducted considering deterministic performance metrics as well as a probabilistic evaluation based on conformal intervals, with the aim of addressing the main objective of the study.

The results presented below correspond to the evaluation of the VEGA-RAD model under a temporal validation scheme and an independent test set, according to the experimental protocol described in Section 2.3.6.

3.1. Evaluated Configurations: Base Model and Optimized Model

To assess the impact of the different components of the VEGA-RAD model, two distinct experimental configurations were analyzed. The base configuration corresponds to a simplified formulation of the model, in which the solar radiation estimate is obtained from the physical–astronomical proxy and basic temporal lags, without incorporating advanced statistical correction mechanisms or explicit seasonal terms.

In contrast, the optimized configuration integrates the full formulation of the VEGA-RAD model, incorporating the stochastic memory component, Fourier harmonic terms to capture annual seasonality, and an adaptive machine-learning-based statistical correction applied to the logarithmic residual. Both configurations were evaluated under the same temporal validation scheme and independent test set, ensuring a fair and consistent comparison of their predictive performance.

3.2. Predictive Performance of the Model

Table 3 summarizes the performance of the VEGA-RAD model under the two evaluated configurations for Tarapoto during the period 2017–2025.

Table 3. Predictive Performance of the VEGA-RAD Model in Two Configurations

Configuration	MAE	RMSE	R^2
Base version	1.699	2.309	0.635
Optimized version	0.477	1.459	0.854

Note: MAE and RMSE are expressed in kWh/m^2/d. The base version uses a clear-sky proxy and meteorological variables with simple lags. The optimized version incorporates the Clarity Index (CMF), cloud and residual lags, moving averages, and Fourier harmonic terms. Values were calculated over 3155 valid days from 2017–2025.

The base version achieved a mean absolute error (MAE) of 1.699 kWh/m^2/d, a root mean square error (RMSE) of 2.309 kWh/m^2/d, and a coefficient of determination R^2 = 0.635. In contrast, the optimized version reduced the MAE to 0.477 kWh/m^2/d and the RMSE to 1.459 kWh/m^2/d, increasing the coefficient of determination to R^2 = 0.854.

3.3. Inferential analysis of predictive performance

To evaluate whether the observed differences between the base and optimized configurations of the VEGA-RAD model are statistically significant, an inferential analysis was conducted based on the paired comparison of daily errors on the independent test set.

Given the temporal nature of the series and the lack of normality assumptions in the error distribution, the non-parametric Wilcoxon signed-rank test was applied. The analysis used the daily absolute errors from both model configurations, considering coincident time points in the test set. The alternative hypothesis stated that the optimized configuration systematically presents lower error than the base configuration.

The test results showed a statistically significant difference in favor of the optimized model version ($p < 0.01$). Additionally, a bootstrap analysis of the mean absolute error difference confirmed an average reduction of ΔMAE = 1.23 kWh/m^2/d, with a 95% confidence interval [1.18; 1.27] kWh/m^2/d, which does not include zero. These results confirm that the improvements observed in the global metrics are not due to random fluctuations but correspond to a structural effect associated with the incorporation of stochastic memory, harmonic seasonality, and adaptive statistical correction via machine learning.

The results of the inferential analysis are summarized in Table 4, highlighting a statistically significant and robust improvement of the optimized VEGA-RAD model compared to the base version.

Table 4. Inferential analysis of the predictive performance of the VEGA-RAD model

Inferential analysis	Evaluated metric	Result	Interpretation
Paired Wilcoxon test	Daily absolute error (MAE)	$p < 0.001$	Statistically significant difference in favor of the optimized version
Bootstrap (B = 5000)	Δ Mean MAE	1.23 kWh/m^2/d	Average reduction in absolute error
Bootstrap (IC 95 %)	ΔMAE	[1.18, 1.27] kWh/m^2/d	Interval does not include 0; robust improvement

Note: The inferential analysis was conducted on paired daily errors from the independent test set (n = 3170), considering coincident time points for both model configurations.

3.4. Temporal Analysis of Predictions

Figure 3 presents the temporal comparison between solar radiation observations from ERA5 and the central prediction generated by VEGA-RAD, along with the 90% conformal interval. The model accurately reproduced both the annual seasonality and daily variability of solar radiation, maintaining consistency with the characteristic climatic patterns of the region.

During periods of high variability associated with increased cloudiness, the optimized version of the model showed greater stability and lower error dispersion, confirming the contribution of the stochastic memory and adaptive statistical correction components.

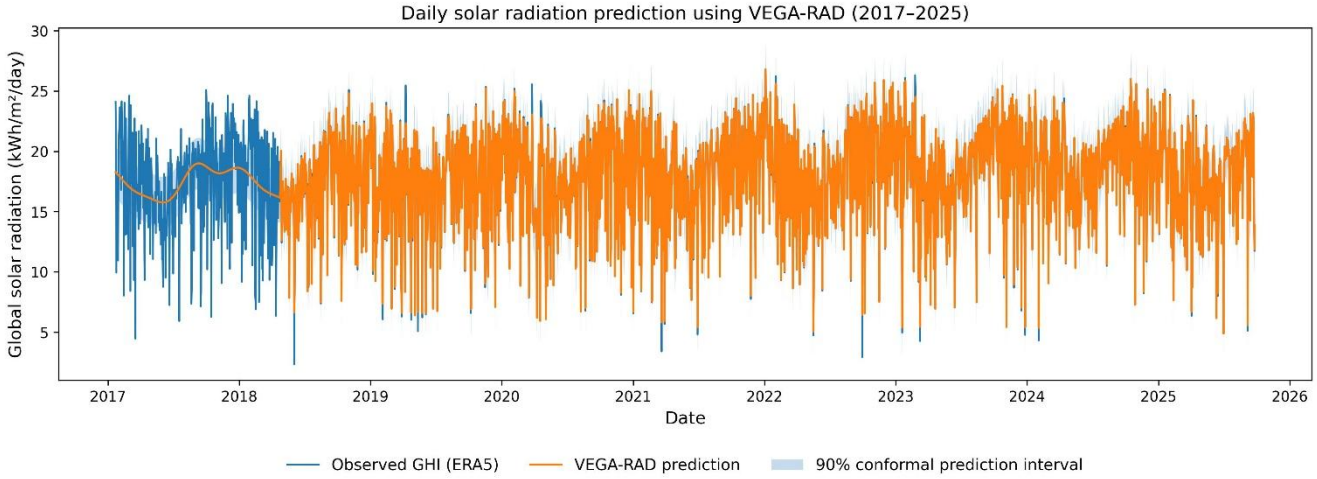

Figure 3. Daily Solar Radiation Prediction Using the VEGA-RAD Model for Tarapoto (2017–2025). The plot shows ERA5 observed values, the model's central prediction, and the 90% conformal prediction interval

To complement the visual analysis presented in Figure 3, the performance of the VEGA-RAD model was further evaluated by disaggregating it according to characteristic climatic periods of the Amazon region. Specifically, error metrics were analyzed for the dry and wet seasons using the independent test set. Results show that the mean absolute error (MAE) and root mean squared error (RMSE) are lower during the dry season (MAE = 0.323 kWh/m^2/d; RMSE = 1.163 kWh/m^2/d) and increase during the wet season (MAE = 0.660 kWh/m^2/d; RMSE = 1.757 kWh/m^2/d), consistent with higher cloudiness and atmospheric variability. Nevertheless, the model maintains predictive stability and a consistent advantage in both climatic regimes, demonstrating the robustness of the VEGA-RAD approach under contrasting atmospheric conditions. The disaggregated values for each climatic period are summarized in Table 5.

Table 5. Predictive performance of the VEGA-RAD model by climatic period (optimized version)

Climatic period	MAE (kWh/m^2/d)	RMSE (kWh/m^2/d)	n
Dry season	0.323	1.163	1621
Wet season	0.66	1.757	1549

Note: The analysis was performed on the independent test set, considering temporal partitions consistent with the validation scheme described in Section 2.3.6.

3.5. Probabilistic evaluation and uncertainty

The probabilistic evaluation was conducted through the analysis of the coverage of the conformal intervals and their overall behavior, while uncertainty was assessed in terms of calibration, degree of conservatism, and temporal stability.

In this context, the probabilistic evaluation using conformal intervals showed coverage consistent with the nominal 90 % and 95 % levels on the test set, as summarized in Table 6. In both cases, the coverage was equal to or greater than the nominal levels, indicating a conservatively calibrated but reliable performance of the conformal intervals. Consequently, VEGA-RAD not only provides accurate point estimates but also delivers prediction intervals that are reliable and well-calibrated from a frequentist perspective.

Table 6. Probabilistic evaluation of the conformal intervals of the VEGA-RAD model on the test set for Tarapoto (2017–2025)

Nominal level	Observed coverage	Mean interval width (kWh/m^2/d)	CV of width	Dataset
90%	1	4.23	≈ 0	Test
95%	1	4.23	≈ 0	Test

Although the observed coverages for both nominal levels reached unity in the test set, this result can be interpreted as indicative of a slightly conservative behavior of the conformal intervals. Such conservatism aligns with the high atmospheric variability characteristic of the Amazonian context and the non-stationary

nature of daily solar radiation. Nevertheless, this behavior is desirable from a risk management perspective, as it prioritizes reliable coverage over underestimation of uncertainty.

Additionally, the mean width of the conformal intervals remained stable throughout the test period, with a virtually null coefficient of variation, indicating a consistent quantification of uncertainty over time. This outcome reinforces the practical usefulness of the proposed approach for planning, designing, and operating photovoltaic systems in Amazonian contexts with high atmospheric variability.

3.6. Discussion of Findings

The results confirmed that integrating interpretable physical foundations with stochastic temporal memory and statistical correction via machine learning constitutes an effective strategy to address the high atmospheric variability characteristic of tropical regions. In particular, the reduction of over 70% in mean absolute error (MAE) between the base and optimized versions of VEGA-RAD highlighted the significant impact of incorporating memory terms, explicit seasonality, and adaptive learning, in line with recent trends reported in the literature on hybrid solar radiation prediction models and machine learning (Bakır, 2024; Demir, 2025; Yadav et al., 2025; Zerouali et al., 2025).

Recent studies have shown that machine learning–based approaches consistently outperform conventional statistical models by capturing complex nonlinearities in the atmosphere–surface system (Ghareeb et al., 2025; Tandon et al., 2025).

However, such works often rely on deep architectures with high complexity and limited physical interpretability. In contrast, the proposed model explicitly introduces an interpretable physical–astronomical proxy coupled with a stochastic memory structure, addressing a recurring limitation identified in recent state-of-the-art reviews (Shringi et al., 2025).

From a methodological perspective, the findings of this study are conceptually consistent with advances reported in hybrid CNN–SVR and CNN–LSTM approaches for daily solar radiation prediction (Ghimire et al., 2022; Hamdaouy et al., 2025; Şener & Tuğal, 2025). Nevertheless, unlike these approaches, VEGA-RAD does not rely exclusively on automatic latent feature extraction but combines, in a parsimonious manner, physical knowledge, autoregressive lags, and gradient boosting–based statistical correction, achieving a favorable balance between performance, interpretability, and computational cost, as suggested by recent studies on lightweight hybrid models (Arseven & Çınar, 2025; Solano & Affonso, 2023).

In the Amazonian context, the results align with studies using regional numerical models, such as WRF–Solar, applied to estimate global horizontal solar radiation in northern Brazil (Amorim et al., 2024; Krishnan & Ravi Kumar, 2025; Alves et al., 2025). These works also highlight persistent cloudiness and intra- and interannual variability as the main challenges for solar radiation prediction in the region. However, while numerical models depend on complex physical parameterizations and high computational cost, VEGA-RAD demonstrates that a hybrid approach based on ERA5 reanalysis data can efficiently capture such variability with lower structural complexity, consistent with recent studies based on satellite and reanalysis data (Wu et al., 2025).

A distinctive aspect of this work is the explicit incorporation of conformal intervals for predictive uncertainty quantification. The observed coverages, consistent with nominal levels of 90% and 95%, confirm the adequate probabilistic calibration of the model, extending the analysis beyond point metrics. This probabilistic approach responds to a growing need in energy planning and climate risk assessment, as discussed in recent studies on energy resilience and hybrid renewable systems (Cui et al., 2025; Jadhav & Bhawar, 2025; Vignesh Kumar et al., 2025).

Overall, the results suggest that VEGA-RAD represents a relevant and original methodological contribution in the specific context of the Peruvian Amazon, coherently integrating an interpretable physical–

astronomical proxy, stochastic memory, machine learning–based statistical correction, and explicit uncertainty quantification within a unified framework.

While individual components of this approach have been previously explored in the literature, their systematic integration under a temporal validation scheme, rigorous inferential analysis, and probabilistic evaluation using conformal intervals has not been jointly reported for Amazonian regions characterized by high atmospheric variability. In this sense, VEGA-RAD positions itself as a robust, interpretable, and reproducible alternative to purely physical or data-driven approaches, helping to fill methodological gaps identified in recent area reviews (Shringi et al., 2025; Zerouali et al., 2025).

CONCLUSIONS

In this study, a hybrid model named VEGA-RAD (Vega Radiative Adaptive Dynamics) was formulated and evaluated, designed for daily solar radiation prediction in the Peruvian Amazon. The model coherently integrates an interpretable physical–astronomical proxy, a stochastic temporal memory component, and an adaptive machine learning–based statistical correction stage. This formulation enabled the simultaneous capture of the physical structure of the radiative process, temporal persistence, and residual nonlinearities associated with local meteorological variables, while maintaining a parsimonious and reproducible architecture.

Predictive and inferential analyses demonstrated substantial improvements in the performance of the optimized version compared to the base configuration, with reductions of over 70% in mean absolute error and significant increases in the coefficient of determination. These improvements were supported by paired inferential analysis (Wilcoxon test) and nonparametric bootstrap resampling, confirming that the observed differences are not attributable to chance. Additionally, the inclusion of conformal intervals allowed for explicit quantification of predictive uncertainty, achieving coverages consistent with nominal levels and exhibiting conservative and stable behavior over time.

Overall, VEGA-RAD positions itself as a robust, interpretable, and reliable alternative for solar radiation prediction in tropical regions with high climatic variability, offering direct utility for the planning and management of photovoltaic systems.

FINANCING

The authors received no sponsorship to conduct this study-article.

CONFLICT OF INTEREST

There is no conflict of interest related to this work.

AUTHORSHIP CONTRIBUTION

Conceptualization, data Curation, formal analysis, research, methodology, project administration, resources, software, visualization, writing - original draft: Agreda-Vega, J. F. Supervision and writing - review and editing: Sare-Lara, E. and Rosales-Huamani, A. R. Validation: Agreda-Vega, J. F. and Sare-Lara, E.

AVAILABILITY OF DEPOSITED DATA

Data processing, implementation of the VEGA-RAD model, and experimental evaluation were carried out using the Python programming language in a reproducible computational environment based on Google Colab. The workflow included automated downloading of ERA5 reanalysis data via the Open-Meteo API, variable preprocessing, model training, and the generation of performance metrics and conformal prediction intervals.

The data used in this study are publicly available and can be reproduced through the Open-Meteo API by specifying the geographic coordinates, temporal period, and variables described in Section 2.2. The code used for model implementation is available from the corresponding author upon reasonable request.

REFERENCES

Abad-Alcaraz, V., Castilla, M., Carballo, J. A., Bonilla, J., & Álvarez, J. D. (2025). Multimodal deep learning for solar radiation forecasting. *Applied Energy*, *393*, 126061. https://doi.org/10.1016/j.apenergy.2025.126061

Aladwani, S. M., Almutairi, A., Alolayan, M. A., Abdullah, H., & Abraham, L. M. (2025). Prediction of solar radiation as a function of particulate matter pollution and meteorological data using machine learning models. *Journal of Engineering Research*, *13*(4), 2818–2825. https://doi.org/10.1016/j.jer.2024.11.005

Alves, P. V., Bourscheidt, V., Fabrício dos Santos, L. O., & Humbelino de Melo, P. R. (2025). Seasonal variations and trends in solar UV spectral irradiances based on data from the Ozone Monitoring Instrument at solar noon in Southern Amazonas, Brazil. *Remote Sensing Applications: Society and Environment*, *37*, 101423. https://doi.org/10.1016/j.rsase.2024.101423

Amorim, A. C. B., de Almeida Dantas, V., dos Reis, J. S., de Assis Bose, N., Emiliavaca, S. de A. S., Cruz Bezerra, L. A., de Matos, M. de F. A., de Mello Nobre, M. T. C., Oliveira, L. de L., & de Medeiros, A. M. (2024). Analysis of WRF-solar in the estimation of global horizontal irradiation in Amapá, northern Brazil. *Renewable Energy*, *235*, 121361. https://doi.org/10.1016/j.renene.2024.121361

Arseven, B., & Çınar, S. M. (2025). A novel hybrid solar radiation forecasting algorithm based on discrete wavelet transform and multivariate machine learning models integrated with clearness index clusters. *Journal of Atmospheric and Solar-Terrestrial Physics*, *267*, 106417. https://doi.org/10.1016/j.jastp.2025.106417

Bakır, H. (2024). Prediction of daily global solar radiation in different climatic conditions using metaheuristic search algorithms: a case study from Türkiye. *Environmental Science and Pollution Research*, *31*(30), 43211–43237. https://doi.org/10.1007/s11356-024-33785-x

Celik, A. N., Sarman, B., & Polat, K. (2025). Horizontal-to-tilted conversion of solar radiation data using machine learning algorithms. *Engineering Applications of Artificial Intelligence*, *153*, 110951. https://doi.org/10.1016/j.engappai.2025.110951

Cui, X., Yin, S., Chen, H., & Niu, D. (2025). A temporal–image parallel hybrid solar radiation–wind speed–green hydrogen production potential prediction model based on federated learning and rolling real-time decomposition. *Energy*, *337*, 138516. https://doi.org/10.1016/j.energy.2025.138516

Demir, V. (2025). Evaluation of Solar Radiation Prediction Models Using AI: A Performance Comparison in the High-Potential Region of Konya, Türkiye. *Atmosphere*, *16*(4), 398. https://doi.org/10.3390/atmos16040398

Gao, Y., Hu, Z., Chen, W.-A., Liu, M., & Ruan, Y. (2025). A revolutionary neural network architecture with interpretability and flexibility based on Kolmogorov–Arnold for solar radiation and temperature forecasting. *Applied Energy*, *378*, 124844. https://doi.org/10.1016/j.apenergy.2024.124844

Ghareeb, N., Alanazi, A., Sedaghat, A., Farhat, M. H., Mehdizadeh, A., Salem, H., Nazififard, M., & Mostafaeipour, A. (2025). Integrating experimental and theoretical approaches for enhanced machine learning modeling of solar radiation. *Engineering Science and Technology, an International Journal*, *70*, 102156. https://doi.org/10.1016/j.jestch.2025.102156

Ghimire, S., Bhandari, B., Casillas-Pérez, D., Deo, R. C., & Salcedo-Sanz, S. (2022). Hybrid deep CNN-SVR

algorithm for solar radiation prediction problems in Queensland, Australia. *Engineering Applications of Artificial Intelligence*, *112*, 104860. https://doi.org/10.1016/j.engappai.2022.104860

Hamdaouy, H., Benghoulam, E. M., Chaibi, M., Berrada, M., & Hmaidi, A. El. (2025). Estimating daily global solar radiation using deep learning. *Results in Engineering*, *27*, 106132. https://doi.org/10.1016/j.rineng.2025.106132

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hou, X., Fountoulakis, I., Blanc, P., Aebi, C., & Kazadzis, S. (2025). Intrahour solar radiation forecasting based on sun visibility for different cloud types. *Solar Energy*, *294*, 113477. https://doi.org/10.1016/j.solener.2025.113477

Hu, Z., Wan, Z., Wang, Z., Zhang, H., Liu, S., Fan, X., & Zheng, W. (2025). Machine learning modeling of indoor thermal sensation under solar radiation considering skin temperatures. *Building and Environment*, *275*, 112822. https://doi.org/10.1016/j.buildenv.2025.112822

Huang, L., Kang, J., Wan, M., Fang, L., Zhang, C., & Zeng, Z. (2021). Solar Radiation Prediction Using Different Machine Learning Algorithms and Implications for Extreme Climate Events. *Frontiers in Earth Science*, *9*. https://doi.org/10.3389/feart.2021.596860

Jadhav, A. V., & Bhawar, R. L. (2025). Future changes in surface solar radiation over India: A bias-corrected and downscaled assessment of CMIP6 projections for renewable energy planning. *Energy and Climate Change*, *6*, 100213. https://doi.org/10.1016/j.egycc.2025.100213

Krishnan, N., & Ravi Kumar, K. (2025). Impact of shortwave radiation parameterization schemes in predicting global horizontal irradiance for various climatic zones by WRF-Solar: A case study in India. *Journal of Atmospheric and Solar-Terrestrial Physics*, *274*, 106590. https://doi.org/10.1016/j.jastp.2025.106590

Open-Meteo. (2025). *Open-Meteo Historical Weather API (ERA5 Reanalysis Data)*. https://open-meteo.com/en/docs/historicalweather-api

Rajput, J., Kushwaha, N. L., Srivastava, A., Vishwakarma, D. K., Mishra, A. K., Sahoo, P. K., Suna, T., Rana, L., Jatav, M. S., Kumar, J., Dimple, Shaloo, Bisht, H., Rai, A., Zerouali, B., Pande, C. B., & Elbeltagi, A. (2025). Developing machine learning models for predicting daily relative humidity and solar radiation using lagged time series data inputs in a semi-arid climate. *Journal of Atmospheric and Solar-Terrestrial Physics*, *276*, 106619. https://doi.org/10.1016/j.jastp.2025.106619

Raju, V. A. G., Nayak, J., Dash, P. B., & Mishra, M. (2025). Short-term solar irradiance forecasting model based on hyper-parameter tuned LSTM via chaotic particle swarm optimization algorithm. *Case Studies in Thermal Engineering*, *69*, 105999. https://doi.org/10.1016/j.csite.2025.105999

Ruan, G., Chen, X., Li, Y., Lim, E. G., Fang, L., Jiang, L., Du, Y., & Wang, F. (2026). SkyNet: A Deep Learning Architecture for Intra-hour Multimodal Solar Forecasting with Ground-based Sky Images. *Renewable Energy*, *256*, 124354. https://doi.org/10.1016/j.renene.2025.124354

Şener, İ. F., & Tuğal, İ. (2025). Optimized CNN-LSTM with hybrid metaheuristic approaches for solar radiation forecasting. *Case Studies in Thermal Engineering*, *72*, 106356. https://doi.org/10.1016/j.csite.2025.106356

Shringi, S., Saini, L. M., & Aggarwal, S. K. (2025). A review of data-driven deep learning models for solar and wind energy forecasting. *Renewable Energy Focus*, *55*, 100739.

https://doi.org/10.1016/j.ref.2025.100739

Solano, E. S., & Affonso, C. M. (2023). Solar Irradiation Forecasting Using Ensemble Voting Based on Machine Learning Algorithms. *Sustainability*, *15*(10), 7943. https://doi.org/10.3390/su15107943

Tandon, A., Awasthi, A., Pattnayak, K. C., Tandon, A., Choudhury, T., & Kotecha, K. (2025). Machine learning-driven solar irradiance prediction: advancing renewable energy in Rajasthan. *Discover Applied Sciences*, *7*(2), 107. https://doi.org/10.1007/s42452-025-06490-8

Vignesh Kumar, V., Madhesh, K., Sanjay, K., Guru Prasath, P., Pavish Karthik, A., & Praveen Kumar, G. (2025). A novel ensemble machine learning approach for optimizing sustainability and green hydrogen production in hybrid renewable-based organic Rankine cycle-operated proton exchange membrane electrolyser system. *Renewable Energy*, *242*, 122369. https://doi.org/10.1016/j.renene.2025.122369

Wu, H., Zhang, C., Xue, J., Niu, X., Zhao, B., Pei, G., & Liu, C. (2025). Machine learning forecasts of short wave radiation from geostationary satellite measurements to optimize solar photovoltaic and concentrated solar power systems. *Solar Energy*, *299*, 113718. https://doi.org/10.1016/j.solener.2025.113718

Y.H., H., S.Y., T., & J.Q., G. (2024). Machine Learning-Based Solar Irradiance Forecasting Model Using GPS. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *16*(4), 31–36. https://doi.org/10.54554/jtec.2024.16.04.005

Yadav, A. K., Kumar, R., Wang, M., Fekete, G., & Singh, T. (2025). Comparative analysis of daily global solar radiation prediction using deep learning models inputted with stochastic variables. *Scientific Reports*, *15*(1), 10786. https://doi.org/10.1038/s41598-025-95281-7

Yan, Z., Lu, X., Wu, L., Zhang, H., Liu, F., Wang, X., Xu, W., & Liu, W. (2025). Enhancing short-term solar radiation forecasting with hybrid VMD and GraphCast-based machine learning models. *Expert Systems with Applications*, *285*, 128042. https://doi.org/10.1016/j.eswa.2025.128042

Zerouali, B., Bailek, N., Qaysi, S., Difi, S., Alarifi, N., Elbeltagi, A., Santos, C. A. G., He, K., & Youssef, Y. M. (2025). Hybrid machine learning optimization for solar radiation forecasting. *Physics and Chemistry of the Earth, Parts A/B/C*, *140*, 104052. https://doi.org/10.1016/j.pce.2025.104052

Zhu, L., Huang, X., Zhang, Z., Li, C., & Tai, Y. (2025). A novel U-LSTM-AFT model for hourly solar irradiance forecasting. *Renewable Energy*, *238*, 121955. https://doi.org/10.1016/j.renene.2024.121955

APPENDIX

Appendix A.1. Example of an Open-Meteo API (ERA5) query

In order to ensure the reproducibility of the study, an example of the HTTP request used to download daily ERA5 reanalysis data via the Open-Meteo API is presented below, corresponding to the city of Tarapoto (Peru) and the period from January 1, 2017 to September 26, 2025.

The query example is provided for illustrative purposes only, with the aim of documenting access to the data source used. The complete computational implementation, including variable preprocessing, model training, and experimental evaluation, was carried out in Python within a reproducible Google Colab–based environment and is available upon reasonable request to the corresponding author.

https://archive-api.open-meteo.com/v1/era5?latitude=-6.5&longitude=-76.3&start_date=2017-01-01&end_date=2025-09-26&daily=shortwave_radiation_sum,cloudcover_mean,temperature_2m_mean,relative_humidity_2m_mean,windspeed_10m_max&timezone=UTC

Note: For execution in Python, this query was implemented using the requests library, as shown in the reproducible code used in this study.

```
import requests
url = "https://archive-api.open-meteo.com/v1/era5"
params = {
  "latitude": -6.5,
  "longitude": -76.3,
  "start_date": "2017-01-01",
  "end_date": "2025-09-26",
  "daily": [
    "shortwave_radiation_sum",
    "cloudcover_mean",
    "temperature_2m_mean",
    "relative_humidity_2m_mean",
    "windspeed_10m_max"
  ],
  "timezone": "UTC"
}
r = requests.get(url, params=params)
data = r.json()
```