



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

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

Agreda-Vega, Juan Francisco<sup>1\*</sup>

Sare-Lara, Evergisto<sup>1</sup>

Rosales-Huamani, Jimmy Aurelio<sup>2,3</sup>

<sup>1</sup>Statistics and Applied Machine Learning Group, Universidad Nacional de San Martín, Tarapoto, Perú

<sup>2</sup>Department of Computer Science, Higher Polytechnic School, Universidad de Alcalá, Madrid, España

<sup>3</sup>Multidisciplinary Sensing, Universal Accessibility and Machine Learning Group, Universidad Nacional de Ingeniería, Lima, Perú

**Recibido:** 05 Oct. 2025 | **Aceptado:** 27 Dic. 2025 | **Publicado:** 20 Ene. 2026

**Autor de correspondencia\*:** [juan.agreda@unsm.edu.pe](mailto:juan.agreda@unsm.edu.pe)

**Cómo citar este artículo:** Agreda-Vega, J. F., Sare-Lara, E. & Rosales-Huamani, J. A. (2026). VEGA-RAD: Modelo híbrido físico-estadístico para la predicción diaria de radiación solar en la Amazonía. *Revista Científica de Sistemas e Informática*, 6(1), e1454.

<https://doi.org/10.51252/rcsi.v6i1.1454>

### 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<sup>2</sup>/d y un aumento del R<sup>2</sup> 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

### 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<sup>2</sup>/d and an increase in the coefficient of determination (R<sup>2</sup>) 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



## 1. INTRODUCCIÓN

La predicción de la radiación solar constituye un problema central para la integración eficiente de energías renovables en los sistemas eléctricos modernos. La elevada variabilidad espacio-temporal de la radiación solar condiciona el dimensionamiento de plantas fotovoltaicas, la estabilidad de las redes eléctricas y la formulación de políticas energéticas sostenibles, particularmente en regiones de alta complejidad climática (Bakır, 2024; Demir, 2025; Shringi et al., 2025; Tandon et al., 2025; Yadav et al., 2025; Zerouali et al., 2025).

En regiones tropicales como la Amazonía peruana, esta problemática se intensifica debido a la persistente nubosidad, la estacionalidad marcada y la escasez de estaciones meteorológicas confiables. En este contexto, los productos de reanálisis global se consolidan como una alternativa robusta para caracterizar la radiación solar y las variables atmosféricas asociadas (Hersbach et al., 2020; Huang et al., 2021).

Entre estas fuentes, ERA5 destaca por su consistencia temporal, cobertura espacial global y validación extensiva en múltiples zonas climáticas. Su acceso abierto mediante interfaces de programación facilita la reproducibilidad y la transparencia metodológica en estudios de predicción solar (Demir, 2025; Open-Meteo, 2025).

De forma paralela, los avances en aprendizaje automático y aprendizaje profundo han impulsado modelos capaces de capturar relaciones no lineales complejas entre la radiación solar y los factores atmosféricos. En particular, enfoques híbridos basados en CNN-SVR, CNN-LSTM y optimización metaheurística muestran mejoras sustanciales en la predicción diaria de radiación solar en distintos climas (Ghimire et al., 2022; Hamdaouy et al., 2025; Y.H. et al., 2024; Raju et al., 2025; Şener & Tuğal, 2025).

Estudios de revisión recientes confirman que los modelos híbridos dominan el estado del arte al integrar conocimiento físico, optimización algorítmica y aprendizaje adaptativo, superando consistentemente a enfoques puramente estadísticos o físicos (Celik et al., 2025; Ghareeb et al., 2025; Rajput et al., 2025; Shringi et al., 2025).

Para horizontes intra-hora e intradía, arquitecturas multimodales que combinan imágenes del cielo con variables meteorológicas permiten anticipar cambios nubosos rápidos con alta precisión. Propuestas como SkyNet y otros modelos multimodales demuestran una elevada robustez frente a la variabilidad atmosférica (Abad-Alcaraz et al., 2025; Hou et al., 2025; Ruan et al., 2026).

De manera complementaria, técnicas de descomposición multiescala, transformadas de onda y métodos de ensamble refuerzan la estabilidad predictiva. Asimismo, el uso de potencia fotovoltaica, productos satelitales y variables exógenas mejora la representación de la dinámica radiativa (Arseven & Çınar, 2025; Cui et al., 2025; Solano & Affonso, 2023; Wu et al., 2025; Yan et al., 2025).

Factores ambientales adicionales, como la calidad del aire, emergen como predictores relevantes de la atenuación de la radiación solar. Estudios recientes evidencian que la inclusión de contaminantes atmosféricos mejora significativamente la precisión de los modelos basados en aprendizaje automático (Aladwani et al., 2025; Hu et al., 2025).

En la planificación energética de mediano y largo plazo, evaluaciones basadas en proyecciones CMIP6 con corrección de sesgo y downscaling aportan evidencia regional sobre cambios futuros de la radiación solar. En este contexto, modelos numéricos regionales como WRF-Solar y estudios de radiación UV confirman la alta sensibilidad climática de regiones tropicales (Amorim et al., 2024; Chen & Gao, 2022; Jadhav & Bhawar, 2025; Krishnan & Ravi Kumar, 2025; Vignesh Kumar et al., 2025; Alves et al., 2025; Zhu et al., 2025).

A pesar de estos avances, persiste un vacío científico en el desarrollo de modelos formulados específicamente para la Amazonía peruana que integren fundamentos físicos interpretables, memoria temporal estocástica, corrección estadística adaptativa y cuantificación explícita de la incertidumbre. En respuesta a este vacío, el presente trabajo propone y formula un modelo híbrido original denominado

VEGA-RAD (Vega Radiative Adaptive Dynamics), orientado a la predicción diaria de la radiación solar mediante datos ERA5 y estimaciones puntuales y probabilísticas bien calibradas.

2. MATERIALES Y MÉTODOS

2.1. Área de estudio

El área de estudio corresponde a la ciudad de Tarapoto, ubicada en la Amazonía peruana (latitud  $-6,5^{\circ}$  , longitud  $-76,3^{\circ}$  ). La región presenta un clima tropical húmedo, caracterizado por elevada nubosidad persistente, alta humedad relativa y una marcada variabilidad intra e interanual de la radiación solar. Estas condiciones convierten a Tarapoto en un entorno particularmente desafiante para la predicción de radiación solar y la planificación de sistemas fotovoltaicos.

2.2. Datos

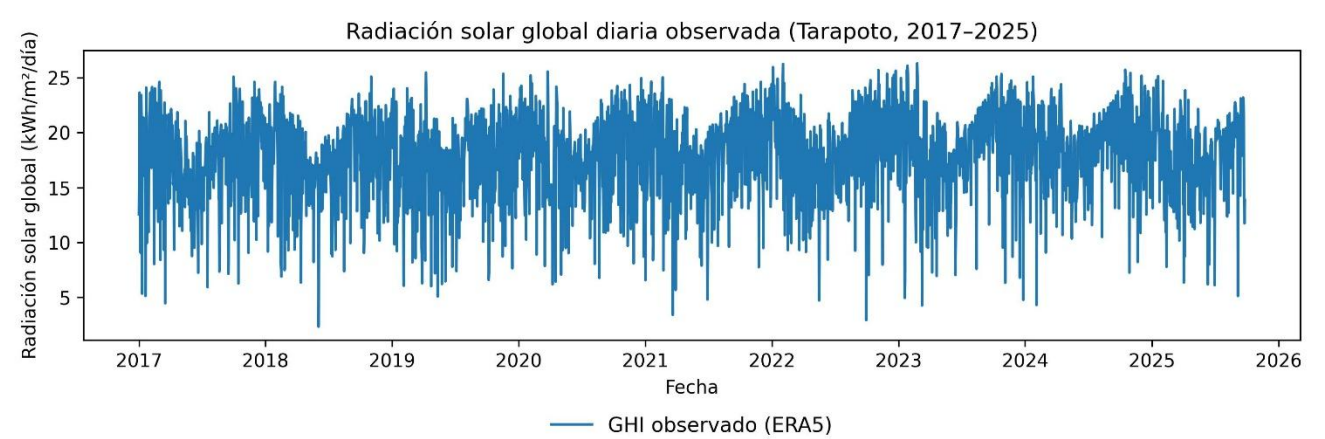
Los datos diarios del reanálisis ERA5 fueron obtenidos mediante la API de Open-Meteo, una plataforma de acceso abierto que proporciona productos de reanálisis climático a través de servicios web tipo REST. En particular, se utilizó el endpoint de acceso programático a ERA5 disponible en: <https://archive-api.open-meteo.com/v1/era5>.

El acceso a los datos se realizó mediante solicitudes HTTP parametrizadas, especificando las coordenadas geográficas, el periodo temporal y las variables meteorológicas de interés. La documentación oficial de la API y ejemplos de consulta se encuentran disponibles en el portal de Open-Meteo (<https://open-meteo.com/>). La descarga de los datos se efectuó el 26 de septiembre de 2025.

**Tabla 1.** Muestra de registros diarios del reanálisis ERA5 para Tarapoto (01.01.2017–26.09.2025). Unidades: GHI en kWh/m²/d, nubosidad en %, temperatura en °C, humedad relativa en % y viento en m/s

|   | Fecha   | 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  |

La Tabla 1 presenta una muestra representativa de los registros diarios del reanálisis ERA5, evidenciando la variabilidad conjunta entre la radiación solar global y las variables meteorológicas asociadas.



**Figura 1.** Serie temporal de la radiación solar global horizontal observada (ERA5) para Tarapoto (2017–2025). Se aprecia la estacionalidad anual y la elevada variabilidad diaria asociada a la nubosidad y a las condiciones atmosféricas

La Figura 1 muestra la evolución continua de la GHI diaria, revelando patrones estacionales e interanuales característicos del clima amazónico, lo que justifica el uso de un enfoque híbrido con memoria temporal.

### 2.2.1. Preprocesamiento de datos

Previo a la formulación y entrenamiento del modelo VEGA-RAD, los datos diarios del reanálisis ERA5 fueron sometidos a un proceso de preprocesamiento orientado a garantizar coherencia temporal, estabilidad numérica y reproducibilidad del análisis. En primer lugar, los registros descargados fueron organizados cronológicamente y se descartaron los días con valores faltantes en alguna de las variables consideradas, de modo que el conjunto final de datos estuvo compuesto exclusivamente por observaciones completas.

Dado que los productos ERA5 incorporan procedimientos de control de calidad a escala global, no se aplicaron métodos adicionales de detección o eliminación de valores atípicos, con el fin de preservar extremos físicamente plausibles asociados a eventos atmosféricos intensos característicos del clima amazónico. Asimismo, no se realizó imputación de valores faltantes, ya que el análisis se restringió a días con información completa.

Como parte de la agregación temporal, se trabajó directamente con series diarias, coherentes con el horizonte de predicción del estudio. A partir de estas series se construyó un proxy físico-astronómico de radiación solar de referencia, utilizado como componente determinista del modelo, y se definió un residuo logarítmico entre la radiación solar observada y dicha referencia, el cual constituye la variable objetivo del bloque de aprendizaje automático.

Para capturar la dependencia temporal y la persistencia atmosférica, se generaron retardos (*lags*) de la nubosidad y del residuo en ventanas de corto y mediano plazo, así como medias móviles que resumen la dinámica reciente del sistema. Adicionalmente, se incorporaron términos armónicos de Fourier basados en el día del año, con el propósito de modelar explícitamente la estacionalidad anual de la radiación solar.

No se aplicaron técnicas de reducción de dimensionalidad, como análisis de componentes principales (PCA), ya que uno de los objetivos del modelo VEGA-RAD es preservar la interpretabilidad física de las variables atmosféricas y de sus transformaciones. Todas las variables fueron utilizadas en escalas coherentes con su significado físico, asegurando una integración transparente entre los componentes físicos, estocásticos y estadísticos del modelo.

### 2.3. Metodología

En esta sección se presenta la formulación metodológica del modelo híbrido propuesto, denominado VEGA-RAD (Vega Radiative Adaptive Dynamics), desarrollado para la predicción diaria de la radiación solar en la Amazonía peruana. La metodología se estructura en cinco componentes secuenciales claramente definidos: (i) la formulación híbrida VEGA-RAD, que integra fundamentos físicos y memoria estocástica con una etapa de corrección estadística adaptativa; (ii) el bloque de proxy físico-astronómico, encargado de modelar la radiación solar de referencia a partir de principios de la mecánica celeste; (iii) el bloque de memoria estocástica, que captura la dependencia temporal y la estacionalidad anual de la radiación solar; (iv) el bloque de corrección estadística adaptativa basada en aprendizaje automático, aplicado sobre el residuo logarítmico del modelo físico-memoria, junto con el procedimiento de entrenamiento, validación y prueba del modelo y la estrategia de ajuste de hiperparámetros bajo un esquema de validación temporal progresiva; y (v) el esquema conceptual del modelo, que sintetiza visualmente la arquitectura propuesta, el flujo de información entre sus componentes y la generación de predicciones puntuales e intervalos conformales.

Como etapa previa al modelado, las series diarias fueron alineadas temporalmente, depuradas de valores faltantes y transformadas mediante retardos, medias móviles y términos armónicos, con el objetivo de preservar la coherencia temporal del proceso y evitar fugas de información entre los conjuntos de entrenamiento y prueba.

Con fines comparativos, se evaluaron dos configuraciones del modelo propuesto. La configuración base considera únicamente el proxy físico-astronómico y retardos temporales simples, sin incorporar términos

armónicos explícitos ni corrección estadística adaptativa. La configuración optimizada corresponde a la formulación completa del modelo VEGA-RAD, integrando memoria estocástica, estacionalidad armónica mediante series de Fourier y una corrección estadística basada en aprendizaje automático. Ambas configuraciones fueron evaluadas bajo un mismo protocolo experimental, permitiendo una comparación directa y consistente de su desempeño predictivo en la sección de Resultados.

### 2.3.1. Formulación híbrida VEGA-RAD

La radiación solar en el punto  $i$  y tiempo  $t$  se define como:

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

Donde  $\Phi_i(t)$  es el flujo astronómico de referencia,  $Z_i^C(t)$  representa la contribución de nubosidad,  $Z_i^X(t)$  modela la composición atmosférica (absorción, aerosoles y vapor de agua), y  $Z_i^H(t)$  describe la dinámica local mediante memoria estocástica.

### 2.3.2. Bloque 1: Proxy físico-astronómico

El componente de referencia se calcula como:

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

Donde  $\Phi_i(t) = I_{sc} \left( 1 + 0.033 \cos \frac{2\pi d}{365} \right)$  representa la radiación extraterrestre corregida por la excentricidad orbital ( $d$  es el día Juliano),  $\tau_a(t)$  es el coeficiente de transmisión atmosférica a gran escala y  $\theta_z(t)$  es el ángulo cenital solar. Este bloque constituye la componente determinista del modelo, gobernada por la mecánica celeste y la atenuación atmosférica de gran escala.

### 2.3.3. Bloque 2: Memoria estocástica

Para capturar la dependencia temporal y la estacionalidad anual de la radiación solar, se define:

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

Donde  $p$  es el orden autorregresivo,  $\alpha_k$  mide la influencia de valores pasados de radiación solar,  $M$  es el número de armónicos de Fourier utilizados, y  $(\beta_m, \gamma_m)$  capturan la variación estacional intra e interanual.

### 2.3.4. Bloque 3: Corrección estadística adaptativa

Se define el residuo logarítmico como:

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

Donde  $G_i^{obs}(t)$  es la radiación solar observada y  $\hat{G}_i^{proxy}(t) = \Phi_i(t) \exp(Z_i^H(t))$  corresponde a la estimación generada por el bloque físico-memoria. Este residuo se modela mediante un algoritmo de aprendizaje automático:

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

Donde  $x_i(t) = \{cloud(t), T_{2m}(t), RH(T), wind(t), R_i(t-k)\}$  es el vector de predictores atmosféricos y retardos temporales, y  $f_{\theta}$  corresponde a un regresor *HistGradientBoosting*.

### 2.3.5. Predicción final

La radiación solar estimada se reconstruye como:

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

En términos generales, el modelo VEGA-RAD combina fundamentos físicos, memoria estocástica y corrección estadística adaptativa para producir predicciones robustas de radiación solar en contextos amazónicos.

### 2.3.6. Entrenamiento, validación y prueba del modelo

El procedimiento experimental del modelo VEGA-RAD fue diseñado de acuerdo con la naturaleza temporal de los datos y con el objetivo de evitar fugas de información (data leakage). El conjunto completo de observaciones diarias fue ordenado cronológicamente y particionado siguiendo un esquema de validación temporal, en el cual los datos de entrenamiento corresponden siempre a instantes anteriores a los datos de validación y prueba.

La etapa de corrección estadística basada en aprendizaje automático se aplicó sobre el residuo logarítmico entre la radiación solar observada y el proxy físico-astronómico. Para el entrenamiento del modelo se empleó un algoritmo de Histogram-based Gradient Boosting Regressor, seleccionado por su capacidad para modelar relaciones no lineales complejas manteniendo estabilidad numérica y eficiencia computacional.

El ajuste del modelo se realizó mediante validación cruzada temporal (time-series cross-validation), utilizando un esquema de partición progresiva (forward chaining), garantizando una evaluación realista del desempeño predictivo.

El desempeño del modelo fue evaluado en un conjunto de prueba independiente, utilizando métricas estándar de regresión, incluyendo el error absoluto medio (MAE), el error cuadrático medio (RMSE) y el coeficiente de determinación ( $R^2$ ). Adicionalmente, para la evaluación probabilística, se construyeron intervalos conformales sobre el conjunto de prueba, permitiendo cuantificar la incertidumbre asociada a las predicciones y evaluar la calibración del modelo desde una perspectiva frecuentista.

Para evitar sobreajuste en la etapa de corrección estadística, se incorporaron mecanismos de regularización y detención temprana en el componente de aprendizaje automático.

#### 2.3.6.1. Ajuste de hiperparámetros

El ajuste de hiperparámetros se realizó de manera sistemática y controlada, considerando la naturaleza temporal de los datos y el objetivo de evitar sobreajuste. En el componente de aprendizaje automático, basado en Histogram Gradient Boosting, se ajustaron los hiperparámetros asociados a la complejidad del modelo y la regularización, incluyendo la profundidad máxima de los árboles, la tasa de aprendizaje, el número de iteraciones de boosting y los términos de regularización. La selección se efectuó mediante validación temporal progresiva (forward chaining), buscando un compromiso entre capacidad predictiva y estabilidad del modelo. Adicionalmente, los hiperparámetros estructurales del modelo VEGA-RAD, tales como el orden autorregresivo de la memoria estocástica, el número de armónicos de Fourier y las ventanas de medias móviles, se determinaron mediante análisis exploratorio y evaluación comparativa preliminar sobre el conjunto de entrenamiento. Una vez seleccionados, todos los hiperparámetros se mantuvieron fijos durante la evaluación final sobre el conjunto de prueba independiente, garantizando la validez del protocolo experimental y la reproducibilidad de los resultados.

La configuración final de los hiperparámetros utilizados en el modelo VEGA-RAD se resume en la Tabla 2.

**Tabla 2.** Configuración final de hiperparámetros del modelo VEGA-RAD

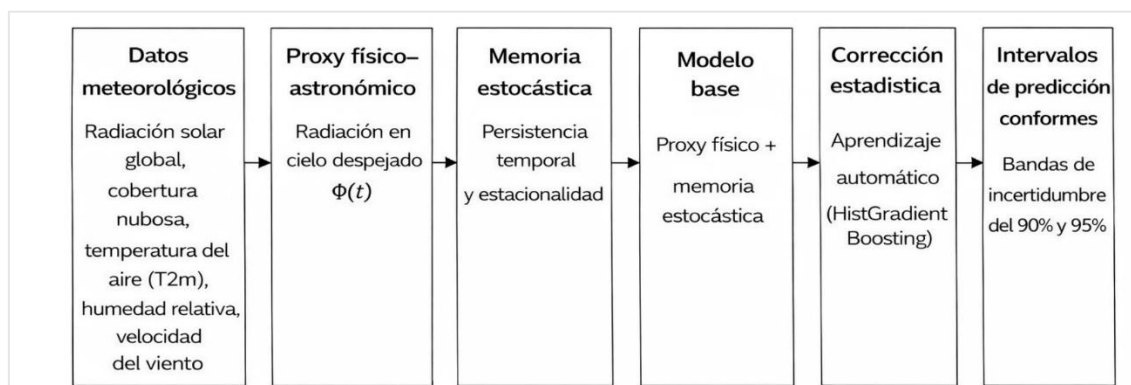
| Componente del modelo | Hiperparámetro                     | Valor utilizado                  | Descripción                                  | Criterio de selección                    |
|-----------------------|------------------------------------|----------------------------------|--|--|
| Memoria estocástica   | Orden autorregresivo $p$ (residuo) | $p = 4$ (lags: 1, 2, 7, 14 días) | Dependencia temporal del residuo logarítmico | Evaluación exploratoria en entrenamiento |
| Memoria estocástica   | Lags de nubosidad                  | 1, 2, 7, 14 días                 | Persistencia de nubosidad                    | Estabilidad del residuo                  |



|                          |                          |                    |   |                            |
|--------------------------|--------------------------|--------------------|---|----------------------------|
| Índice de claridad (CMF) | Lags CMF                 | 1, 2, 7 días       | Suavizado del estado atmosférico              | Robustez predictiva        |
| Estacionalidad           | Armónicos de Fourier (M) | 3                  | Captura de estacionalidad anual e intra-anual | Análisis espectral         |
| Suavizado temporal       | Ventana de media móvil   | 7 días             | Reducción de ruido de alta frecuencia         | Estabilidad temporal       |
| HGBR                     | max_depth                | 8                  | Profundidad máxima del árbol                  | Control de complejidad     |
| HGBR                     | max_iter                 | 2200               | Número de iteraciones de boosting             | Convergencia del error     |
| HGBR                     | learning_rate            | 0,045              | Tasa de aprendizaje                           | Compromiso sesgo-varianza  |
| HGBR                     | min_samples_leaf         | 15                 | Mínimo de muestras por hoja                   | Regularización estructural |
| HGBR                     | l2_regularization        | $5 \times 10^{-3}$ | Penalización L2                               | Prevención de sobreajuste  |
| HGBR                     | early_stopping           | Activado           | Detención temprana automática                 | Estabilidad numérica       |
| HGBR                     | validation_fraction      | 0,1                | Fracción interna de validación                | Control del entrenamiento  |
| HGBR                     | n_iter_no_change         | 60                 | Paciencia del early stopping                  | Convergencia robusta       |

## 2.4. Esquema del modelo

En esta sección se presenta el esquema conceptual del modelo VEGA-RAD, con el objetivo de sintetizar de manera visual la arquitectura del modelo, el flujo de información entre sus componentes y la secuencia de procesamiento seguida desde los datos meteorológicos de entrada hasta la generación de predicciones puntuales e intervalos conformales.



**Figura 2.** Esquema conceptual del modelo VEGA-RAD, mostrando la interacción entre el proxy físico-astronómico, la memoria estocástica y la corrección estadística mediante aprendizaje automático, así como la inclusión de intervalos conformes para la cuantificación de la incertidumbre

Los detalles completos del preprocesamiento de datos, el esquema de entrenamiento, la selección de hiperparámetros y la construcción de intervalos conformes se describen en las Secciones 2.3.6, 2.3.6.1 y 3.

## 3. RESULTADOS Y DISCUSIÓN

En esta sección se presentan y analizan los resultados obtenidos con el modelo VEGA-RAD, aplicado a la predicción diaria de la radiación solar en la ciudad de Tarapoto durante el período 01 de enero de 2017 al 26 de septiembre de 2025. El análisis se desarrolló considerando métricas de desempeño deterministas y una evaluación probabilística basada en intervalos conformes, con el objetivo de responder al propósito central del estudio.

Los resultados que se presentan a continuación corresponden a la evaluación del modelo VEGA-RAD bajo un esquema de validación temporal y un conjunto de prueba independiente, de acuerdo con el protocolo experimental descrito en la Sección 2.3.6.

### 3.1. Configuraciones evaluadas: modelo base y modelo optimizado

Con el fin de evaluar el impacto de los distintos componentes del modelo VEGA-RAD, se analizaron dos configuraciones experimentales diferenciadas. La configuración base corresponde a una formulación simplificada del modelo, en la cual la estimación de la radiación solar se obtiene a partir del proxy físico-astronómico y retardos temporales básicos, sin incorporar mecanismos avanzados de corrección estadística ni términos estacionales explícitos.

Por su parte, la configuración optimizada integra la formulación completa del modelo VEGA-RAD, incorporando el componente de memoria estocástica, términos armónicos de Fourier para capturar la estacionalidad anual y una corrección estadística adaptativa basada en aprendizaje automático, aplicada sobre el residuo logarítmico. Ambas configuraciones fueron evaluadas bajo el mismo esquema de validación temporal y conjunto de prueba independiente, garantizando una comparación justa y consistente de su desempeño predictivo.

### 3.2. Desempeño predictivo del modelo

La Tabla 3 resume el desempeño del modelo bajo dos configuraciones evaluadas para Tarapoto en el periodo 2017–2025.

**Tabla 3.** Desempeño predictivo del modelo VEGA-RAD en dos configuraciones

| Configuración      | MAE   | RMSE  | R <sup>2</sup> |
|--------------------|-------|-------|----------------|
| Versión base       | 1,699 | 2,309 | 0,635          |
| Versión optimizada | 0,477 | 1,459 | 0,854          |

Nota: MAE y RMSE se expresan en kWh/m<sup>2</sup>/d. La versión base utiliza proxy de cielo claro y variables meteorológicas con retardos simples. La versión optimizada incorpora índice de claridad (CMF), retardos de nubosidad y residuo, medias móviles y términos armónicos de Fourier. Los valores se calcularon sobre 3155 días válidos del periodo 2017–2025.

La versión base alcanzó un error absoluto medio (MAE) de 1,699 kWh/m<sup>2</sup>/d, un error cuadrático medio (RMSE) de 2,309 kWh/m<sup>2</sup>/d y un coeficiente de determinación R<sup>2</sup> = 0,635. En contraste, la versión optimizada redujo el MAE a 0,477 kWh/m<sup>2</sup>/d y el RMSE a 1,459 kWh/m<sup>2</sup>/d, incrementando el coeficiente de determinación hasta R<sup>2</sup> = 0,854.

### 3.3. Análisis inferencial del desempeño predictivo

Con el objetivo de evaluar si las diferencias observadas entre la configuración base y la configuración optimizada del modelo VEGA-RAD son estadísticamente significativas, se realizó un análisis inferencial basado en la comparación pareada de los errores diarios sobre el conjunto de prueba independiente.

Dado el carácter temporal de la serie y la ausencia de supuestos de normalidad en la distribución de los errores, se aplicó la prueba no paramétrica de Wilcoxon para muestras pareadas. El análisis se realizó sobre el error absoluto diario de ambas configuraciones del modelo, considerando instantes temporales coincidentes del conjunto de prueba. El contraste se formuló bajo la hipótesis alternativa de que la configuración optimizada presenta un error sistemáticamente menor que la configuración base.

Los resultados de la prueba evidenciaron una diferencia estadísticamente significativa a favor de la versión optimizada del modelo ( $p < 0,01$ ). Adicionalmente, un análisis de bootstrap sobre la diferencia media del error absoluto confirmó una reducción promedio de  $\Delta\text{MAE} = 1,23 \text{ kWh/m}^2/\text{d}$ , con un intervalo de confianza del 95 % [1,18; 1,27] kWh/m<sup>2</sup>/d, el cual no incluye el valor nulo. Estos resultados confirman que las mejoras observadas en las métricas globales no responden a fluctuaciones aleatorias, sino a un efecto estructural



asociado a la incorporación de memoria estocástica, estacionalidad armónica y corrección estadística adaptativa mediante aprendizaje automático.

Los resultados del análisis inferencial se resumen en la Tabla 4, donde se evidencia una mejora estadísticamente significativa y robusta de la versión optimizada del modelo VEGA-RAD respecto a la versión base.

**Tabla 4.** Análisis inferencial del desempeño predictivo del modelo VEGA-RAD

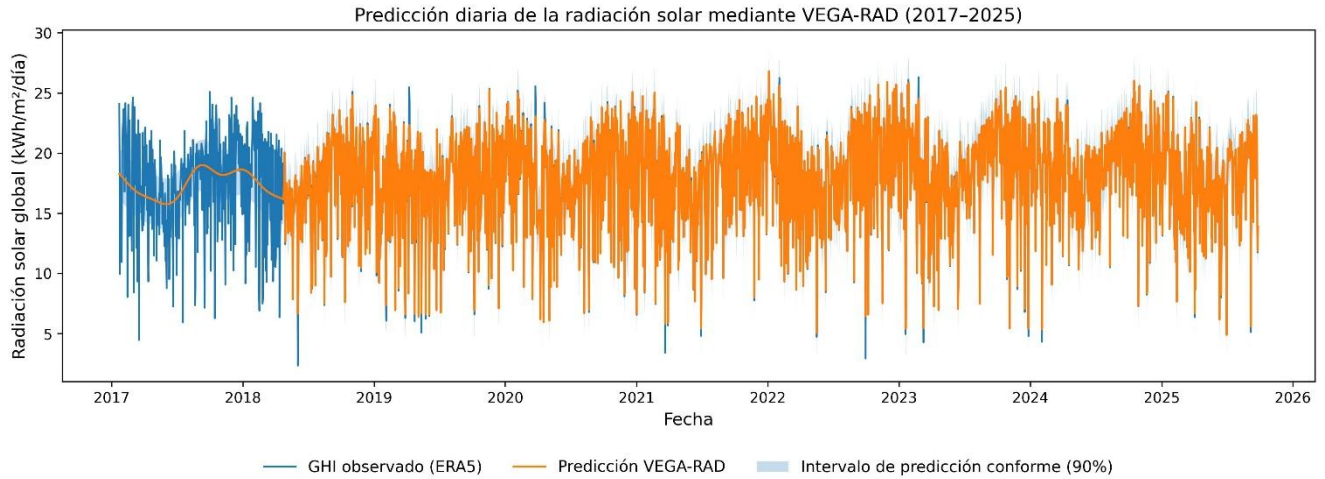
| Análisis inferencial | Métrica evaluada            | Resultado                          | Interpretación   |
|----------------------|-----------------------------|------------------------------------|--|
| Wilcoxon pareado     | Error absoluto diario (MAE) | $p < 0,001$                        | Diferencia estadísticamente significativa a favor de la versión optimizada |
| Bootstrap (B = 5000) | $\Delta$ MAE medio          | 1,23 kWh/m <sup>2</sup> /d         | Reducción promedio del error absoluto                                      |
| Bootstrap (IC 95 %)  | $\Delta$ MAE                | [1,18; 1,27] kWh/m <sup>2</sup> /d | Intervalo no incluye 0; mejora robusta                                     |

Nota: El análisis inferencial se realizó sobre errores diarios pareados correspondientes al conjunto de prueba independiente (n= 3170), considerando instantes temporales coincidentes para ambas configuraciones del modelo.

3.4. Análisis temporal de las predicciones

La Figura 3 presenta la comparación temporal entre las observaciones de radiación solar derivadas de ERA5 y la predicción central generada por VEGA-RAD, junto con el intervalo conformal del 90 %. Se observó que el modelo reprodujo adecuadamente tanto la estacionalidad anual como la variabilidad diaria de la radiación solar, manteniendo coherencia con los patrones climáticos característicos de la región.

Durante episodios de alta variabilidad asociados a incrementos de nubosidad, la versión optimizada del modelo mostró una mayor estabilidad y menor dispersión del error, lo que confirmó el aporte de los componentes de memoria estocástica y corrección estadística adaptativa.



**Figura 3.** Predicción diaria de la radiación solar mediante el modelo VEGA-RAD para Tarapoto (2017–2025). Se muestran los valores observados de ERA5, la predicción central del modelo y el intervalo de predicción conformal del 90 %

Con el fin de complementar el análisis visual presentado en la Figura 3, se evaluó adicionalmente el desempeño del modelo VEGA-RAD de forma desagregada por periodos climáticos característicos de la región amazónica. En particular, se analizaron las métricas de error correspondientes a la estación seca y la estación húmeda sobre el conjunto de prueba independiente. Los resultados muestran que el error absoluto medio (MAE) y el error cuadrático medio (RMSE) son menores durante la estación seca (MAE = 0,323 kWh/m<sup>2</sup>/d; RMSE = 1,163 kWh/m<sup>2</sup>/d) y se incrementan durante la estación húmeda (MAE = 0,660 kWh/m<sup>2</sup>/d; RMSE = 1,757 kWh/m<sup>2</sup>/d), en coherencia con la mayor nubosidad y variabilidad atmosférica. No obstante, el modelo mantiene estabilidad predictiva y una ventaja consistente en ambos regímenes

climáticos, lo que evidencia la robustez del enfoque VEGA-RAD bajo condiciones atmosféricas contrastantes. Los valores desagregados correspondientes a cada periodo climático se resumen en la Tabla 5.

**Tabla 5.** Desempeño predictivo del modelo VEGA-RAD por periodo climático (versión optimizada)

| Periodo climático | MAE (kWh/m <sup>2</sup> /d) | RMSE (kWh/m <sup>2</sup> /d) | n    |
|-------------------|-----------------------------|------------------------------|------|
| Estación seca     | 0,323                       | 1,163                        | 1621 |
| Estación húmeda   | 0,66                        | 1,757                        | 1549 |

Nota: El análisis se realizó sobre el conjunto de prueba independiente, considerando particiones temporales coherentes con el esquema de validación descrito en la Sección 2.3.6.

3.5. Evaluación probabilística e incertidumbre

La evaluación probabilística se aborda mediante el análisis de la cobertura de los intervalos conformales y su comportamiento global, mientras que la incertidumbre se analiza a partir de su calibración, grado de conservadurismo y estabilidad temporal.

En este contexto, la evaluación probabilística basada en intervalos conformales mostró coberturas coherentes con los niveles nominales del 90 % y 95 % en el conjunto de prueba, tal como se resume en la Tabla 6. En ambos casos, las coberturas fueron iguales o superiores a los niveles nominales, lo que indica una calibración conservadora pero confiable de los intervalos conformales. En consecuencia, VEGA-RAD no solo proporciona estimaciones puntuales precisas, sino también intervalos de predicción confiables y bien calibrados desde un punto de vista frecuentista.

**Tabla 6.** Evaluación probabilística de los intervalos conformales del modelo VEGA-RAD en el conjunto de prueba para Tarapoto (2017–2025)

| Nivel nominal | Cobertura observada | Ancho medio del intervalo (kWh/m <sup>2</sup> /d) | CV del ancho | Conjunto |
|---------------|---------------------|---|--------------|----------|
| 90%           | 1                   | 4,23  | ≈ 0          | Test     |
| 95%           | 1                   | 4,23  | ≈ 0          | Test     |

Si bien las coberturas observadas para ambos niveles nominales alcanzan el valor unitario en el conjunto de prueba analizado, este resultado puede interpretarse como indicativo de un comportamiento ligeramente conservador de los intervalos conformales. Dicho conservadurismo es coherente con la elevada variabilidad atmosférica característica del contexto amazónico y con la naturaleza no estacionaria de la radiación solar diaria. No obstante, este comportamiento resulta deseable desde una perspectiva de gestión del riesgo, ya que prioriza la fiabilidad de la cobertura frente a la subestimación de la incertidumbre.

Adicionalmente, el ancho medio de los intervalos conformales se mantuvo estable a lo largo del periodo de prueba, con un coeficiente de variación prácticamente nulo, lo que indica una cuantificación homogénea de la incertidumbre en el tiempo. Este resultado refuerza la utilidad práctica del enfoque propuesto para aplicaciones de planificación, diseño y operación de sistemas fotovoltaicos en contextos amazónicos de alta variabilidad atmosférica.

3.6. Discusión de los hallazgos

Los resultados obtenidos confirmaron que la integración de fundamentos físicos interpretables con memoria temporal estocástica y corrección estadística mediante aprendizaje automático constituye una estrategia eficaz para abordar la elevada variabilidad atmosférica característica de regiones tropicales. En particular, la reducción superior al 70 % del error absoluto medio (MAE) entre la versión base y la versión optimizada de VEGA-RAD evidenció el impacto significativo de incorporar términos de memoria, estacionalidad explícita y aprendizaje adaptativo, en línea con tendencias recientes reportadas en la literatura sobre predicción de radiación solar basada en modelos híbridos y aprendizaje automático (Bakır, 2024; Demir, 2025; Yadav et al., 2025; Zerouali et al., 2025).

Diversos estudios recientes han demostrado que los enfoques basados en aprendizaje automático superan sistemáticamente a modelos estadísticos convencionales al capturar no linealidades complejas del sistema atmósfera-superficie (Ghareeb et al., 2025; Tandon et al., 2025).

Sin embargo, dichos trabajos suelen apoyarse en arquitecturas profundas de alta complejidad y limitada interpretabilidad física. En contraste, el modelo propuesto introduce explícitamente un proxy físico-astronómico interpretable acoplado a una estructura de memoria estocástica, respondiendo a una limitación recurrente identificada en revisiones recientes del estado del arte (Shringi et al., 2025).

Desde una perspectiva metodológica, los hallazgos de este estudio son conceptualmente consistentes con avances reportados en enfoques híbridos CNN-SVR y CNN-LSTM para la predicción diaria de radiación solar (Ghimire et al., 2022; Hamdaouy et al., 2025; Şener & Tuğal, 2025). No obstante, a diferencia de estas aproximaciones, VEGA-RAD no se basa exclusivamente en la extracción automática de características latentes, sino que combina de forma parsimoniosa conocimiento físico, retardos autorregresivos y corrección estadística basada en gradient boosting, logrando un equilibrio favorable entre desempeño, interpretabilidad y costo computacional, tal como sugieren estudios recientes sobre modelos híbridos ligeros (Arseven & Çınar, 2025; Solano & Affonso, 2023).

En el contexto amazónico, los resultados obtenidos presentan coherencia con estudios basados en modelos numéricos regionales, como WRF-Solar, aplicados a la estimación de radiación solar global horizontal en el norte de Brasil (Amorim et al., 2024; Krishnan & Ravi Kumar, 2025; Alves et al., 2025). Estos trabajos coinciden en señalar que la nubosidad persistente y la variabilidad intra e interanual representan los principales desafíos para la predicción de radiación solar en la región. Sin embargo, mientras los modelos numéricos dependen de parametrizaciones físicas complejas y alto costo computacional, VEGA-RAD demuestra que un enfoque híbrido apoyado en datos de reanálisis ERA5 puede capturar eficientemente dicha variabilidad con menor complejidad estructural, en concordancia con estudios recientes basados en datos satelitales y de reanálisis (Wu et al., 2025).

Un aspecto distintivo de este trabajo es la incorporación explícita de intervalos conformales para la cuantificación de la incertidumbre predictiva. Las coberturas observadas, coherentes con los niveles nominales del 90 % y 95 %, confirman la adecuada calibración probabilística del modelo, ampliando el análisis más allá de métricas puntuales. Este enfoque probabilístico responde a una necesidad creciente en la planificación energética y la evaluación de riesgo climático, tal como se discute en estudios recientes sobre resiliencia energética y sistemas renovables híbridos (Cui et al., 2025; Jadhav & Bhawar, 2025; Vignesh Kumar et al., 2025).

Conjuntamente, los resultados sugieren que VEGA-RAD representa una contribución metodológica relevante y original en el contexto específico de la Amazonía peruana, al integrar de manera coherente un proxy físico-astronómico interpretable, memoria estocástica, corrección estadística mediante aprendizaje automático y cuantificación explícita de la incertidumbre en un marco unificado.

Si bien componentes individuales de este enfoque han sido explorados previamente en la literatura, su integración sistemática bajo un esquema de validación temporal, análisis inferencial riguroso y evaluación probabilística basada en intervalos conformales no ha sido reportada de forma conjunta para regiones amazónicas caracterizadas por alta variabilidad atmosférica. En este sentido, VEGA-RAD se posiciona como una alternativa robusta, interpretable y reproducible frente a enfoques puramente físicos o basados en datos, contribuyendo a cubrir vacíos metodológicos identificados en revisiones recientes del área (Shringi et al., 2025; Zerouali et al., 2025).

## CONCLUSIONES

En este estudio se formuló y evaluó un modelo híbrido denominado VEGA-RAD (Vega Radiative Adaptive Dynamics), diseñado para la predicción diaria de la radiación solar en la Amazonía peruana. El modelo integra de manera coherente un proxy físico-astronómico interpretable, un componente de memoria temporal estocástica y una etapa de corrección estadística adaptativa basada en aprendizaje automático. Esta formulación permitió capturar simultáneamente la estructura física del proceso radiativo, la persistencia temporal y las no linealidades residuales asociadas a variables meteorológicas locales, manteniendo una arquitectura parsimoniosa y reproducible.

Los resultados del análisis predictivo e inferencial evidenciaron mejoras sustanciales en el desempeño de la versión optimizada frente a la configuración base, con reducciones superiores al 70 % en el error absoluto medio y aumentos significativos en el coeficiente de determinación. Estas mejoras fueron respaldadas mediante análisis inferencial pareado (prueba de Wilcoxon) y técnicas de remuestreo bootstrap no paramétrico, confirmando que las diferencias observadas no son atribuibles al azar. Adicionalmente, la incorporación de intervalos conformales permitió cuantificar explícitamente la incertidumbre predictiva, obteniendo coberturas coherentes con los niveles nominales y un comportamiento conservador y estable en el tiempo. En conjunto, VEGA-RAD se posiciona como una alternativa robusta, interpretable y confiable para la predicción de radiación solar en regiones tropicales de alta variabilidad climática, con utilidad directa en la planificación y gestión de sistemas fotovoltaicos.

## FINANCIAMIENTO

Los autores no recibieron patrocinio para realizar este estudio-artículo.

## CONFLICTO DE INTERESES

No existe ningún conflicto de intereses relacionado con este trabajo.

## CONTRIBUCIÓN DE AUTORES

Conceptualización, curación de datos, análisis formal, investigación, metodología, administración de proyectos, recursos, software, visualización, redacción - borrador original: Agreda-Vega, J. F. Supervisión y Redacción - revisión y edición: Sare-Lara, E. y Rosales-Huamani, A. R. Validación: Agreda-Vega, J. F. y Sare-Lara, E.

## DISPONIBILIDAD DE DATOS

El procesamiento de los datos, la implementación del modelo VEGA-RAD y la evaluación experimental fueron realizados utilizando el lenguaje de programación Python en un entorno computacional reproducible basado en Google Colab. El flujo de trabajo incluyó la descarga automatizada de datos del reanálisis ERA5 mediante la API de Open-Meteo, el preprocesamiento de las variables, el entrenamiento del modelo y la generación de métricas de desempeño e intervalos de predicción conformales.

Los datos empleados en este estudio son de acceso público y pueden reproducirse mediante la API de Open-Meteo, especificando las coordenadas geográficas, el periodo temporal y las variables descritas en la Sección 2.2. El código utilizado para la implementación del modelo está disponible previa solicitud razonable al autor correspondiente.

## REFERENCIAS

- 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>
- Chen, Y., & Gao, Y. (2022). Comparative analysis of digital trade development strategies and governance approaches. *Journal of Digital Economy*, 1(3), 227–238. <https://doi.org/10.1016/j.jdec.2023.02.001>
- 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>
- Ghareeb, N., Alanazi, A., Sedaghat, A., Farhat, M. H., Mehdizadeh, A., Salem, H., Nazififard, M., & Mostafaiepour, 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>
- 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>

## ANEXOS

### Anexo A.1. Ejemplo de consulta a la API de Open-Meteo (ERA5)

Con el fin de garantizar la reproducibilidad del estudio, a continuación, se presenta un ejemplo de solicitud HTTP utilizada para la descarga de los datos diarios del reanálisis ERA5 mediante la API de Open-Meteo, correspondiente a la ciudad de Tarapoto (Perú) y al período comprendido entre el 1 de enero de 2017 y el 26 de septiembre de 2025.

El ejemplo de consulta se proporciona únicamente con fines ilustrativos, con el objetivo de documentar el acceso a la fuente de datos empleada. La implementación computacional completa, incluyendo el preprocesamiento de las variables, el entrenamiento del modelo y la evaluación experimental fue realizada en Python dentro de un entorno reproducible basado en Google Colab, y se encuentra disponible previa solicitud al autor correspondiente.

[https://archive-api.open-meteo.com/v1/era5?latitude=-6.5&longitude=-76.3&start\\_date=2017-01-01&end\\_date=2025-09-](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)

[26&daily=shortwave\\_radiation\\_sum,cloudcover\\_mean,temperature\\_2m\\_mean,relative\\_humidity\\_2m\\_mean,windspeed\\_10m\\_max&timezone=UTC](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)

Nota: Para la ejecución en Python, esta consulta fue implementada mediante la librería requests, tal como se muestra en el código reproducible utilizado en este estudio.

```
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()
```