

## ADAPTACIÓN AL SITIO DE IRRADIANCIA SOLAR GLOBAL USANDO APRENDIZAJE AUTOMÁTICO. CASO DE ESTUDIO: COTA COTA Y CHIVAY (PERÚ)

Miriam Manrique<sup>1,2</sup>, Rubén Ledesma<sup>2,3</sup>, Nicolás Rivera-Lera<sup>2,3</sup>, Germán Salazar<sup>2,3</sup>

<sup>1</sup>Universidad Nacional de Ingeniería (UNI) - Av. Tupac Amaru 210-Rímac-Lima-Perú.

<sup>2</sup>Grupo de Estudio y Evaluación de la Radiación Solar (GEERS)-INENCO-CONICET-Av. Bolivia 5150 - Salta - Argentina.

<sup>3</sup>Facultad de Ciencias Exactas - Universidad Nacional de Salta -Av. Bolivia 5150 – Salta - Argentina.  
e-mail: miriam.manrique.m@uni.pe

**RESUMEN:** En este estudio aplicamos la técnica de Adaptación al Sitio (AaS) para mejorar la estimación de la irradiancia global horizontal (GHI) de la base de datos satelital National Solar Radiation Database (NSRDB) en dos estaciones radiométricas del sur del Perú (Cota Cota y Chivay), para los años 2021 y 2022, con una frecuencia temporal horaria. Se evaluaron diferentes funciones adaptivas mediante modelos estadísticos, como la Regresión Lineal Simple (SLR), Regresión Lineal Múltiple (MLR) y un modelo de aprendizaje automático, (Perceptrón Multicapa, MLP). Las métricas utilizadas para evaluar el performance de estas funciones fueron el error medio de sesgo (rMBE) y el error cuadrático medio (rRMSE) porcentuales, así como sus versiones absolutas. Los resultados muestran que SLR y MLR corrigen el sesgo ( $rMBE \approx 0$ ), pero no los desvíos de las estimaciones (rRMSE), mientras que el MLP la reduce en aproximadamente un 4% en Chivay y en casi un 0,5% en Cota Cota. Esta mínima diferencia para Cota Cota podría deberse a que la base de datos satelital NSRDB tendría una buena correlación con los datos medidos, lo que no ocurría en Chivay.

**Palabras clave:** Irradiancia solar global, adaptación al sitio, aprendizaje automático, Perú.

### INTRODUCCIÓN

Conocer la distribución espacial y temporal del recurso solar en un sitio o región de interés es la primera tarea antes de cualquier otra al momento de diseñar o dimensionar un equipo con el que se pretenda aprovechar la energía solar para generar energía eléctrica.

Si no se dispone de datos de irradiancia solar (por ejemplo, Irradiancia Global Horizontal (GHI)) medidos para una región o sitio, se pueden usar datos estimados por modelos satelitales, como la National Solar Radiation Database de Estados Unidos (NSRDB, 2018), el Copernicus Atmosphere Monitoring Service de radiación solar (CAMS-Rad, 2004), el Land Surface Analysis Satellite Application Facility (LSA-SAF, 2005) o Solcast (2016), empresa que proporciona datos y pronósticos de la irradiancia solar y energía fotovoltaica. Sin embargo, esos modelos tienen desviaciones de los valores reales, por lo que siempre debe analizarse el error que presentan localmente (Yang y Bright, 2020; Salazar et al., 2020; Fernández-Peruchena et al., 2020). Existe un método que permite disminuir el error de los datos estimados por los modelos de las bases de datos satelitales, a través del uso de una serie de datos medidos en el sitio de interés: la Adaptación al Sitio (AaS). Esto permite utilizar los valores estimados (ya adaptados) con un sesgo menor al original, mejorando la situación general (Polo et al., 2020, 2016).



Tradicionalmente, se han utilizado métodos como la regresión lineal o el mapeo de cuantiles para minimizar el sesgo y mejorar el desempeño de estas técnicas en áreas geográficas específicas (Salazar et al., 2021). Sin embargo, estudios recientes sugieren que los modelos de aprendizaje automático ofrecen un enfoque más efectivo para la AaS, superando a las técnicas estadísticas convencionales (Salazar et al., 2024). Dado que los modelos de aprendizaje automático pueden adaptarse mejor al analizar las diferencias en las condiciones en cada localidad, su aplicación en la AaS para la estimación de irradiancia solar tiene un gran potencial. No obstante, la performance de estos modelos depende de las condiciones climáticas en cada sitio, de la orografía, de la calidad de las mediciones de irradiación y de la calidad del modelo satelital usado, lo que hace necesario desarrollar y comparar múltiples algoritmos para identificar la configuración de adaptación más adecuada (Esposito et al., 2024).

En el caso de Perú son pocas las estaciones que disponen de datos medidos de GHI. Por tanto, el presente trabajo evalúa y compara los modelos regresión lineal simple (SLR), múltiple (MLR), el Multilayer Perceptron (MLP) para la adaptación al sitio de datos de GHI provenientes de la base de datos satelital (NSRDB), utilizando registros de dos estaciones radiométricas ubicadas en Cota Cota y Chivay, ambas localizadas en el sur del Perú. El objetivo es analizar el desempeño de los modelos para mejorar la calidad de la información radiométrica en la región.

## METODOLOGÍA

### *Lugar de estudio*

Nuestro análisis y estudio se centra en dos estaciones radiométricas, Cota Cota y Chivay, ubicadas en el departamento de Arequipa, Perú (Figura 1). Las coordenadas geográficas de cada estación se presentan en la Tabla 1.

*Tabla 1: Coordenadas de ubicación de las estaciones radiométricas Cota Cota y Chivay. El signo negativo indica latitud sur y longitud oeste*

Estación	Latitud(°)	Longitud (°)	Altitud (m.s.n.m)
Cota Cota	-15,21	-71,39	4240
Chivay	-15,64	-71,60	3644



*Figura 1: Mapa de ubicación de las estaciones radiométricas de Cota Cota y Chivay.*

### **Datos medidos**

La GHI se obtuvo a partir de registros proporcionados por el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) de los años completos 2021 y 2022, usando un piranómetro de la marca Kipp & Zonen CMP3, con una frecuencia de registro horaria correspondiente a las estaciones de estudio.

### **Datos estimados**

Los datos estimados de GHI provienen del servicio NSRDB, que proporciona la irradiancia solar. La utilizada en este trabajo tiene una resolución espacial de 2 km y una frecuencia temporal de 10 min, y está disponible desde el año 2018. Dicha irradiancia es calculada por el Modelo Físico Solar (PSM) del Laboratorio Nacional de Energía Renovable (NREL) (Sengupta et al., 2018). Se empleó este conjunto de datos porque cubre toda Sudamérica, lo que resulta especialmente útil para nuestros lugares de estudio.

### **Preprocesamiento**

Antes de aplicar la adaptación al sitio, convertimos los datos satelitales estimados a escala horaria: promediamos los datos de 10 minutos para poder compararlos con los medidos. Luego, sincronizamos ambas fuentes a la misma zona horaria (GMT-5).

Los datos medidos en tierra de ambas estaciones radiométricas, fueron sometidos a un control de calidad. Se aplicaron los filtros mencionados en (Nollas et al., 2023) como se muestra en las Ecs. (1-3) para eliminar valores atípicos, datos nulos y anomalías instrumentales que puedan presentarse.

$$GHI < 1,5 S E_{SC} (\cos\theta_z)^{1,2} + 100 W/m^2 \quad (1)$$

$$GHI < (6,5331 - 0,065502 \theta_z + 1,8312 \times 10^{-4} \theta_z^2) / (1 + 0,01113 \theta_z) \quad (2)$$

$$K_t < 1,4 \quad (3)$$

donde:

$E_{SC}$  = Irradiancia Solar Total ( $1361 W/m^2$ )

$\theta_z$  = Ángulo cenital solar

$K_t$  = Índice de claridad

### **Adaptación al Sitio**

La AaS es un procedimiento que permite corregir estadísticamente series de irradiancia solar derivadas de bases de datos satelitales (BDS) utilizando como referencia datos medidos en tierra de menor extensión temporal. La corrección se realiza en el periodo de simultaneidad entre ambas series, en el cual se estima una función de calibración que relaciona los valores de GHI medidos en tierra y los estimados por el satélite; minimizando la diferencia entre ellos y generando una serie corregida. Una vez hallada, esta función se aplica al resto de la BDS fuera del periodo de simultaneidad. (Gueymard, 2012; Polo et al., 2016).

A continuación, se describen los modelos utilizados en este artículo para la realización de la AaS en los dos sitios peruanos mencionados:

La SLR asume una relación lineal entre la variable independiente y una dependiente. En este estudio, la variable independiente corresponde a la GHI estimada por la base de datos NSRDB ( $GHI_{NSRDB}$ ), mientras que la variable dependiente, es la GHI medida en las estaciones ( $GHI_{Medido}$ ).

La MLR incorpora más variables predictoras. En nuestro caso, se consideraron la GHI de cielo claro, la temperatura, la humedad relativa, el ángulo cenital solar y el coseno del ángulo cenital solar. Cabe resaltar que estas variables fueron seleccionadas porque reducen el sesgo y la dispersión de los datos.

Las expresiones matemáticas de la SLR y MLR, se presentan en (Salazar et al., 2024). Estos dos modelos lo que hacen es ajustar los datos de NSRDB a los datos medidos en tierra.

El MLP es posiblemente el método más interesante para encontrar una función adaptativa. Se trata de un tipo de red neuronal artificial (RNA) con una estructura feedforward compuesta por capas de entrada, ocultas y de salida, donde las neuronas intermedias aplican funciones de activación no lineales (Rumelhart et al., 1986). El proceso de entrenamiento se realiza mediante aprendizaje supervisado, comparando las salidas generadas por la red con los valores medidos y calculando un error, que se retropropaga a través de la red para ajustar los pesos mediante un algoritmo de optimización (Salazar et al., 2024).

Es importante tener en cuenta que, en SLR y MLR, no requiere etapas de entrenamiento y prueba en el mismo sentido que un MLP, porque su ajuste es analítico y directo. Sin embargo, en esta investigación se tomó el año 2021 para ajuste (Train) y el año 2022 para prueba (Test), no porque sea obligatorio para el modelo, sino para evaluar su capacidad de generalización y comparar con método de aprendizaje automático. Este punto de vista se debe tomar en cuenta al analizar la Tabla 3.

### ***Entrenamiento, Validación y Prueba para el MLP***

En el caso del MLP, es indispensable separar los datos en conjuntos de entrenamiento, validación y prueba, ya que el modelo ajusta sus pesos mediante un proceso iterativo de optimización y existe el riesgo de sobreajuste (overfitting). Esto ocurre cuando el modelo se especializa excesivamente en una función de correlación sobre la serie de datos de entrenamiento, logrando métricas muy favorables en dicho periodo, pero mostrando un desempeño deficiente fuera de él. Por ello, resulta fundamental verificar la capacidad de generalización del modelo (Abu-Mostafa, Magdon-Ismael y Lin, 2012).

Durante la etapa de entrenamiento, el objetivo es estimar la función que relaciona los valores de irradiancia medidos y los estimados. Para ello, en este trabajo se utilizaron los datos del año 2021, de los cuales se asignó un 80 % al entrenamiento y un 20 % a la validación, seleccionados de forma aleatoria con el fin de evitar sesgos estacionales. En el entrenamiento, el modelo ajusta sus parámetros para minimizar el error entre los valores observados y estimados, mientras que en la validación se evalúa el desempeño del modelo frente a datos no vistos en el entrenamiento.

Finalmente, para comprobar la verdadera capacidad de generalización y reducir el riesgo de sobreajuste, se implementó una tercera fase: la prueba (Test). En esta etapa se utilizaron los datos del año 2022 en ambas estaciones, completamente independientes de los utilizados en el entrenamiento y validación.

### ***Métricas de desempeño***

Para comparar y evaluar los resultados de los modelos se realizó un análisis de sus métricas de desempeño, siguiendo la metodología aplicada en estudios previos (Ledesma et al., 2023; Salazar et al., 2021; Fernández-Peruchena et al., 2020), tal como se muestra en la Tabla 2.

*Tabla 2: Métricas de desempeño*

<b>Métricas de desempeño utilizadas en este análisis</b>	<b>Ecuación</b>
Error medio de sesgo (Mean Bias Error)	$MBE = \frac{1}{N} \sum_{i=1}^N y_{est,i} - y_{med,i}$
Error cuadrático medio	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{est,i} - y_{med,i})^2}$
Error medio sesgo porcentual o relativo	$rMBE = \frac{MBE}{\bar{y}_{med}} \times 100\%$
Error cuadrático medio porcentual o relativo	$rRMSE = \frac{RMSE}{\bar{y}_{med}} \times 100\%$

donde:

$y_{est,i}$  = Valor estimado

$y_{med,i}$  = Valor medido  
 $\bar{y}_{med}$  = Media de los valores medidos

## RESULTADOS Y DISCUSIÓN

### Resultados de las métricas de desempeño

La Tabla 3 muestra los resultados de la base satelital NSRDB comparada con mediciones de GHI en las dos estaciones de estudio: Cota Cota y Chivay, aplicando tres modelos de ajuste: SLR, MLR y MLP. La comparación se realiza con las métricas RMSE y MBE y sus valores porcentuales respectivamente, tanto para el conjunto de entrenamiento (Train) año 2021, como para el conjunto de prueba (Test) año 2022.

En Cota Cota, el modelo satelital NSRDB presenta un error rRMSE inicial de 20,04 % en entrenamiento y 16,36 % en prueba. Tras la corrección con SLR y MLR, se observa una reducción no tan marcada en rRMSE, tanto en Train como en el Test, con sesgos (rMBE) prácticamente nulos, lo que indica que la corrección lineal logra centrar las estimaciones, pero no mejora la dispersión de los datos. El MLP alcanza un error rRMSE reducido a 19,54 % (Train) y 15,88 % (Test), además de un sesgo mínimo (<0,1 %). En los tres casos la reducción del error es mínimo, esto puede indicarnos que la base NSRDB es buena para este lugar.

En Chivay, se presenta el mismo patrón, aunque con errores absolutos mayores en la estimación satelital inicial (rRMSE de 23,67 % en Train y 21,93 % en Test). Tanto SLR como MLR logran notorias mejoras, con reducciones de aproximadamente 3% en rRMSE y sesgos cercanos a cero. Sin embargo, el MLP ofrece una mejor métrica, alcanzando un rRMSE de 18,73% (Train) y 17,39 % (Test), reduciendo el error de dispersión en más de 4% respecto al NSRDB sin ajuste. Por tanto, de los tres modelos usados para la técnica de adaptación al sitio, el más recomendable para mejorar la GHI estimada por NSRDB en Chivay es el MLP.

Tabla 3: Métricas de desempeño de las estaciones radiométricas Cota Cota y Chivay

Estación	Modelo	Etapas	RMSD (W/m <sup>2</sup> )	MBE (W/m <sup>2</sup> )	rRMSE (%)	rMBE (%)
Cota Cota	NSRDB	Train	97,28	-6,23	20,04	-1,18
		Test	85,76	-8,68	16,36	-1,66
	SLR	Train	97,26	0,00	19,80	0,00
		Test	85,40	3,14	16,29	0,60
	MLR	Train	96,15	0,00	19,57	0,00
		Test	84,17	2,78	16,06	0,53
	MLP	Train	95,09	-1,83	19,54	-0,62
		Test	83,43	-5,06	15,88	-0,01
Chivay	NSRDB	Train	111,83	53,35	23,67	11,24
		Test	110,59	53,72	21,93	10,64
	SLR	Train	98,03	0,00	20,46	0,00
		Test	96,66	-0,17	19,15	-0,03
	MLR	Train	95,70	0,00	19,98	0,00
		Test	93,66	-2,92	18,55	-0,58
	MLP	Train	91,13	4,34	18,73	-0,40
		Test	90,05	5,40	17,39	-0,81

### Comparación de los gráficos de dispersión de las estaciones radiométricas Cota Cota y Chivay

Analizadas y observadas las métricas de desempeño se realizó los gráficos de dispersión presentados en las Figuras 2 y 3 que muestran la comparación entre  $GHI_{Medido}$  y los valores estimados por el satélite

( $GHI_{NSRDB}$ ), así como su versión corregida mediante el modelo de Perceptrón Multicapa ( $GHI_{Adaptado-MLP}$ ), para las estaciones radiométricas Cota Cota y Chivay.

En la estación Cota Cota (Figura 2), los datos estimados por el satélite NSRDB (puntos azules) muestran una ligera subestimación en la pendiente de 0,94 y un intercepto 39,5  $W/m^2$ . Con la adaptación al sitio mediante MLP (puntos naranjas), la pendiente permanece igual y el intercepto baja a 33,2  $W/m^2$ , reduciendo así el sesgo y mejorando la alineación de los datos respecto a la línea ideal  $y=x$  de color fucsia. Por tanto, el modelo adaptado logra un mejor ajuste a las mediciones reales, aunque todavía se aprecia una ligera subestimación en los valores más altos de radiación.

En la estación Chivay (Figura 3), la comparación muestra que los datos originales de NSRDB (puntos azules) tienden a sobreestimar la  $GHI_{Medido}$ , especialmente en valores bajos, con un intercepto alto de 83,3  $W/m^2$ . Tras aplicar la adaptación al sitio con MLP (puntos naranjas), la pendiente se mantiene en 0,94 pero el intercepto baja a 27,7  $W/m^2$ , lo que significa que el sesgo se reduce de forma notable. Por tanto, la nube de puntos adaptada (naranja) se alinea más con la línea  $y=x$  en comparación con la nube original (azul), lo cual evidencia una buena corrección de sesgo.

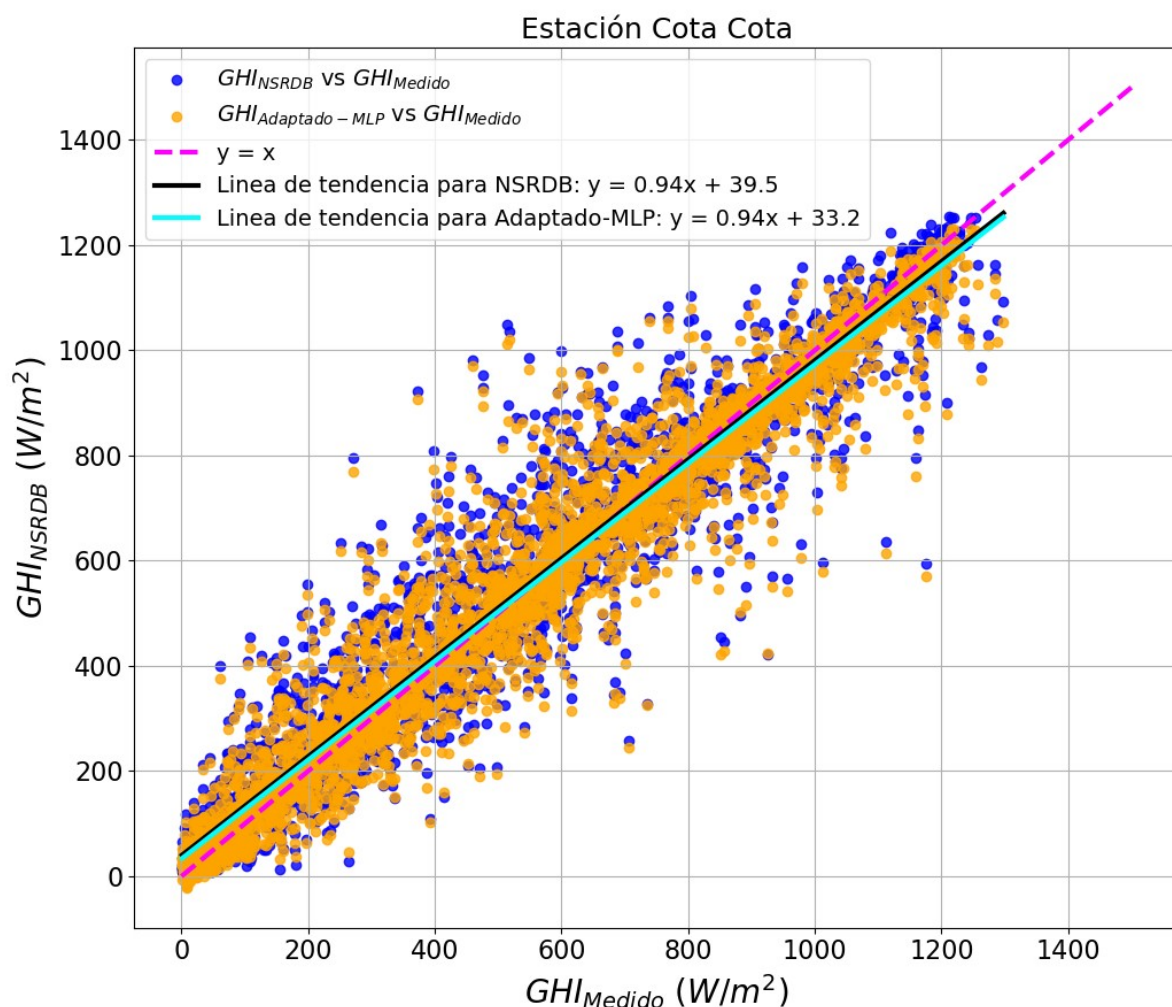


Figura 2: Diagrama de dispersión para la estación Cota Cota entre  $GHI_{NSRDB}$  y  $GHI_{Adaptado-MLP}$  VS  $GHI_{Medida}$ .

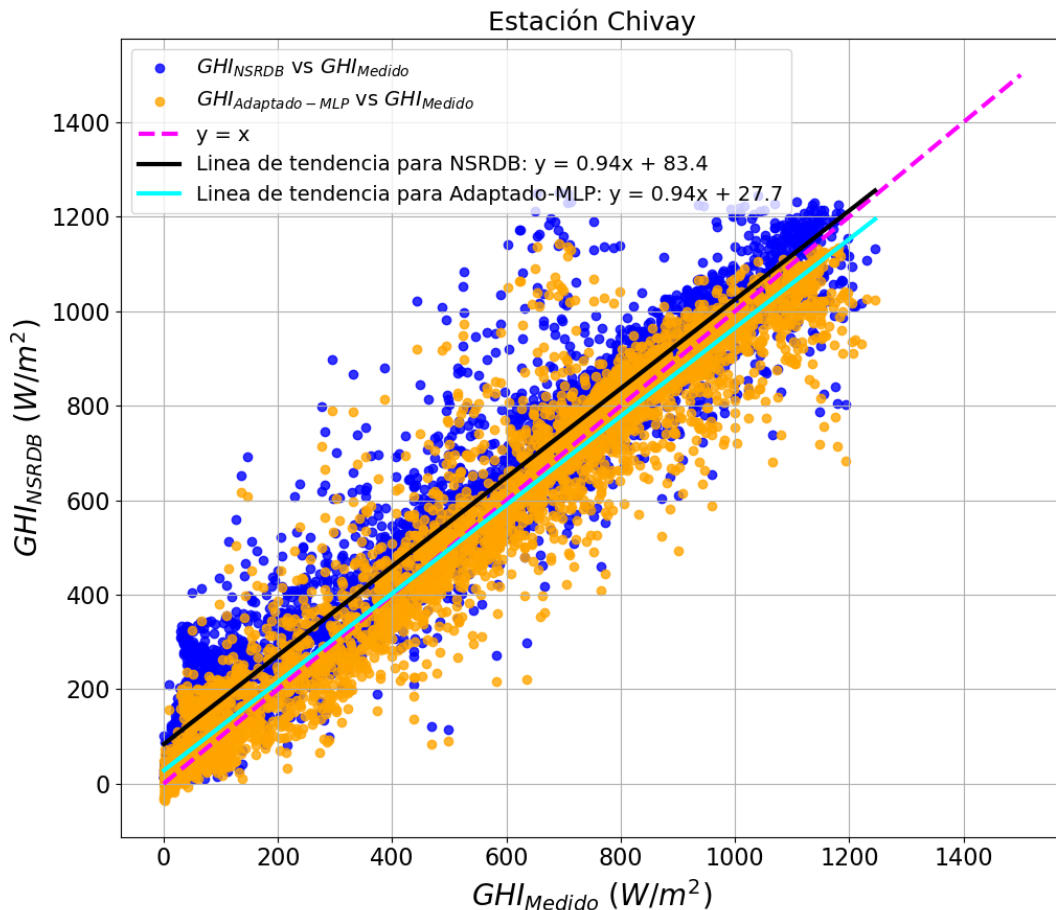


Figura 3: Diagrama de dispersión para la estación Chivay entre  $GHI_{NSRDB}$  y  $GHI_{Adaptada-MLP}$  vs  $GHI_{Medida}$ .

#### **Análisis de la comparación temporal diaria de GHI medido, estimado y ajustado mediante MLP**

En las Figuras 4 y 5 se muestra la evolución diaria de la GHI para diferentes fechas representativas en las estaciones Cota Cota y Chivay durante el año 2022, comparando las mediciones en tierra ( $GHI_{Medido}$ ) con las estimaciones satelitales originales de NSRDB ( $GHI_{NSRDB}$ ) y las series ajustadas mediante perceptrón multicapa ( $GHI_{Ajustado-MLP}$ ). Cabe señalar que no se incluyeron las curvas de ajuste de los modelos SLR y MLR, con el fin de evitar solapamientos y facilitar la visualización de las diferencias.

En Cota Cota (Figura 4), se observa que las curvas estimadas por NSRDB reproducen en general el patrón horario de la irradiancia, pero con una ligera sobreestimación en las horas cercanas al mediodía solar, particularmente en días despejados como el 08 y 09 de junio. En días como el 12 y 13 de noviembre se aprecia una buena adaptación. En fechas con presencia de nubosidad (por ejemplo, 21 y 22 de octubre), se muestra mayor diferencia entre ambas mediciones. El ajuste con el MLP reduce estas diferencias y aproxima la curva estimada a la medida.

En Chivay (Figura 5), las diferencias iniciales entre  $GHI_{NSRDB}$  y  $GHI_{Medido}$  son más pronunciadas, especialmente en días con alta nubosidad (03 y 05 de enero), donde las estimaciones satelitales tienden a sobrestimar significativamente en las primeras horas y subestimar en las últimas. El MLP atenúa estas diferencias, aunque persisten ligeras diferencias en días parcialmente nublados, como el 12 y el 13 de noviembre. En días como el 20 y 21 de octubre, se logra un ajuste muy cercano entre  $GHI_{Medido}$  y  $GHI_{Ajustado-MLP}$  a lo largo de casi todo el ciclo diario.

En conclusión, el análisis visual confirma que el MLP no solo reduce el sesgo medio, sino que mejora la representación de la forma de la curva diaria de GHI, especialmente en días con presencia de nubosidad; debido a que aprende patrones específicos del sitio.

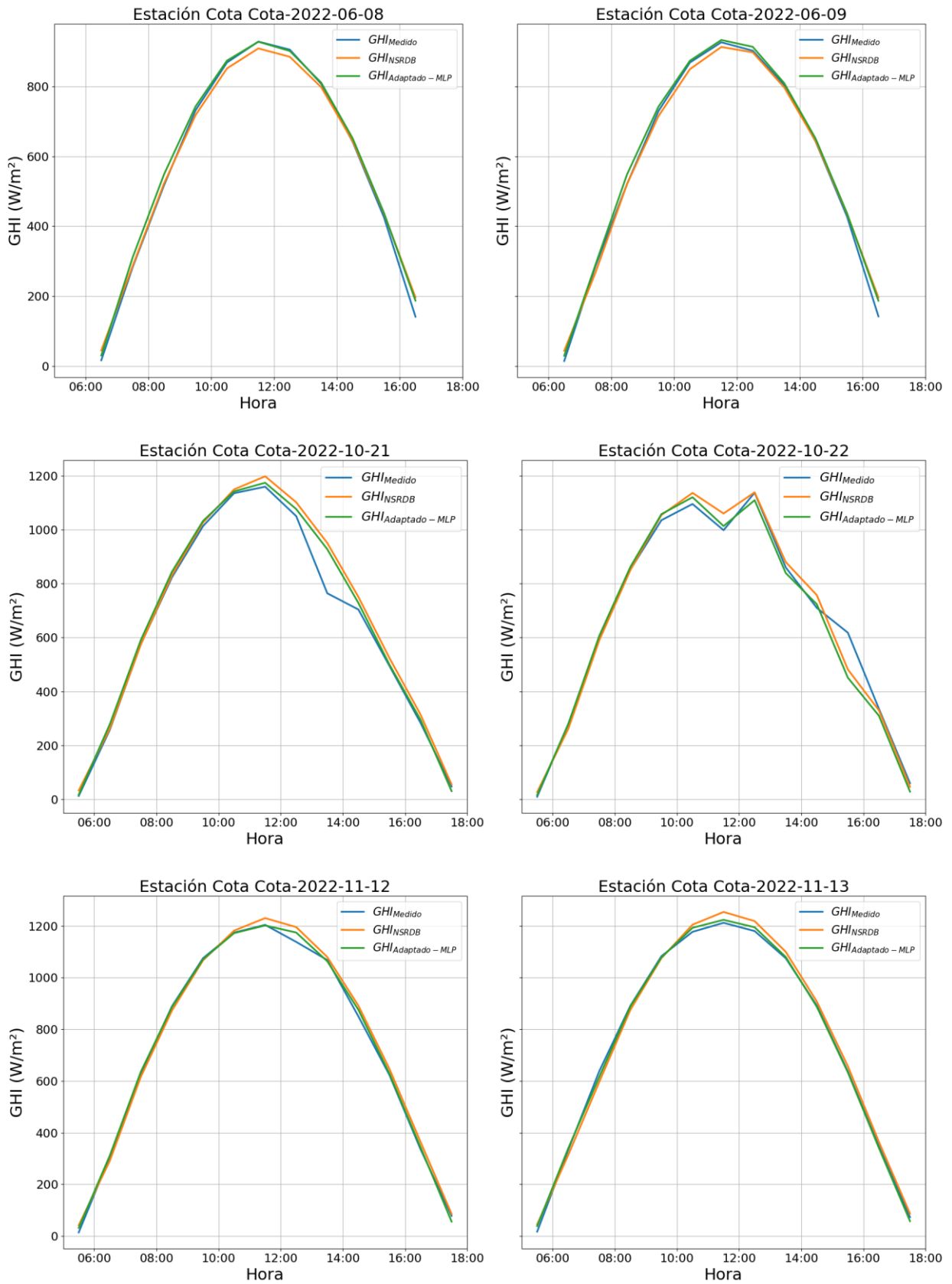


Figura 4: Representación de la evolución diaria de la GHI de la estación Cota Cota para 6 días del año 2022.

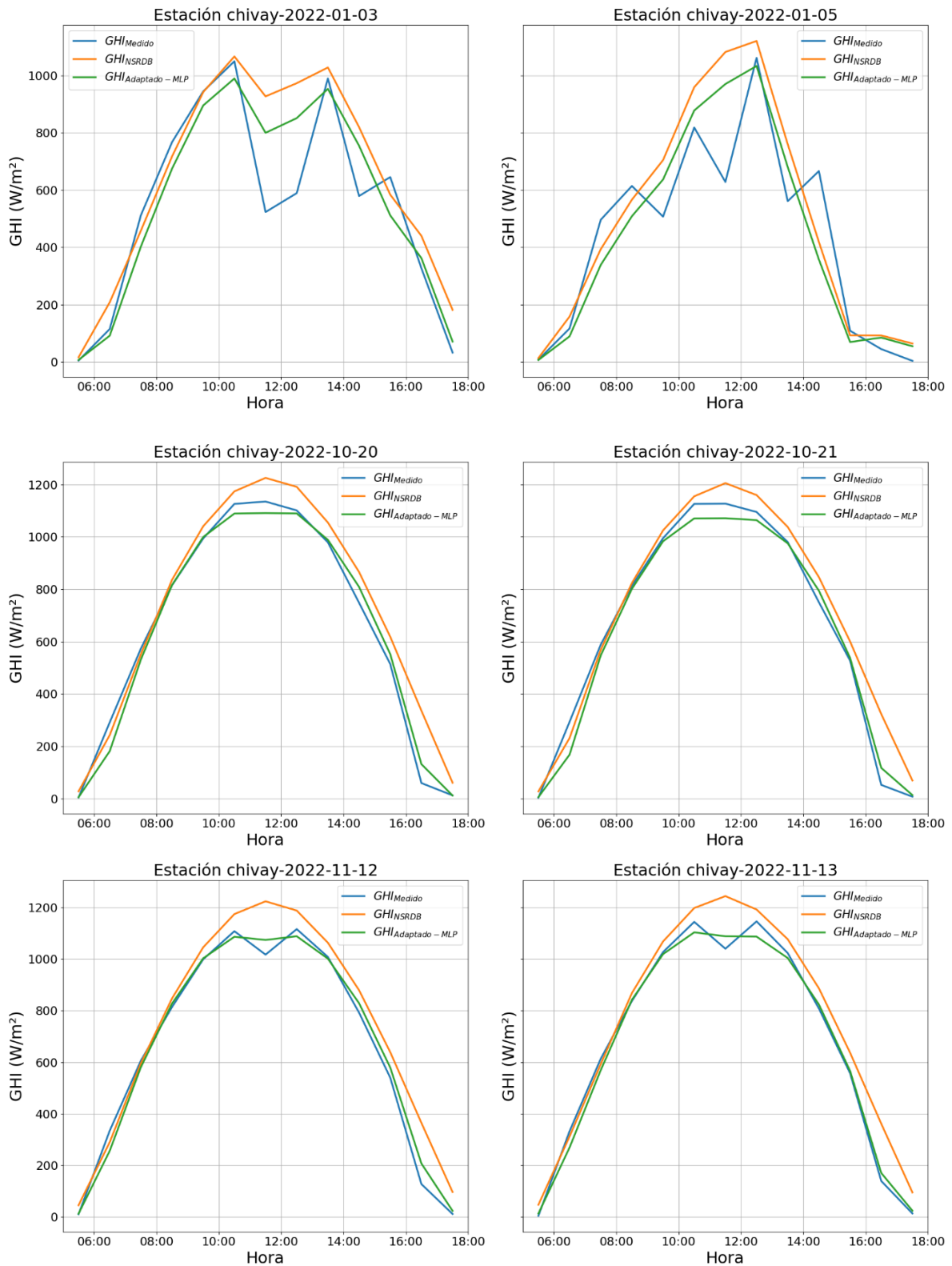


Figura 5: Representación de la evolución diaria de la GHI de la estación Chivay para 6 días del año 2022.

## CONCLUSIONES

Los resultados muestran que la AaS puede mejorar las métricas entre los valores de GHI medidos y adaptados. Particularmente, para los datos analizados de las dos estaciones, tanto SLR como MLR mostraron corrección del sesgo ( $rMBE \approx 0$ ) sin disminuciones en el rRMSE. El MLP mostró una corrección similar en el sesgo y una corrección en el rRMSE, de 0,5% en Cota Cota y de 4% en Chivay.

En el caso de Cota Cota, los resultados podrían indicar que la base satelital NSRDB ya presenta una buena correlación con los datos medidos, por lo que no es posible una mejora significativa (reducción del error) en los valores adaptados referidos a esta estación.

En el caso de Chivay, los modelos reducen todos el rRMSE, destacándose el MLP, que lo hace hasta un 4%, respecto al valor original del NSRDB. Esta diferencia hace razonable el uso de esta metodología para mejorar las estadísticas a largo plazo de la base de datos NSRDB adaptada para esta estación.

Este análisis muestra que la base de datos NSRDB puede ser confiable para la realización de cálculos previos de dimensionamiento de sistemas solares. También muestra que las metodologías utilizadas pueden servir para mejorar la precisión de datos locales obtenidos de bases satelitales, usando la metodología de la AaS. Sin embargo, esta es una línea de investigación que aún debe ser profundamente estudiada, mediante el análisis de bases de datos de radiación solar medidos en la región de interés.

## FUENTES DE FINANCIAMIENTO

CONCYTEC subvencionó este trabajo a través del Programa PROCIENCIA en el marco del concurso «Becas en programas de doctorado educativo a través de alianzas interinstitucionales», según contrato PE501090126-2024, y del Concurso de Alianzas Interinstitucionales para Programas de Doctorado correspondientes a la UNIVERSIDAD NACIONAL DE INGENIERÍA, según contrato PE501084296-2023- PROCIENCIA-BM.

## AGRADECIMIENTOS

Agradecemos al Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI), por la provisión de los datos de GHI de las estaciones Cota Cota y Chivay, porque sin ellos no hubiera sido posible realizar este análisis.

## REFERENCIAS

- Abu-Mostafa, Y., Magdon-Ismael, M., y Lin, H. (2012). *Learning from data*. 1ª edición. pp, 119-151. AMLBook, USA.
- CAMS Rad [en línea] Copernicus Atmosphere Monitoring Service - solar radiation. Dirección URL: <<https://atmosphere.copernicus.eu/solar-radiation-supplementary-products> [consulta: julio de 2025]
- Esposito, E., Leanza, G., y Di Francia, G. (2024). Comparative analysis of ground-based solar irradiance measurements and Copernicus satellite observations. *Energies*, 17(7), 1579. <https://doi.org/10.3390/en17071579>.
- Fernández-Peruchena, C. M., Polo, J., Martín, L., y Mazorra, L. (2020). Site-adaptation of modeled solar radiation data: The SiteAdapt procedure. *Remote Sensing*, 12(13). <https://doi.org/10.3390/rs12132127>
- Gueymard, C. A. (2012). Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models. In *Solar Energy* (Vol. 86, Issue 8, pp. 2145–2169). <https://doi.org/10.1016/j.solener.2011.11.011>
- Ledesma, R., Alonso-Suárez, R., Monetta, A., Laguarda, A., Vilela, O., & Salazar, G. (2023). Evaluación en Uruguay del producto DSR GOES-16 de irradiancia solar global horizontal. *Avances en Energías Renovables y Medio Ambiente (AVERMA)*

- LSA-SAF [en línea] Land Surface Analysis Satellite Application Facility. Dirección URL:< <https://lsa-saf.eumetsat.int/en/s>> [consulta: julio de 2025]
- Nollas, F. M., Salazar, G. A., & Gueymard, C. A. (2023). Quality control procedure for 1-minute pyranometric measurements of global and shadowband-based diffuse solar irradiance. *Renewable Energy*, 202, 40–55. <https://doi.org/10.1016/j.renene.2022.11.056>
- NSRDB [en línea] National Solar Radiation Database. Dirección URL: < <https://nsrdb.nrel.gov/> > [consulta: mayo de 2025]
- Polo J, Wilbert S, Ruiz-Arias JA, Meyer R, Gueymard C, Súrri M, et al. Preliminary survey on site adaptation techniques for satellite-derived and reanalysis solar radiation datasets. *Solar Energy* (2016); 132:25–37. <https://doi.org/10.1016/j.solener.2016.03.001>
- Polo, J., Fernández-Peruchena, C. M., Vindel, J. M., González, J. A., & Gueymard, C. A. (2020). Benchmarking on improvement and site-adaptation techniques for modeled solar radiation datasets. *Solar Energy*, 201, 469–479. <https://doi.org/10.1016/j.solener.2020.03.040>
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088), 533-536.
- Salazar, G., Ledesma, R. D., Ruiz, C. L., y De Castro Vilela, O. (2024). Análisis de desempeño de diferentes técnicas de aprendizaje automático en una adaptación al sitio de irradiancia solar global para Salta (Argentina). *Avances en Energías Renovables y Medio Ambiente (AVERMA)*.
- Salazar, G., Gueymard, C., Galdino, J. B., de Castro Vilela, O., & Fraidenraich, N. (2020). Solar irradiance time series derived from high-quality measurements, satellite-based models, and reanalyses at a near-equatorial site in Brazil. *Renewable and Sustainable Energy Reviews*, 117, 109478. <https://doi.org/10.1016/j.rser.2019.109478>.
- Salazar, G., Alonso Suárez, R., Laguarda Cirigliano, A., Ledesma, R. D. (2021). Evaluación del proceso de adaptación al sitio aplicado a la irradiancia solar global medida en la ciudad de Salta, Argentina. *Avances en Energías Renovables y Medio Ambiente (AVERMA)*.
- Solcast [en línea] empresa que proporciona datos y pronósticos de la irradiancia solar y energía fotovoltaica. Dirección URL:< <https://solcast.com/>> [consulta: julio de 2025]
- Sengupta, M., Xie, Y., Lopez, A., Habte, A., Maclaurin, G., y Shelby, J. (2018). The National Solar Radiation Data Base (NSRDB). In *Renewable and Sustainable Energy Reviews* (Vol. 89, pp. 51–60). Elsevier Ltd. <https://doi.org/10.1016/j.rser.2018.03.003>
- Yang, D., y Bright, J. M. (2020). Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27 years. *Solar Energy*, 210, 3–19. <https://doi.org/10.1016/j.solener.2020.04.06>

#### **SITE ADAPTATION OF GLOBAL SOLAR IRRADIANCE USING MACHINE LEARNING. CASE STUDY: COTA COTA AND CHIVAY (PERU)**

**ABSTRACT;** In this study, we applied the Site Adaptation (AaS) technique to improve the estimation of global horizontal irradiance (GHI) from the National Solar Radiation Database (NSRDB) at two radiometric stations in southern Peru (Cota Cota and Chivay) for the years 2021 and 2022, with hourly temporal frequency. Different adaptive functions were evaluated using statistical models, such as Simple Linear Regression (SLR), Multiple Linear Regression (MLR), and a machine learning model (Multilayer Perceptron, MLP). The metrics used to evaluate the performance of these functions were the mean bias error (rMBE) and the mean square error (rRMSE) in percentage terms, as well as their absolute versions. The results show that SLR and MLR correct the bias (rMBE  $\approx$  0), but not the deviations from the estimates (rRMSE), while MLP reduces it by approximately 4% in Chivay and by almost 0.5% in Cota Cota. This minimal difference for Cota Cota could be due to the fact that the NSRDB satellite database has a good correlation with the measured data, which was not the case in Chivay.

**Keywords:** Global solar irradiance, site adaptation, machine learning, Peru.