ASADES

Avances en Energías Renovables y Medio Ambiente - AVERMA
Vol. 29, pp. 556-567 2025, ISSN 2796-8111

IDENTIFICACIÓN DE MUESTRAS DE IRRADIANCIA SOLAR DE CIELO CLARO MEDIANTE APRENDIZAJE AUTOMÁTICO NO SUPERVISADO

Nicolás Rivera-Lera[1,2], Miriam Manrique[3], Germán Salazar[1,2], Rubén Ledesma[1,2], Agustín Laguarda[4]
[1]Instituto de investigaciones en Energía no Convencional (INENCO)
[2]Departamento de Física, Facultad de Ciencias Exactas, Universidad Nacional de Salta
[3]Universidad Nacional de Ingeniería (UNI), Lima Perú
[4]Laboratorio de Energía Solar, Instituto de Física, Facultad de Ingeniería, UdelaR
E-mail: nicolasrivera2297@gmail.com

RESUMEN: Este trabajo presenta un enfoque para la detección de muestras de cielo claro a partir del análisis de series temporales de Irradiancia Global Horizontal (GHI) y técnicas del tipo no supervisadas de machine learning. Además, propone un modelo local de cielo claro ajustado mediante regresión tipo potencia sobre la Irradiancia Horizontal en el tope de la atmósfera, sobre las muestras obtenidas. Se desarrolló un detector de muestras de cielo claro basado en umbrales de derivadas (primera y segunda), clustering no supervisado utilizando el modelo Gaussian Mixture Models y análisis de densidad de datos vecinos para ser aplicados sobre una serie de GHI. La aplicación de este detector se realizó sobre datos de GHI provenientes del sitio Desert Rock (Estados Unidos) a escala de 1 minuto, extraídos de la red Surface Radiation Budget Network (SURFRAD). Se contrastaron las muestras detectadas con el algoritmo de Reno y Hansen (2016), utilizado como referencia. Estas muestras mostraron más del 90% de coincidencia. El modelo de cielo claro obtenido presentó métricas menores a un 10% en rRMSE con el modelo McClear.

Palabras clave: irradiancia solar en plano horizontal**,** cielo claro, clústeres, modelos de cielo claro.

INTRODUCCIÓN

La radiación solar que ingresa a los sistemas que aprovechan esta energía puede verse atenuada por su interacción con nubes o aerosoles en la atmósfera (Shen et al., 2018). Los modelos de cielo claro (CC) estiman la radiación máxima que alcanza la superficie terrestre en ausencia de obstrucciones meteorológicas, por lo que son esenciales en el diseño y dimensionamiento de sistemas solares (Gueymard, 2008).

Estos modelos se basan en la geometría solar del sistema Sol-Tierra (Gueymard, 2012) y consideran factores como latitud, época del año, hora del día y altitud. Dichos parámetros se emplean en funciones que incluyen la masa de aire relativa para una trayectoria específica y el ángulo de incidencia solar sobre un plano horizontal. Además, pueden incorporar variables como la transmitancia de los distintos componentes atmosféricos.

Entre los modelos de CC más utilizados se encuentran McClear (Lefevre et al., 2013), de naturaleza física y basado en información atmosférica procedente de reanálisis satelital: ESRA (Rigollier et al., 2000), combina parametrizaciones empíricas con datos de turbidez de Linke (Louche et al., 1986; Linke, 1992); REST2 (Gueymard, 2008), es un modelo físico que resuelve la ecuación de transferencia radiativa con alta precisión para condiciones de CC; y modelos como ARGP y ARGP-v2 (Ledesma et al., 2022), diseñados específicamente para sitios de altura en Argentina, con ajustes empíricos adaptados a condiciones regionales, a los que se les realizó una validación para Salta en Rivera et al. (2024). La elección del modelo depende del objetivo de estudio, la disponibilidad de datos de entrada y la necesidad

de representar condiciones atmosféricas específicas.

La versatilidad de los modelos de CC permite estimar de forma rápida la radiación solar sin nubosidad en un sitio concreto, incluso sin mediciones locales. Son útiles en la planificación de proyectos solares y como referencia para construir modelos en condiciones atmosféricas diversas (Gueymard, 2012).

Aunque su rendimiento puede variar según la ubicación, la validación requiere mediciones en tierra que correspondan a condiciones de CC. Una forma común de obtenerlas es inspeccionar visualmente series temporales de irradiancia global horizontal (GHI) medidas con piranómetros, descartando valores afectados por nubes procedentes del disco solar. Una alternativa más simple y de bajo costo computacional es el algoritmo RENO-2016 (Reno y Hansen, 2016), que detecta automáticamente muestras de CC comparando la GHI medida con un modelo de CC y aplicando métricas como desvíos estándar en ventanas temporales, entre otras. Pese a sus buenos resultados, su lógica al momento de realizar validaciones puede resultar sesgada si el modelo de CC que requiere de entrada no representa adecuadamente las condiciones locales, ya que las muestras obtenidas se usan en la validación de otros modelos.

Frente a esto, este artículo propone el siguiente enfoque: un detector de muestras de CC a escala de 1 minuto basado en herramientas de machine learning (ML) y el análisis de series temporales para un determinado sitio, que mantiene un bajo costo computacional y de tipo no supervisado. Se complementan con las muestras detectadas, un nuevo modelo empírico local utilizando un ajuste del tipo potencia a partir de las muestras. En este trabajo, se utiliza el modelo McClear (Lefevre et al., 2013) como referencia para la validación y comparación con el nuevo modelo, dado que es de fácil acceso, provee estimaciones de alta resolución temporal y espacial respaldadas por datos globales y fue validado en diferentes estudios como en Antonanzas-Torres et al. (2019); Engerer y Mills (2015); Gueymard (2012); Rivera et al., (2024); Laguarda y Abal (2017); Laguarda et al. (2020); Russo et al. (2022). En este contexto, adoptamos la definición de cielo claro como aquellas condiciones atmosféricas en las que el disco solar no se ve afectado por nubosidad ni sombreamiento directo. Esta definición, en línea con la utilizada por McClear, incluye la atenuación producida por aerosoles y otros constituyentes atmosféricos, por lo que días con elevada turbidez pueden ser clasificados igualmente como cielo claro en ausencia de nubes.

METODOLOGÍA

El procedimiento propuesto para la detección de muestras de CC y el ajuste del modelo local se desarrolla en varias etapas secuenciales. La Figura 1 presenta un diagrama de flujo que resume de manera esquemática el procedimiento, desde la adquisición de los datos hasta las validaciones finales.

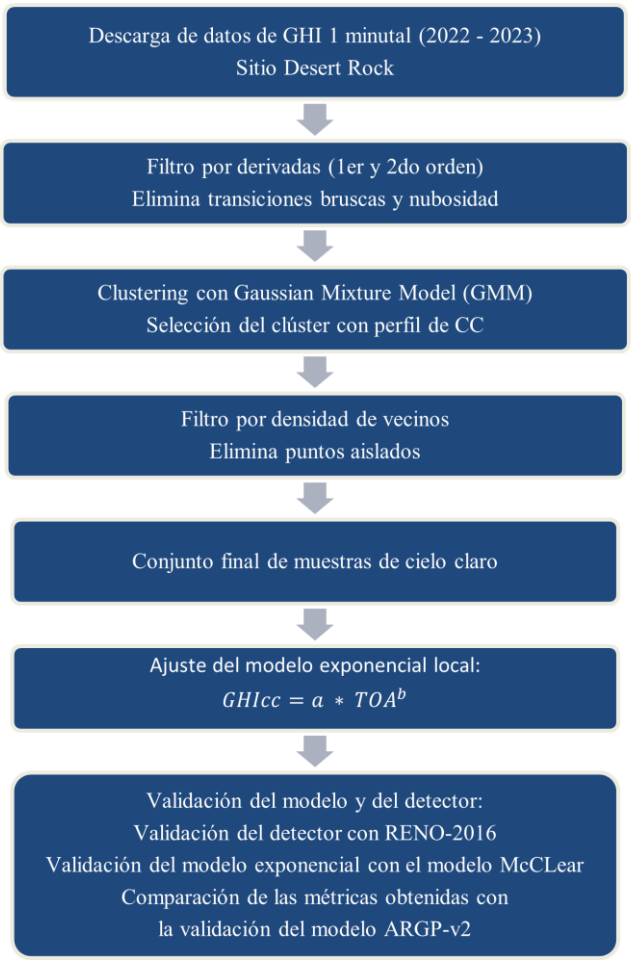


Figura 1: Diagrama de flujo del procedimiento propuesto

Datos

Se descargaron dos años (2022 y 2023) de datos con frecuencia de 1 minuto de GHI (Wm^{-2}) de la red SURFRAD (NOAA, 2025) para el sitio Desert Rock (DRA) (36,62373° N, 116,01947° O, 1007 m s. n. m.) ubicado en Nevada, Estados Unidos. Desert Rock se encuentra en una zona desértica como se observa en la Figura 2, con una clasificación climática BWh según Köppen-Geiger (Peel et al., 2007), que corresponde a un clima desértico cálido.

Figura 2: Estación SURFRAD, Desert Rock, Estados Unidos (Google Earth)

Tanto para obtener un modelo de CC como para la validación de los mismos, es imprescindible contar con muestras de mediciones que correspondan a momentos de cielo claro. Para detectar estas muestras, se trabajó con herramientas de ML y el análisis de series temporales.

Algoritmo (Reno y Hansen, 2016)

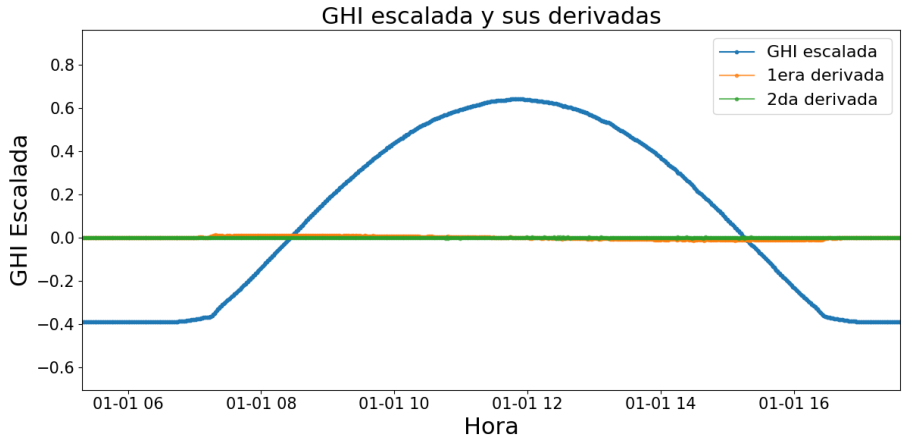
El algoritmo RENO-2016 es un algoritmo diseñado para identificar períodos de CC en series temporales de GHI. Toma como entrada la GHI medida y un modelo de CC de referencia, con el que compara las mediciones a través de cinco criterios calculados en ventanas temporales: (1) diferencia de medias, (2) diferencia de máximos, (3) longitud de línea, (4) variabilidad normalizada de la pendiente y (5) cambio máximo sucesivo. Solo los intervalos que cumplen simultáneamente todos los criterios se clasifican como muestras de CC.

La implementación de este algoritmo se realizó a través de la librería pvlib de Python con la función detect_clearsky, donde el algoritmo incluye una iteración que ajusta un factor de escala global (α) sobre el modelo de CC de entrada. Este factor se calcula con las muestras detectadas en la iteración previa para compensar sesgos multiplicativos y se aplica antes de volver a ejecutar los criterios. De esta manera, en cada iteración se actualiza tanto el valor de α como el conjunto de muestras detectadas como CC, repitiendo el proceso hasta converger o alcanzar un número máximo de iteraciones.

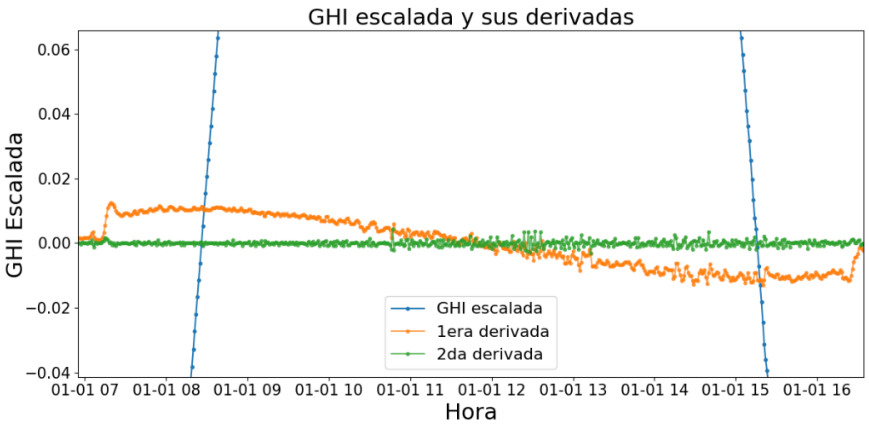
Cabe destacar que esta iteración solo ajusta el modelo de CC y no corrige errores en su forma o en cómo representa las condiciones locales. Por ello, las muestras detectadas pueden seguir reflejando ciertos sesgos del modelo original.

Análisis de la derivada

Mediante la inspección visual de un día característico y completo de CC sobre la serie temporal escalada (con una normalización estándar), se observó que la derivada discreta de primer orden toma valores dentro de un umbral acotado, aproximadamente entre -0,02 y 0,02. En el caso de la derivada de segundo orden, los valores típicos se ubican entre -0,003 y 0,003, como se muestra en la Figura 3. Estos umbrales se justifican porque, en condiciones de CC, la irradiancia solar presenta variaciones suaves y continuas a lo largo del día, determinadas principalmente por la geometría solar. En consecuencia, las derivadas de primer y segundo orden muestran valores bajos y estables, mientras que la presencia de nubes genera cambios abruptos que se traducen en derivadas de mayor magnitud. A partir de esta observación, se filtraron los valores que no se encontraban dentro de esos rangos.



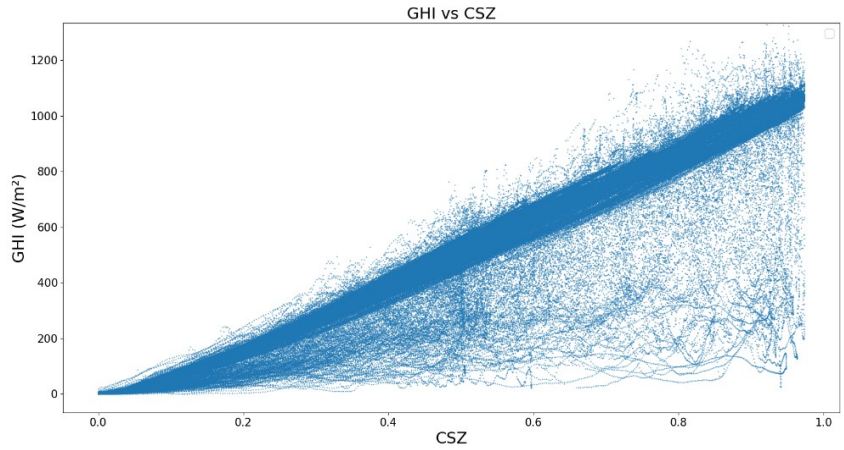
(a)



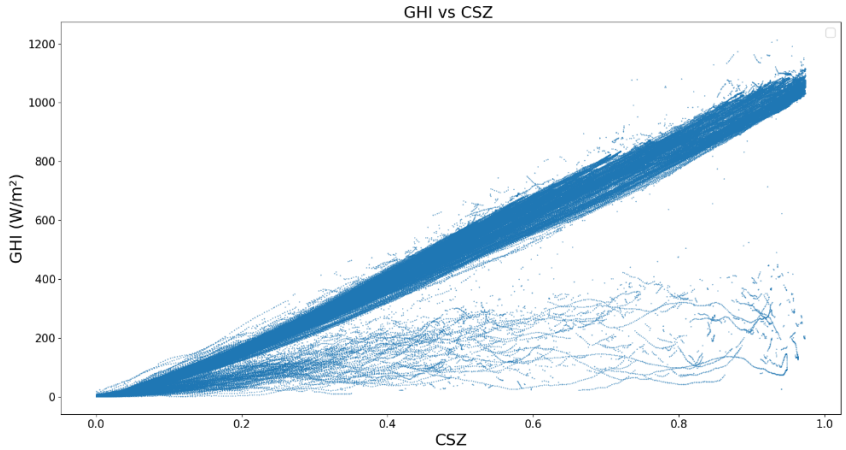
(b)

Figura 3: Día de CC donde se observan los umbrales definidos para las derivadas, a) Día 1-01-2022, b) Mismo día con zoom

Este primer filtro permitió eliminar momentos del día afectados por picos de nubosidad o caídas abruptas de la GHI. En las Figuras 4a y 4b se muestra la GHI en función del coseno del ángulo cenital (CSZ), antes y después de aplicar este criterio. Se utilizó este tipo de gráfico ya que, en condiciones de cielo claro, la relación de GHI vs CSZ describe una envolvente superior que representa un límite o cota superior de irradiancia posible para una determinada posición solar (sin tener en cuenta fenómenos de sobreirradiancia).



(a)



(b)

Figura 4: Comparación de datos de GHI en función del CSZ antes y después del primer filtro aplicado sobre derivadas para el año 2022. (a) Datos sin filtrar; (b) Datos filtrados aplicando el criterio de derivadas

Separación de muestras I: Clasificación mediante Gaussian Mixture

A los datos que superaron el primer filtro de derivadas se les aplicó un algoritmo de clustering basado en la clase Gaussian Mixture Models (GMM) de la librería sklearn.mixture (Pedregosa et al., 2011). Este algoritmo permite especificar la cantidad deseada de clústeres y se encarga de realizar distribuciones gaussianas a los datos y separarlos en grupos distintos. La Figura 5 muestra el resultado de la separación en dos clústeres.

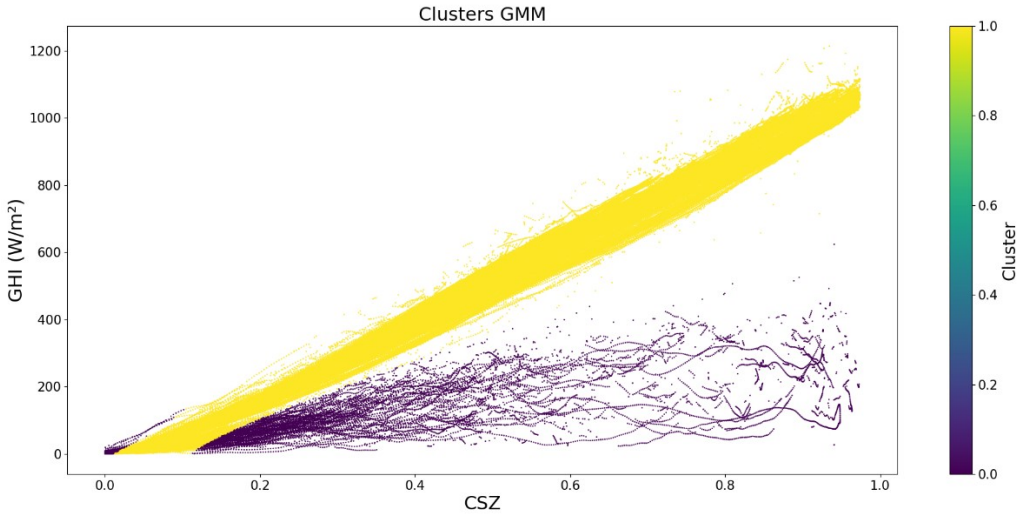


Figura 5: Separación de dos clústeres de datos GHI utilizando Gaussian Mixture sobre la base de datos del año 2022

De estos clústeres, se seleccionó el clúster correspondiente con la región superior, representada en color amarillo en la Figura 5, ya que se ajusta mejor a un perfil típico de cielo claro, actuando como cota superior para un determinado valor de GHI en función del ángulo cenital.

Separación de muestras II: Filtrado por densidad de vecinos

Luego de aplicar el segundo filtro, aún se observaban algunos subconjuntos de datos aislados que no representaban condiciones reales de CC. La mayoría de estos correspondían a situaciones de sobreirradiancia. Para realizar un tercer filtrado más minucioso se implementó un criterio basado en densidad de vecindad. Este método evaluó, sobre los datos previamente escalados y filtrados, cuántas muestras se encontraban dentro de un radio de 0.03 unidades normalizadas. Solo se conservaron aquellos puntos que tenían al menos 20 vecinos dentro de dicho radio, este criterio también fue realizado por inspección visual.

Este paso permitió eliminar subconjuntos de puntos aislados y mejorar la robustez del conjunto final de muestras representativas de CC. En la Figura 6 se muestran los resultados obtenidos.

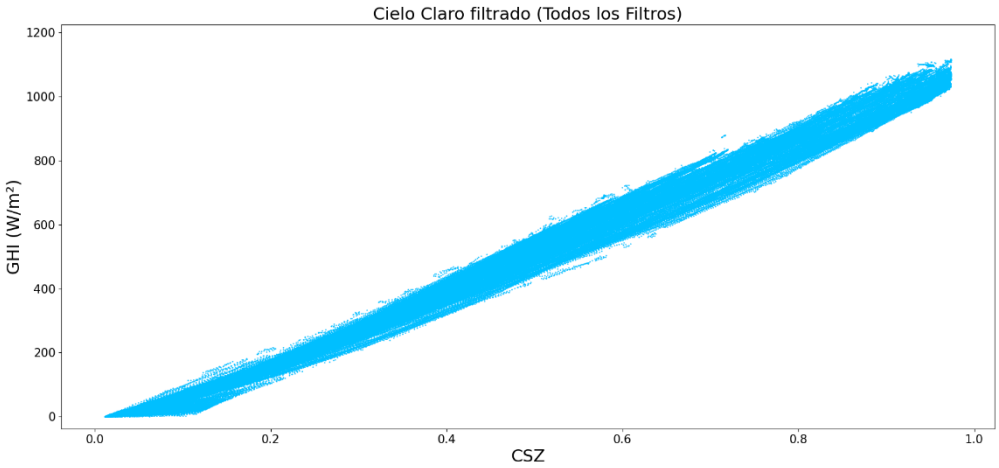


Figura 6: Conjunto final de muestras de cielo claro luego del filtrado basado en densidad de vecinos tomando un radio de 0.03 para el año 2022

Modelo de cielo claro local

Una vez obtenido el detector de muestras de CC, se procedió al desarrollo e implantación de un modelo empírico de cielo claro ajustado mediante regresión no lineal univariada. El modelo representa la GHI bajo condiciones de cielo claro (GHIcc) como una función potencia de la GHI en el tope de la atmósfera (TOA), dado por la siguiente ecuación:

$$GHI_{cc} = a * (TOA)^b \quad (1)$$

donde a y b son parámetros ajustados a partir de las muestras clasificadas como cielo claro.

Se utilizó el año 2022 como conjunto de ajuste para obtener los parámetros a y b del modelo, mientras que el año 2023 se utilizó como conjunto de testeo. La validación se realizó de manera local mediante la comparación del modelo de potencia con el modelo McClear. De forma análoga, se validó también el modelo ARGP-v2 frente a McClear, y finalmente se compararon las métricas de ambas validaciones.

RESULTADOS

Análisis del desempeño del detector de muestras de CC

Las muestras de CC obtenidas fueron comparadas con las identificadas por el algoritmo RENO-2016 como referencia, alimentado con mediciones de GHI y el modelo McClear de CC.

La comparación entre ambos detectores se realizó mediante una matriz de confusión (agregando además un nuevo filtro de altura solar de 7° para no tener en cuenta la noche (o sombreamiento al amanecer o atardecer) como se muestra en la Figura 7.

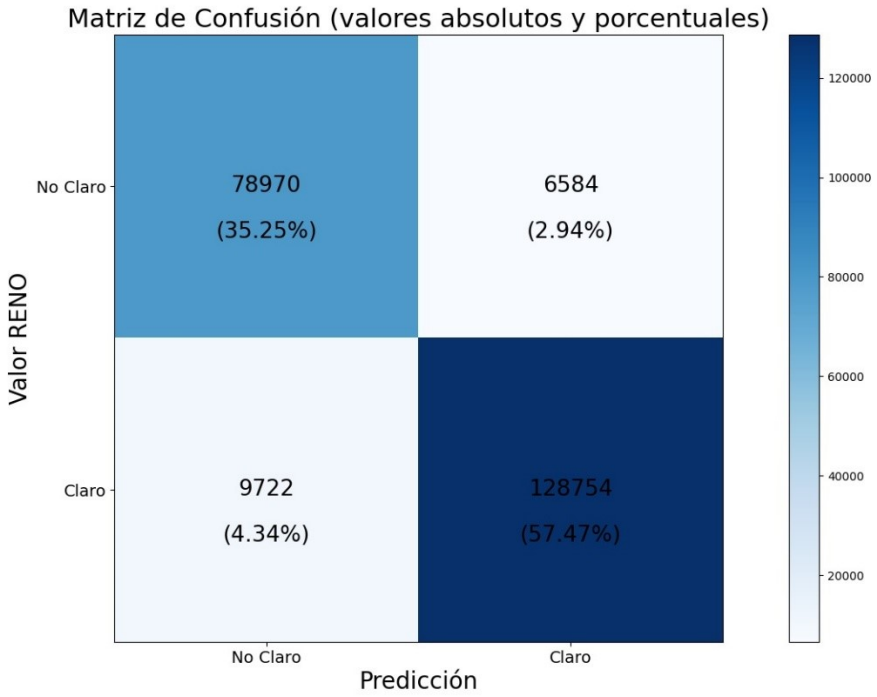
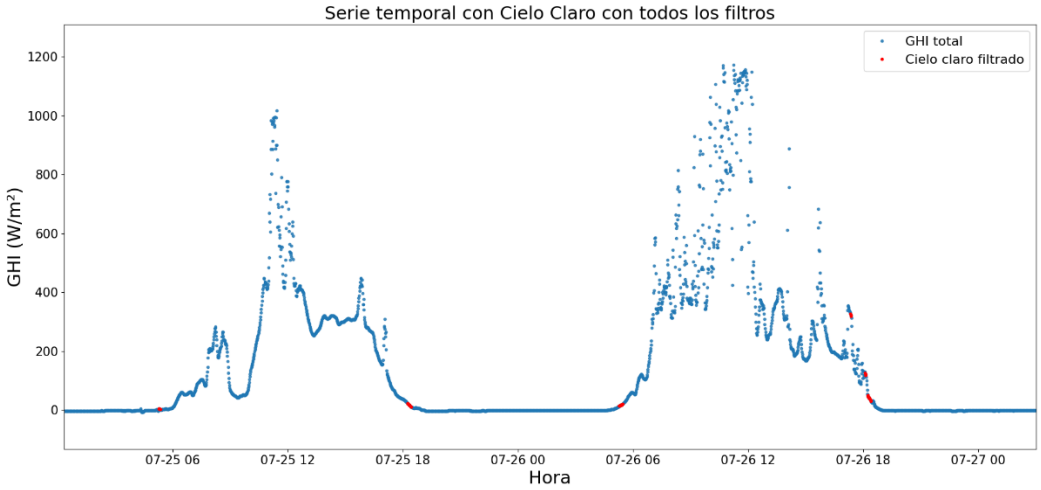


Figura 7: Matriz de confusión de la comparación entre las muestras de CC filtradas con el detector desarrollado y RENO-2016. La escala indica el número de datos

Se observa que el detector propuesto presenta una coincidencia superior al 90% respecto al algoritmo de RENO-2016, lo cual representa un buen indicador de desempeño y valida su utilidad como herramienta para la clasificación de muestras de CC. Además, ofrece posibilidades de mejoras y adaptación a otros contextos mediante el ajuste de sus parámetros.

En las Figuras 8a y 8b se muestran en azul los valores medidos de GHI en el sitio, mientras que en rojo se representan los valores que el detector clasifica como potenciales muestras de CC. En el caso de días nublados (Figura 8b), se aprecia que el detector filtra adecuadamente los datos atenuados por la presencia de nubosidad. En contraste, para un día representativo de CC (Figura 8a), el detector logra identificar la mayor parte de los valores como pertenecientes a condiciones de cielo claro.

(a)



(b)

Figura 8: Comportamiento del detector de muestras de CC en distintos tipos de nubosidades, (a) Día correspondiente a CC, (b) Días atenuados por nubosidad

Modelo de cielo claro local

Las muestras de CC obtenidas para el año 2022 se utilizaron para ajustar el modelo de CC de la Ec. (1). Así, se determinaron los parámetros presentados en la Tabla 1.

Tabla 1: Parámetros que mejor ajustan al modelo de potencia para DRA

Parámetros	Valor
a	0,208
b	1,193

Se utilizó el modelo de cielo claro McClear como modelo patrón para validar este modelo tanto en el ajuste como en el testeo y además se comparó con las métricas obtenidas de las validaciones con el modelo ARGP-v2 en el mismo periodo de ajuste y testeo. Los resultados de estas métricas se muestran a continuación en la siguiente Tabla 2. Las métricas utilizadas son el error cuadrático medio porcentual (rRMSE), el coeficiente de Pearson (corr) y el sesgo medio porcentual (rMBD). La reducción del rRMSE en la validación del modelo de potencia respecto al ARGP-v2 implica una mejora en la precisión de las estimaciones de GHIcc.

Tabla 2: Métricas de los modelos de cielo claro a escala minutal con respecto al modelo McClear en períodos de 2022 y 2023

Modelo	**Métrica**	**Ajuste (datos 2022)**	**Testeo (datos 2023)**
Modelo de potencia	rRMSE (%)	8,00	6,54
	rMBD (%)	2,54	1,91
	Corr	0,99	0,99
ARGP-v2	rRMSE (%)	13,52	12,03
	rMBD (%)	7,90	7,24
	Corr	0,99	0,99

La Figura 9 muestra el comportamiento de los modelos, incluido McClear, frente a datos medidos. Se observa que los modelos se encuentran en fase y que el modelo de potencia (en la gráfica abreviado como modelo exp) ajusta mejor que el modelo ARGP-v2.

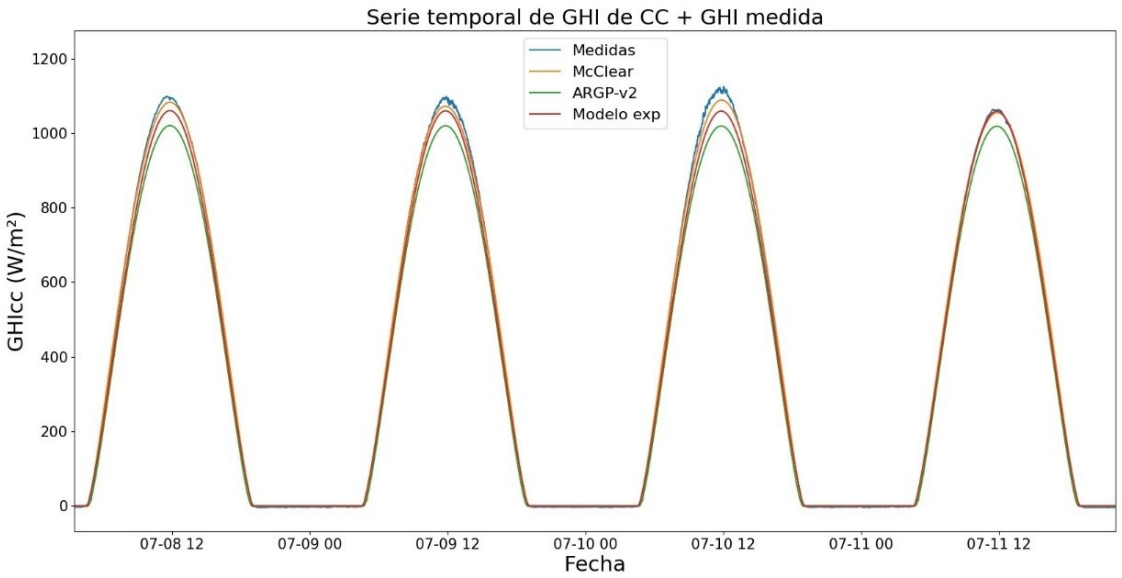


Figura 9: Comportamiento de los modelos de CC en la serie temporal para DRA 2023 (Testeo)

CONCLUSIONES

Tomando el modelo RENO-2016 como referencia, se observa que el detector desarrollado para condición de CC presenta una coincidencia superior al 90% para DRA, utilizando datos del año 2022.

Se observa que las técnicas no supervisadas utilizadas permiten clasificar muestras de cielo claro con buen desempeño, sin requerir un modelo previo como referencia, reduciendo el riesgo de validaciones que puedan estar sesgadas si dicho modelo no representa las condiciones de CC del sitio.

El detector de muestras de CC, además, presenta la ventaja de ser optimizable, ya que cuenta con parámetros ajustables. Esto permite adaptarlo a otros sitios en caso de que se requiera una calibración más precisa.

Dado el bajo costo computacional y la ventaja de ser un algoritmo no supervisado y optimizable presenta una buena alternativa al algoritmo de RENO-2016 que necesita un modelo de CC como entrada para poder hacer la respectiva detección.

Se construyó un modelo de cielo claro local mediante regresión no lineal univariada de tipo potencia, ajustado a partir de las muestras de CC del año 2022. Este modelo resultó empírico y efectivo, mostrando métricas de desempeño (respecto a McClear) superiores a las del modelo ARGP-v2. Asimismo, en el año de testeo se observó que, independientemente del modelo considerado, las métricas fueron mejores.

Para investigaciones futuras se podría realizar un análisis de sensibilidad sobre los umbrales utilizados en las derivadas y sobre la cantidad de vecinos considerados en el filtro de densidad de vecindad. Asimismo, se plantearía incorporar una etapa de control iterativa: utilizar el modelo obtenido de CC para retomar el proceso desde el paso inicial y generar nuevas muestras de CC, tomando a este modelo como referencia y calculando dichas métricas a partir de comparaciones punto a punto o mediante ventanas temporales de la GHI. Adicionalmente, se podría aplicar tanto el detector como el modelo desarrollado en este trabajo en otros sitios, así como llevar a cabo validaciones en regiones con características climáticas diferentes o durante días en los que se cuenten con diferentes niveles de cargas de aerosoles.

REFERENCIAS

Antonanzas-Torres, F., Urraca, R., Polo, J., Perpiñan-Lamigueiro, O. y R, E. (2019). Clear sky solar irradiance models: A review of seventy models. Renewable and Sustainable Energy Reviews, 107, 374-387. doi:https://doi.org/10.1016/j.rser.2019.02.032.

Developers. Gaussian mixture models en scikit-learn. Dirección URL: <https://scikit-learn.org/stable/modules/mixture.html> [consulta: agosto de 2025]

Engerer, N. y Mills, F. (2015). Validating nine clear sky radiation models in Australia. Solar Energy, 120, 9-24. doi:https://doi.org/10.1016/j.solener.2015.06.044.

Gueymard C. A (2008). REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation. Validation with a benchmark dataset. Solar Energy 82(3) 272-85.

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. Solar Energy, 86(8), 2145–2169. https://doi.org/10.1016/j.solener.2011.11.011

Laguarda, A. y Abal, G. (2017). Clear-sky broadband irradiance: first model assessment in Uruguay. Proceedings of Solar World Congress, 29, 10-2.

Laguarda, A., Giacosa, G., Alonso Suárez, R. y Abal, G. (2020). Performance of the site-adapted CAMS database and locally adjusted cloud index models for estimating global solar horizontal irradiation over the Pampa Húmeda. Solar Energy, 199, 295 - 307.

Ledesma, R. D., Salazar, G. A., y Castro Vilela, O. d. (2022). ARGP-v2 un modelo práctico para la estimación de irradiancia global horizontal en condiciones de cielo claro para sitios de altura. Avances en Energías Renovables y Medio Ambiente, 26.

Lefevre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., . . . Arola, A. (2013). McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. Atmospheric Measurement Techniques, 6, 2403-2418.

National Oceanic and Atmospheric Administration. (2025). Surface Radiation Budget Network (SURFRAD). NOAA Global Monitoring Laboratory. Dirección URL: <https://gml.noaa.gov/grad/surfrad/> [consulta: julio de 2025].

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . P. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

Peel, M. C., Finlayson, B. L., y McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. Hydrology and Earth System sciences, 11, 1633-1644.

Reno, M. J., y Hansen, C. W. (2016). Identification of periods of clear sky irradiance in time series of GHI measurements. Renewable Energy, 90, 520-531. doi: https://doi.org/10.1016/j.renene.2015.12.031.

Rivera, N., Salazar, G., y Laguarda, A. (2024). Análisis de la calidad de mediciones de irradiancia solar en plano horizontal (GHI) y evaluación de los modelos de GHI en condiciones de cielo claro para dos sitios en la provincia de salta. (A. A. Ambiente, Ed.) Actas del Congreso de ASADES 2024, 249-260. Obtenido de https://asades.org.ar/wp-content/uploads/2025/03/ACTAS-2024.pdf.

Russo, P., Laguarda, A., Abal, G. y Piccioli, I. (2022). Performance of the REST2 model for 1-minute clear-sky solar irradiance with MERRA-2 atmospheric inputs. Anais Congresso Brasileiro de Energia Solar-CBENS, 1-9. doi:10.59627/cbens.2022.1100.

IDENTIFICATION OF CLEAR-SKY GLOBAL HORIZONTAL IRRADIANCE SAMPLES USING UNSUPERVISED LEARNING

ABSTRACT: This work presents an approach for detecting clear-sky samples from the analysis of Global Horizontal Irradiance (GHI) time series using unsupervised machine learning techniques. In addition, it proposes a local clear-sky model fitted through a power-law regression on top-of-atmosphere (TOA) irradiance, using the detected samples. The detector was developed based on first- and second-order derivative thresholds, unsupervised clustering using the Gaussian Mixture Models (GMM) algorithm, and neighborhood density analysis, applied to a GHI time series. The method was tested on 1-minute GHI data from the Desert Rock site (USA), obtained from the SURFRAD network. The

detected samples were compared to those identified by the RENO-2016 algorithm, used as a reference, showing over 90% agreement. The resulting clear-sky model achieved relative RMSE values below 10% compared to the McClear model.

Keywords: global horizontal irradiance, clear-sky, clustering, clear-sky models