

APLICACIÓN DE UN MODELO DE MACHINE LEARNING PARA CLASIFICAR LA NUBOSIDAD A PARTIR DE DATOS DE IRRADIANCIA GLOBAL HORIZONTAL

Anabela Lusi^{1,2}, Facundo Orte^{1,2}, Elian Wolfram³, José Ignacio Orlando⁴

¹Departamento de Investigaciones en Láseres y Aplicaciones (DEILAP), UNIDEF (CITEDEF-CONICET), IRL-IFAECI-CNRS-3351, Villa Martelli (B1603), Buenos Aires, Argentina.

²Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290 (C1425FQB) CABA, Argentina

³Servicio Meteorológico Nacional (SMN), Av. Dorrego 4019 (C1425GBE), CABA, Argentina

⁴Yatiris Group, Instituto PLADEMA, CONICET, UNCPBA, Tandil (B7000), Argentina
e-mail: anabelalusi@gmail.com

RESUMEN: La radiación solar que alcanza la superficie terrestre está modulada por diversos factores, siendo la nubosidad el de mayor impacto y el que presenta las variaciones más rápidas. Las diferentes clases o tipos de nubes pueden generar patrones característicos en los niveles de irradiancia medidos, los cuales pueden aprovecharse para determinar la clasificación de la propia nubosidad. En este estudio se aplica un modelo de machine learning sobre una base de datos de GHI y un modelo de cielo claro, para clasificar la nubosidad en el sitio de Villa Martelli perteneciente a la red Saver-Net, durante los años 2015 a 2024. Se encontró que la mayor frecuencia relativa corresponde a situaciones de cielo despejado, seguida por la clase mezcla, mientras que la clase de nubes menos frecuente fue Stratocumulus-Stratus-Altocumulus. Esta distribución puede estar relacionada con la precisión del modelo para discriminar cada grupo. Este trabajo representa un avance en la clasificación de nubosidad y aporta información valiosa para estudios orientados a evaluar tendencias y cambios a largo plazo.

Palabras clave: clasificación de nubosidad, GHI, y aprendizaje supervisado.

INTRODUCCIÓN

Las nubes desempeñan un papel fundamental en el balance energético de la Tierra, ya que modulan tanto la radiación solar entrante como la radiación infrarroja saliente (Stephens y L'Ecuyer, 2015). Sus características, como la opacidad, el desarrollo vertical o la altura de la base y el tope, determinan en gran medida el tipo de nube y, con ello, su efecto sobre la cantidad de radiación que alcanza la superficie terrestre. Este efecto se produce a través de procesos de reflexión, absorción y dispersión de la radiación por parte de las partículas que componen la nube, y varía según factores como su volumen, forma, espesor, composición y cobertura (Deneke et al., 2008; Kokhanovsky, 2004; Ododo et al., 1996; Orsini et al., 2002). Por esta razón, la presencia y el movimiento de las nubes generan variaciones significativas, incluso a muy corto plazo, en los valores de Irradiancia Horizontal Global (GHI) medidos en superficie, con implicancias directas en campos como la meteorología, la predicción de energía solar y la gestión de sistemas eléctricos.

La Organización Meteorológica Mundial (OMM) (WMO, 2017) clasifica las nubes en diez géneros principales, agrupados en niveles altos, medios y bajos según la altitud a la que se observan. Esta clasificación es tradicionalmente realizada por observadores meteorológicos, a partir de la apariencia y distribución de las nubes en el cielo (octas). Sin embargo, este método resulta costoso y presenta un



grado inevitable de subjetividad. Dadas las dificultades asociadas con la clasificación visual, los investigadores han buscado desarrollar metodologías más objetivas y rentables para la detección de nubes. Como alternativa, se han desarrollado distintos métodos que utilizan otras herramientas, como datos satelitales (Escrig et al., 2013; Liang y Yuan, 2016; Nespoli et al., 2022), cámaras de todo cielo (Kassianov et al., 2005; Long et al., 2006) o parámetros meteorológicos medidos en superficie (Dürr y Philipona, 2004).

En este contexto, surge la posibilidad de utilizar series temporales de GHI, un parámetro de medición sencilla y de mayor disponibilidad, para inferir el tipo de nubosidad presente en la atmósfera. A fines de la década de 1990, Duchon y O'Malley (1999) (Duchon y O'Malley, 1999) propusieron un método para discriminar el tipo de nube basándose en mediciones de irradiancia de un piranómetro y análisis estadístico. Sin embargo, este enfoque mostró una precisión entre las estimaciones del piranómetro con las observaciones desde tierra aproximadamente el 45% de las veces. Estas diferencias fueron atribuidas a que el piranómetro se enfoca en las nubes que cruzan la trayectoria del Sol y a la susceptibilidad de la interferencia de aerosoles.

Los avances recientes en técnicas de machine learning (ML) o aprendizaje automático, ofrecen un marco más robusto para abordar este desafío, como se ha evidenciado en diversas aplicaciones dentro del campo de las Ciencias de la Tierra (Lary et al., 2018). Estos modelos tienen la ventaja de ser lo suficientemente flexibles para aplicarse en diferentes escenarios, ya que son capaces de aprender automáticamente patrones y relaciones entre sus variables de entrada y salida. Diversos autores han utilizado estas técnicas para la detección de cielos despejados (M. Liu et al., 2021; Mommert, 2020). Otros investigadores han utilizado este método para detectar diferentes condiciones de nubosidad (Lee et al., 2004; Taravat et al., 2015).

Lusi et. al (2024) (Lusi et al., 2024) desarrollaron un modelo de clasificación de nubosidad basado en el algoritmo XGBoost, utilizando únicamente mediciones de GHI y un modelo de cielo despejado. Inspirado en el enfoque de Duchon y O'Malley (1999), el modelo fue entrenado con datos de GHI del año 2019 en Villa Martelli, Buenos Aires, mientras que las etiquetas de referencia provinieron de la clasificación de imágenes todo cielo, realizada por observadores meteorológicos del Servicio Meteorológico Nacional (SMN) de Argentina (<https://www.smn.gob.ar/>). El modelo alcanzó un accuracy de 0,88 y un coeficiente kappa de Cohen de 0,87 en el conjunto de test. Dicho modelo se encuentra disponible públicamente para ser usado.

En el presente estudio, se implementa el modelo de ML desarrollado por Lusi et. al (2024) para clasificar la nubosidad en el sitio de la red Saver-Net de Villa Martelli, Buenos Aires, y obtener estadísticas de la presencia de las diferentes clases de nubes durante los años 2015 a 2024. El modelo, previamente entrenado y validado en este sitio, se aplica aquí con el fin de generar dichas estadísticas sobre un periodo más largo.

MATERIALES Y MÉTODOS

Zona de estudio

La red Sudamericana de Gestión de Riesgos Ambientales (SAVER-Net, por sus siglas en inglés), es una red de monitoreo remoto de la atmósfera, que fue desplegada entre 2013 y 2018 en el marco de un proyecto conjunto entre Argentina, Chile y Japón (Orte et al., 2022). Incluye ocho sitios de medición distribuidos en Argentina, equipados con diversos instrumentos para la caracterización de la atmósfera. Entre ellos, se incluyen piranómetros Kipp&Zonen CMP-21 (285-2800 nm). Los piranómetros registran la GHI en el rango de onda corta sobre un plano horizontal, con una resolución temporal de un minuto. El presente trabajo se llevó a cabo en el sitio de Villa Martelli (VMA) (34,58°S, 58,48°O, 25 m s.n.m.), uno de los nodos de la red, con datos adquiridos durante los años 2015 a 2024.

El conjunto de datos de GHI de la estación fue sometido a un control de calidad. Dentro de los procedimientos aplicados, se incluye la exclusión de medidas de altura solar baja ($<7^\circ$), la inspección visual de las series para identificar periodos defectuosos y el análisis conjunto con otras características,

como el índice de claridad (kt), para poder detectar la presencia de sombras o errores sistemáticos y descartar el periodo en caso de que el registro esté defectuoso.

Modelo de machine learning utilizado para la clasificación

El modelo de clasificación de nubosidad que se utilizó en este trabajo fue desarrollado por Lusi et al. (2024) y es un clasificador supervisado basado en XGBoost con función de pérdida softmax. El modelo es capaz de discriminar seis clases de nubosidad, a partir de la combinación de series temporales de mediciones de GHI y de un modelo de irradiancia de cielo claro (GHI_{CS}). En este trabajo se utilizó el modelo McClear (Lefèvre et al., 2013) para obtener GHI_{CS} . El modelo de ML determina la clasificación de la nubosidad en ventanas temporales de 33 minutos de duración. Una descripción detallada del entrenamiento, validación, testeo y desempeño del modelo puede encontrarse en Lusi et al. (2024).

La Figura 1 muestra las siete categorías de nubosidad (seis categorías más la clase mezcla) capaces de discriminar, agrupadas en dos grandes grupos denominados “not-mixed” (no mezcla) y “mixed” (mezcla). Estas categorías son: Cb-Ns (Cumulonimbus y Nimbostratus), Ac-Cc (Alto cumulus y Cirrocumulus), Cu (Cumulus), Sc-St-As (Stratocumulus, Stratus y Alto cumulus), Ci-Cs (Cirrus y Cirrostratus) y Clear-sky (cielo despejado). Si bien la clase “clear sky” no es una clase de nubosidad estrictamente, se incluye como una clase más dentro de las clasificaciones.

Cabe destacar que la clase “mezcla” se asigna cuando en una ventana temporal coexisten más de uno de los grupos de la clase “no mezcla”. Por ejemplo, si en una ventana se detectan simultáneamente Cb-Ns y Cu, esa ventana se clasifica como mezcla.

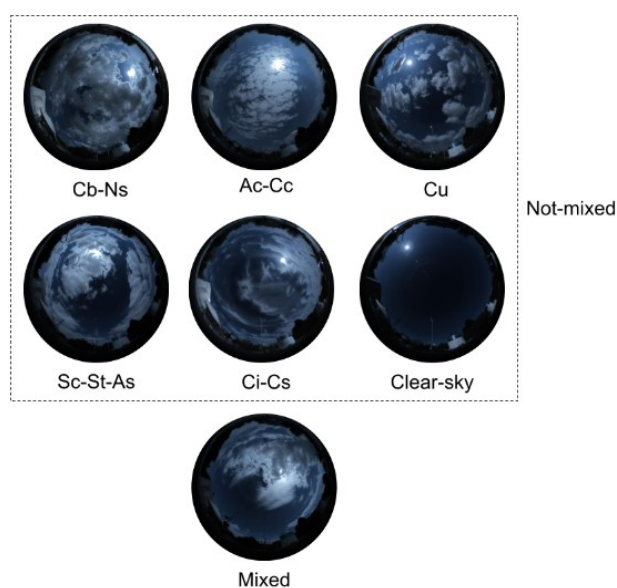


Figura 1: Clasificaciones posibles de nubosidad a partir del modelo de machine learning (Lusi et al., 2024)

Con el fin de obtener las clasificaciones dentro de los siete grupos posibles, se siguió la siguiente metodología. En primer lugar, se calcularon diferentes características (*features*) para intervalos de 33 minutos de irradiancia definidas en Lusi et al. (2024). Estas *features*, son utilizadas como indicadores de la variabilidad de la radiación solar presente, con las cuales fue entrenado el modelo de ML. Se determinaron a través de las siguientes estimaciones dentro de las ventanas temporales de 33 minutos:

- Desviación estándar del índice de variabilidad (Stein et al., 2012).
- Desviación cuadrática media entre GHI y GHI_{CS} .
- Desviación estándar, media y valor máximo del índice de claridad modificado (Perez et al., 1990).
- Asimetría y curtosis de la serie del índice de claridad modificado.
- Variabilidad media de los cambios en el índice de claridad modificado.

Luego de calcular y normalizar las *features*, se utilizó el modelo de ML disponible públicamente en GitHub (<https://github.com/anabelalusi/cloud-classification>) para obtener las clasificaciones de nubosidad. La Figura 2 presenta un esquema de la metodología empleada.

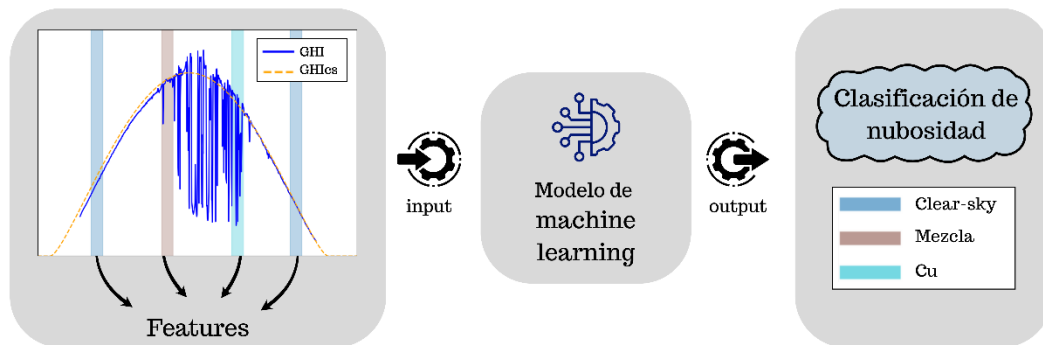


Figura 2: Metodología para la aplicación del modelo de ML para clasificación de nubosidad. El modelo se alimenta con ocho features extraídas de ventanas temporales de 33 minutos, que utilizan únicamente GHI y GHI_{cs}. La salida corresponde a la clasificación del periodo en una de las siete clases de nubosidad.

Es importante destacar que el modelo clasifica en seis grupos principales de nubosidad. Adicionalmente, se incorporó en el post procesamiento la clase “mezcla”, que corresponde a la presencia simultánea de más de un grupo de nubosidad (por ejemplo, si se detecta Ac-Cc y Cu en la ventana de tiempo, se etiqueta como mezcla). Para identificar esta categoría, se calcula la entropía de las probabilidades de clasificación, cuando la entropía supera el umbral de 0,7, la etiqueta de salida se define como “mezcla”, siguiendo lo indicado en Lusi et al. (2024).

RESULTADOS Y DISCUSIÓN

Clasificación de nubosidad

El modelo asignó una clasificación de nubosidad a la serie temporal de GHI registrada en VMA durante el periodo 2015-2024, evaluando ventanas consecutivas de 33 minutos, siempre y cuando la serie esté completa.

La Figura 3 muestra un caso de estudio correspondiente al 28 de enero de 2019. Las líneas verticales de colores sobre la gráfica representan las ventanas clasificadas en cada categoría de nubosidad. En la parte inferior, tres imágenes captadas por la cámara de observación del cielo ilustran momentos específicos del día, coincidiendo con las clasificaciones del modelo.

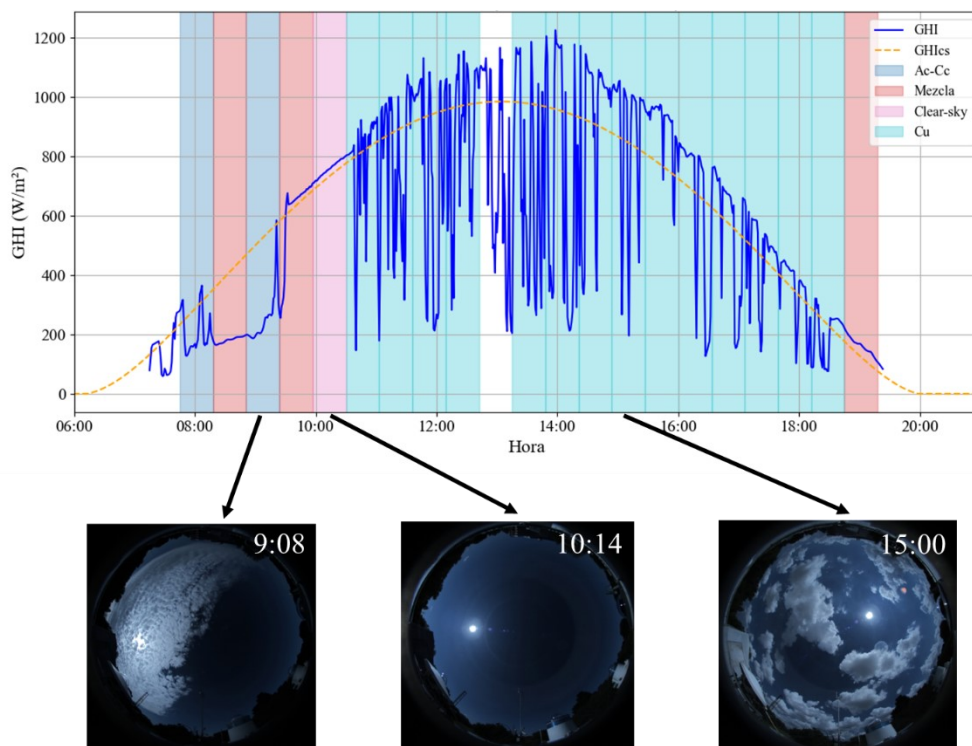


Figura 3: Clasificación de nubosidad para el 28 de enero de 2019 en VMA. Las líneas verticales de colores indican las ventanas de 33 minutos clasificadas por el modelo en distintas categorías de nubosidad: Ac-Cc, Mezcla, Clear-Sky y Cu. En la parte inferior se muestran imágenes de la cámara de observación del cielo que ilustran tres ventanas específicas.

A las 9:08h, la ventana fue etiquetada como Ac-Cc. A las 10:14h, la alta correspondencia entre las curvas GHI y GHI_{CS} respalda la clasificación como cielo despejado (Clear-sky). Finalmente, a las 15:00h, la imagen evidencia cúmulos bien definidos, en concordancia con la clasificación Cu asignada por el modelo.

La Tabla 1 resume la cantidad de ventanas de 33 minutos clasificadas para cada tipo de nubosidad, así como el porcentaje que representa cada clase durante los años 2015 a 2024.

Tabla 1: Resumen de las clasificaciones de nubosidad en VMA durante el periodo 2015 a 2024.

Clase	Número de ventanas	Frecuencia relativa (%)
Clear-sky	22900	37,2
Mezcla	20118	32,7
Cb-Ns	7213	11,7
Ci-Cs	5691	9,3
Ac-Cc	2360	3,8
Cu	1735	2,8
Sc-St-As	1504	2,4

A partir de los resultados se observa que la clase predominante es Clear-sky, condición para la cual el modelo también presenta mayor precisión (Lusi et al., 2024). Esto sugiere que el modelo podría ser una herramienta útil para aplicaciones relacionadas con la radiación solar, donde la identificación de períodos de cielo despejado es fundamental.

En segundo lugar, la clase Mezcla muestra una alta frecuencia relativa, indicando que en la mayoría de las ventanas de 33 minutos se detectó la presencia simultánea de más de uno de los seis grupos “not mixed”. Por su parte, la clase Sc-St-As presenta la frecuencia más baja, lo que indica una menor presencia durante el periodo analizado. Sin embargo, es importante destacar que esta es también la clase con mayor dificultad para la discriminación por parte del modelo de ML (Lusi et al., 2024).

Distribución anual de las clases de nubosidad

La Figura 4 presenta la frecuencia relativa anual de las distintas clases de nubosidad asignadas por el modelo entre 2015 y 2024. La frecuencia relativa se calculó considerando el porcentaje de cada clase sobre la cantidad total de muestras analizadas en un año particular. Se destacan las categorías Clear-sky y Cu en la mayoría de los años, mientras que las clases Ci-Cs y Ac-Cc, Cb-Ns muestran porcentajes más bajos y relativamente estables a lo largo del periodo analizado.

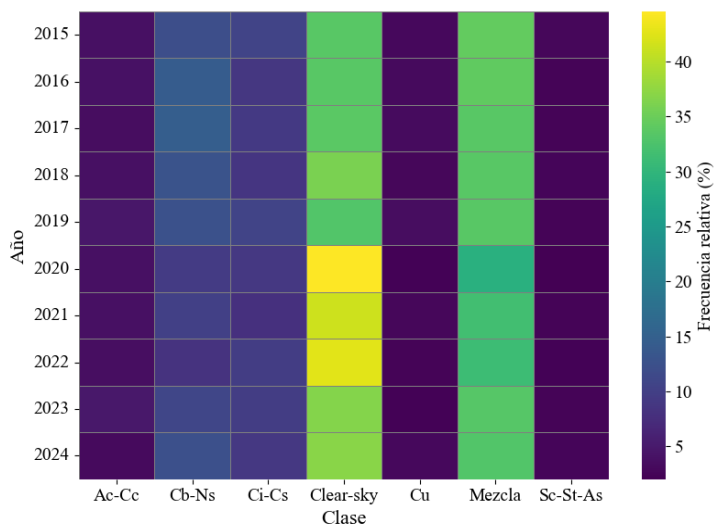


Figura 4: Distribución porcentual anual de las clases de nubosidad observadas entre 2015 y 2024. Cada valor representa el porcentaje relativo al total de ventanas de observación de cada año.

Esta representación permite detectar años con valores atípicos. Por ejemplo, los años 2020 y 2022, presentan una mayor proporción de clasificaciones de cielo despejado en comparación con el resto de los años. Este patrón podría asociarse a condiciones meteorológicas particulares o a variaciones en los forzantes climáticos de escala regional. Por otro lado, la consistencia observada en algunas clases sugiere la presencia de un régimen nuboso persistente, mientras que otras evidencian una mayor variabilidad interanual.

Distribución estacional de las clases de nubosidad

La Figura 5 muestra la distribución porcentual estacional de las diferentes clases de nubosidad asignadas por el modelo en el sitio durante el período analizado.

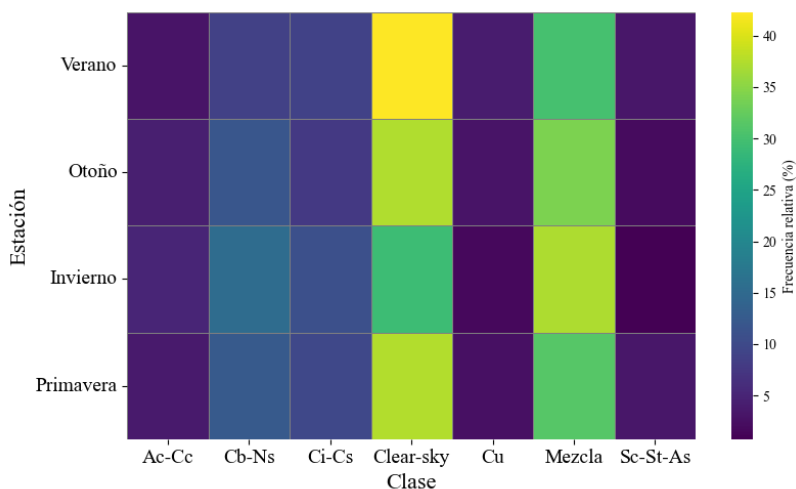


Figura 5: Distribución porcentual por estación de las clases de nubosidad observadas entre 2015 y 2024. Cada valor representa el porcentaje relativo al total de ventanas de observación de cada estación.

Se observa que durante el verano predomina la clase Clear-sky, representando el 42,3% de las clasificaciones, seguida por la clase Mezcla con un 30% y la clase Ci-Cs con un 8,9%. En otoño, Clear-sky y Mezcla mantienen una presencia similar, con 37,2% y 34,0% respectivamente, mientras que las clases Cb-Ns y Ci-Cs aumentan su presencia alcanzando 11,8% y 7,7%. Durante el invierno se observa un incremento en las clases Mezcla (37,1%) y Cb-Ns (15,3%), acompañado de una reducción de la clase de Clear-sky al 29,3%. Además, Ci-Cs también incrementa su frecuencia a un 10,8%. En primavera, la distribución vuelve a ser similar a la predicha en otoño, con Clear-sky en 37,5%, Mezcla en 31,3%, y las clases Cb-Ns y Ci-Cs con 12,3% y 9,7%, respectivamente. Las clases de menor frecuencia, como Ac-Cc y Cu, mantienen porcentajes inferiores al 6% durante todas las estaciones.

CONCLUSIONES

Este trabajo muestra la utilidad de un modelo de ML para clasificar la nubosidad a partir de series temporales de GHI y de un modelo de cielo claro en ventanas de 33 minutos. La aplicación del modelo sobre datos de VMA entre 2015 y 2024 permitió identificar las diferentes clases de nubosidad, resultando la clase de cielo despejado la más frecuente durante el periodo analizado.

La distribución anual de las clases nubosas evidencia patrones consistentes a lo largo de los años, con variaciones que podrían estar asociadas a condiciones meteorológicas particulares, como se observó en los años 2020 y 2022 con un mayor porcentaje de cielo claro. La presencia y frecuencia relativa de clases menos comunes, aunque algunas más difíciles de discriminar por el modelo, también aportan información valiosa sobre la variabilidad nubosa local.

Este tipo de análisis permite caracterizar la climatología nubosa, y además establece una base comparativa para futuras investigaciones orientadas a evaluar tendencias y posibles cambios de largo plazo en la nubosidad.

FUENTES DE FINANCIAMIENTO

Se retira este contenido en la versión actual para respetar los lineamientos de revisión doble-ciego.

AGRADECIMIENTOS

Se retira este contenido en la versión actual para respetar los lineamientos de revisión doble-ciego.

REFERENCIAS

- Deneke, H. M., Feijt, A. J., y Roebeling, R. A. (2008). Estimating surface solar irradiance from METEOSAT SEVIRI-derived cloud properties. *Remote Sensing of Environment*, 112(6), 3131-3141. <https://doi.org/10.1016/j.rse.2008.03.012>
- Duchon, C. E., y O'Malley, M. S. (1999). Estimating Cloud Type from Pyranometer Observations. *Journal of Applied Meteorology and Climatology*, 38(1), 132-141. [https://doi.org/10.1175/1520-0450\(1999\)038<0132:ECTFPO>2.0.CO;2](https://doi.org/10.1175/1520-0450(1999)038<0132:ECTFPO>2.0.CO;2)
- Dürr, B., y Philipona, R. (2004). Automatic cloud amount detection by surface longwave downward radiation measurements. *Journal of Geophysical Research: Atmospheres*, 109(D5). <https://doi.org/10.1029/2003JD004182>
- Escrig, H., Batlles, F. J., Alonso, J., Baena, F. M., Bosch, J. L., Salbidegoitia, I. B., y Burgaleta, J. I. (2013). Cloud detection, classification and motion estimation using geostationary satellite imagery for cloud cover forecast. *Energy*, 55, 853-859. <https://doi.org/10.1016/j.energy.2013.01.054>
- Kassianov, E., Long, C., y Ovchinnikov, M. (2005). Cloud Sky Cover versus Cloud Fraction: Whole-Sky Simulations and Observations. *Journal of Applied Meteorology*, 44. <https://doi.org/10.1175/JAM-2184.1>
- Kokhanovsky, A. (2004). Optical properties of terrestrial clouds. *Earth-Science Reviews*, 64(3), 189-241. [https://doi.org/10.1016/S0012-8252\(03\)00042-4](https://doi.org/10.1016/S0012-8252(03)00042-4)

- Lary, D., Zewdie, G., Liu, X., Wu, D., Levetin, E., Allee, R., Malakar, N., Walker, A., Mussa, H., Mannino, A., y Aurin, D. (2018). Machine Learning Applications for Earth Observation (pp. 165-218). https://doi.org/10.1007/978-3-319-65633-5_8
- Lee, Y., Wahba, G., y Ackerman, S. A. (2004). Cloud Classification of Satellite Radiance Data by Multicategory Support Vector Machines. *Journal of Atmospheric and Oceanic Technology*, 21(2), 159-169. [https://doi.org/10.1175/1520-0426\(2004\)021<0159:CCOSRD>2.0.CO;2](https://doi.org/10.1175/1520-0426(2004)021<0159:CCOSRD>2.0.CO;2)
- Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., Wald, L., Schroedter-Homscheidt, M., Hoyer-Klick, C., Arola, A., Benedetti, A., Kaiser, J. W., y Morcrette, J.-J. (2013). McClear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmospheric Measurement Techniques*, 6(9), 2403-2418. <https://doi.org/10.5194/amt-6-2403-2013>
- Liang, X., y Yuan, C. (2016). Derivation of 3D cloud animation from geostationary satellite images. *Multimedia Tools and Applications*, 75(14), 8217-8237. <https://doi.org/10.1007/s11042-015-2738-7>
- Liu, M., Xia, X., Fu, D., y Zhang, J. (2021). Development and Validation of Machine-Learning Clear-Sky Detection Method Using 1-Min Irradiance Data and Sky Imagers at a Polluted Suburban Site, Xianghe. *Remote Sensing*, 13(18), 3763. <https://doi.org/10.3390/rs13183763>
- Liu, S., Duan, L., Zhang, Z., y Cao, X. (2019). Hierarchical Multimodal Fusion for Ground-Based Cloud Classification in Weather Station Networks. *IEEE Access*, 7, 85688-85695. <https://doi.org/10.1109/ACCESS.2019.2926092>
- Long, C., Sabburg, J., Calbó, J., y Pages, D. (2006). Retrieving Cloud Characteristics from Ground-Based Daytime Color All-Sky Images. *Journal of Atmospheric and Oceanic Technology*, 23. <https://doi.org/10.1175/JTECH1875.1>
- Loyber, P. J. (2022). Entrenamiento de un modelo de IA para el procesamiento de imágenes todo cielo y clasificación de nubes [Instituto Tecnológico de Buenos Aires, ITBA]. <https://ri.itba.edu.ar/handle/123456789/3982>
- Lusi, A. R., Orte, P. F., Wolfram, E., y Orlando, J. I. (2024). Cloud classification through machine learning and global horizontal irradiance data analysis. *Quarterly Journal of the Royal Meteorological Society*, 150(765), 5435-5451. <https://doi.org/10.1002/qj.4880>
- Mommert, M. (2020). Cloud Identification from All-sky Camera Data with Machine Learning. *The Astronomical Journal*, 159(4), 178. <https://doi.org/10.3847/1538-3881/ab744f>
- Nespoli, A., Niccolai, A., Ogliari, E., Perego, G., Collino, E., y Ronzio, D. (2022). Machine Learning techniques for solar irradiation nowcasting: Cloud type classification forecast through satellite data and imagery. *Applied Energy*, 305, 117834. <https://doi.org/10.1016/j.apenergy.2021.117834>
- Ododo, J. C., Agbakwuru, J. A., y Ogbu, F. A. (1996). Correlation of solar radiation with cloud cover and relative sunshine duration. *Energy Conversion and Management*, 37(10), 1555-1559. [https://doi.org/10.1016/0196-8904\(96\)86837-3](https://doi.org/10.1016/0196-8904(96)86837-3)
- Orsini, A., Tomasi, C., Calzolari, F., Nardino, M., Cacciari, A., y Georgiadis, T. (2002). Cloud cover classification through simultaneous ground-based measurements of solar and infrared radiation. *Atmospheric Research*, 61(4), 251.
- Orte, F., Wolfram, E., Luccini, E., D'Elia, R., Lusi, A., Pallotta, J., Nollas, F., Carmona, F., Papandrea, S., Cabezas, M. D., Benítez, G. C., y Mizuno, A. (2022). Red de Monitoreo de la Irradiancia Solar UV-Total en Argentina «Saver-Net». *Meteorológica*, 47(2). <https://doi.org/10.24215/1850468Xe016>
- Perez, R., Ineichen, P., Seals, R., y Zelenka, A. (1990). Making full use of the clearness index for parameterizing hourly insolation conditions. *Solar Energy*, 45(2), 111-114. [https://doi.org/10.1016/0038-092X\(90\)90036-C](https://doi.org/10.1016/0038-092X(90)90036-C)
- Stein, J. S., Hansen, C., y Reno, M. J. (2012, marzo 1). The Variability Index: A New and Novel Metric for Quantifying Irradiance and PV Output Variability. *World Renewable Energy Forum*, United States. <https://www.osti.gov/biblio/1068417>
- Stephens, G. L., y L'Ecuyer, T. (2015). The Earth's energy balance. *Atmospheric Research*, 166, 195-203. <https://doi.org/10.1016/j.atmosres.2015.06.024>
- Taravat, A., Del Frate, F., Cornaro, C., y Vergari, S. (2015). Neural Networks and Support Vector Machine Algorithms for Automatic Cloud Classification of Whole-Sky Ground-Based Images. *IEEE Geoscience and Remote Sensing Letters*, 12(3), 666-670. <https://doi.org/10.1109/LGRS.2014.2356616>
- WMO. (2017). Manual on the Observation of Clouds and Other Meteors – International Cloud Atlas: Vol. Basic Documents No. 2. <https://library.wmo.int/idurl/4/66297>

CLOUD CLASSIFICATION FROM GLOBAL HORIZONTAL IRRADIANCE USING A MACHINE LEARNING MODEL

ABSTRACT Solar radiation reaching the Earth's surface is modulated by various factors, with cloudiness being the most impactful factor, exhibiting the fastest variations. Different cloud classes or types can generate characteristic patterns in measured irradiance levels, which can be leveraged to determine cloud classification. In this study, a machine learning model was applied to a global horizontal irradiance (GHI) database and a clear-sky model to classify cloudiness at the Villa Martelli site, part of the Saver-Net network, from 2015 to 2024. The highest relative frequency corresponded to clear-sky conditions, followed by the mixed cloud class, whereas the least frequent class was the stratocumulus-stratus-alto-cumulus. This distribution may be related to the accuracy of the model in discriminating each group. This study represents an advance in cloud classification and provides valuable information for studies aimed at evaluating long-term trends and changes.

Keywords: cloud classification, GHI and supervised learning algorithms.