



ENANDES Project: Enhancing Adaptive Capacity of Andean Communities through Climate Services (ENANDES) (Chile, Colombia, Peru)

Activity: Energy mini-project on validation and calibration of the SOLEVAPP evaporation model for floating solar systems in Chile (Ref. 21397/2024-1.2 GS)

Deliverable N°4: Operational Manual: Evaporation Model for Floating Solar Panels Using Hydrometeorological Data and Machine Learning

History of draft preparation and review

V1.0- Initial draft by Hernán Alcayaga, Rodrigo Cáceres, and Andres Díaz (UDP), with the contribution of Juan Crespo Fuentes and Marjorie Cataldo Soto (DMC), Camila Vásquez Páez (MinEnergía), Hamid Bastani (WMO)	29/11/2025
V2.0 – Reviewed by Hamid Bastani (WMO)	29/01/2026
V3.0 – Reviewed by Hamid Bastani (WMO)	09/04/2026
V4.0 Reviewed by Raul Polato and Roberta Boscolo (WMO)	09/06/2026



WORLD
METEOROLOGICAL
ORGANIZATION



ADAPTATION FUND



udp UNIVERSIDAD
DIEGO PORTALES



— ENANDES —

Note:

This deliverable (N°4) is composed of the following three main components:

1. The present technical document, developed as an operational manual for the updated SOLEVAPP model.
2. The open-source Jupyter Notebook codes of the model, aligned with the present technical publication, available at: <https://gitlab.com/rodrigo.caceres1/enandes-solevapp/>
3. A recorded training on the methodology and the codes, aligned with the present technical publication and the Jupyter Notebook codes, available at: https://wmoomm-my.sharepoint.com/:f:/g/personal/fguerrero_wmo_int/IgATKnfXCDGET4Jt8lIdgSZgAdFJu169S8vBWGB9k6U_hvE?e=hTIHBz

Nota:

Este entregable se compone de los siguientes tres componentes principales:

1. El presente documento técnico, elaborado como manual operativo para la versión actualizada del modelo SOLEVAPP.
2. Los códigos de acceso abierto en Jupyter Notebook del modelo, alineados con la presente publicación técnica, disponibles en: <https://gitlab.com/rodrigo.caceres1/enandes-solevapp/>
3. Una capacitación grabada sobre la metodología y los códigos, alineada con la presente publicación técnica y los códigos en Jupyter Notebook, disponible en: https://wmoomm-my.sharepoint.com/:f:/g/personal/fguerrero_wmo_int/IgATKnfXCDGET4Jt8lIdgSZgAdFJu169S8vBWGB9k6U_hvE?e=hTIHBz

Manual Operativo: Modelo de evaporación para paneles solares flotantes utilizando datos hidrometeorológicos y aprendizaje automático

Autores: Rodrigo Cáceres González (UDP), Hernán Alcayaga (UDP), Andrés Díaz Alarcón (UDP)

Colaboradores: Juan Crespo Fuentes (DMC), Marjorie Cataldo Soto (DMC), Camila Vásquez Páez (MinEnergía), Fabian Bustos Olavarría (MinEnergía), Hamid Bastani (WMO)

Entidades participantes: Universidad Diego Portales (UDP), Ministerio de Energía de Chile, Dirección Meteorológica de Chile (DMC), Organización Meteorológica Mundial (WMO).

Coordinador técnico del proyecto: Hamid Bastani (WMO)

Financiamiento: Esta publicación fue producida en el marco de un Mini Proyecto de Energía de la Organización Meteorológica Mundial (WMO) desarrollado en Chile dentro del proyecto ENANDES (Enhancing Adaptive Capacity of Andean Communities through Climate Services). El proyecto ENANDES es financiado por el Fondo de Adaptación e implementado por la WMO en colaboración con los Servicios Meteorológicos e Hidrológicos Nacionales de Chile, Colombia y Perú, y el Centro Internacional para la Investigación del Fenómeno de El Niño (CIIFEN).

Contenido

Resumen	6
1. INTRODUCCIÓN	8
1.1. Objetivos del documento	8
1.2. Estudio de caso en Chile	9
2. DESCRIPCIÓN GENERAL DEL SISTEMA	11
2.1. Implementación inicial de SOLEVAPP	12
2.2. Porcentaje de cobertura	13
2.2.1. Definición del porcentaje de cobertura	13
2.2.2. Cálculo de las dos contribuciones a la evaporación	13
2.2.3. Cálculo de la potencial evaporación efectiva	14
2.2.4. Generación de resultados	14
2.3. Implementación de modelo ajustado por Machine Learning para SOLEVAPP ..	15
2.3.1. Paso 1: Cálculo físico de evaporación en superficie libre	15
2.3.2. Paso 2: Construcción de variables de entrada	16
2.3.3. Paso 3: Asignación automática a un clúster climático	16
2.3.4. Paso 4: Predicción ajustada de evaporación potencial en superficie libre y cubierta 16	
2.3.5. Paso 5: Combinación final mediante β	17
2.4. Visión general de los datos requeridos	17
3. ESTRUCTURA MODULAR DEL SOFTWARE DESARROLLADO	19
3.1. Estructura del flujo de trabajo	19
3.2. Flujo de operación: Parte A	20
3.2.1. Configuración inicial (Paso 0)	21
3.2.2. Parámetros del sitio (Paso 1)	21
3.2.3. Carga y preparación de datos (Paso 2)	22
3.2.4. Verificación y conversión de unidades (Paso 3)	23
3.2.5. Exploración de datos (Paso 4, opcional)	24
3.2.6. Clasificación climática del sitio, enrutador (Paso 5)	25
3.2.7. Simulación (Paso 6)	26
3.3. Reentrenamiento del modelo (uso avanzado): Parte B	29
3.3.1. Configuración del entrenamiento (Paso T1)	29

3.3.2.	Carga y preparación de datos (Paso T2).....	30
3.3.3.	Entrenamiento y guardado de modelos (Paso T3).....	33
4.	RECOMENDACIONES TÉCNICAS.....	35
5.	Guía de Formatos de Archivos	37
5.1.	Formato horario (Hourly)	38
5.2.	Formato diario (daily).....	39
5.3.	Formato mensual (monthly)	40
5.4.	Errores comunes y soluciones	41
6.	Figuras generadas por el módulo de entrenamiento	45
6.1.	Evaporación Acumulada Anual (Datos Brutos)	45
6.2.	Correlación entre Features y Objetivo.....	46
6.3.	Series de Tiempo por Estación	47
6.4.	Justificación del Suavizado del Objetivo.....	48
6.5.	Falla del Modelo General en ANT	49
6.6.	Evaporación Acumulada (Real vs ML)	50
6.7.	Dispersión Real vs Predicho.....	51
6.8.	Métricas (R^2 y KGE) por Estación	52
6.9.	Comparación Acumulada Real vs ML vs PM	53
6.10.	Material de apoyo para validación del entrenamiento.....	53
7.	Procedimiento de Reconstrucción de Datos Diarios y mensuales.....	56
7.1.	Detección automática del tipo de archivo.....	56
7.2.	Reconstrucción desde datos diarios.....	56
7.3.	Reconstrucción desde datos mensuales	57
7.4.	Radiación Solar Diaria.....	58
7.5.	Temperatura Media Diaria.....	58
7.5.1.	Amplitud Térmica Diaria.....	58
7.6.	Velocidad del Viento y Humedad	58
7.7.	Radiación	59
7.8.	Temperatura.....	59
7.9.	Humedad Relativa	60
7.10.	Velocidad del Viento	60

8. Funciones del entrenamiento	61
8.1. Preprocesamiento: De Horas a Días	61
8.2. Agrupamiento en Clústeres: Enrutador Climático.....	62
8.3. Entrenamiento por Clúster.....	63
8.4. Construcción del Enrutador Climático	65
Referencias	66

Resumen

Este documento ha sido elaborado en el marco del proyecto ENANDES de la Organización Meteorológica Mundial (WMO) en Chile, en colaboración con la Dirección Meteorológica de Chile (DMC), el Ministerio de Energía y la Universidad Diego Portales (UDP). Su propósito es entregar una guía de implementación paso a paso de la metodología y de la herramienta desarrollada, de manera que pueda ser útil no solo en el contexto nacional, sino también para otros Servicios Meteorológicos e Hidrológicos Nacionales miembros de la WMO en América Latina y otras regiones, así como para instituciones y equipos de investigación interesados en aplicar o adaptar esta aproximación.

Aquí se describen los ajustes metodológicos, operativos y de implementación incorporados a la herramienta SOLar EVaporation APplication (SOLEVAPP), una plataforma computacional de código abierto orientada a la estimación de tasas de evaporación en cuerpos de agua considerando la instalación de paneles fotovoltaicos flotantes. La presente actualización se construye sobre la versión previa de SOLEVAPP¹, cuyos fundamentos científicos se encuentran publicados² y para la cual se elaboró un manual operativo específico para la integración en el Explorador Solar del Ministerio de Energía de Chile³. El presente documento pone foco en la representación de coberturas parciales, la calibración local del modelo y su aplicabilidad bajo distintas condiciones climáticas.

En particular, el documento introduce un enfoque híbrido que combina el modelo físico de Penman–Monteith con técnicas de aprendizaje automático orientadas a modelar el residuo entre estimaciones teóricas y observaciones experimentales. Esta última corresponde a mediciones de evaporación en bandeja y variables hidrometeorológicas observadas en estaciones distribuidas a lo largo del territorio chileno, lo que permite evaluar el desempeño del modelo bajo distintas condiciones climáticas (climas secos y templados).

La herramienta se organiza en una arquitectura modular compuesta por un módulo de entrenamiento y un módulo de operación. El primero permite la calibración regional del modelo mediante agrupamiento climático y entrenamiento por clúster, mientras que el segundo permite la aplicación operativa del modelo a nuevos sitios, la clasificación climática automática y la simulación de escenarios con distintos niveles de cobertura de paneles flotantes.

Se exponen los procedimientos implementados para operar con datos meteorológicos horarios, diarios y mensuales, incluyendo métodos de reconstrucción de series horarias físicamente coherentes cuando la resolución temporal es limitada. En conjunto, se presentan los alcances, supuestos y condiciones de uso de la versión actualizada de la herramienta SOLEVAPP.

¹ <https://solar.minenergia.cl/fotovoltaico>

² Publicación correspondiente a la versión previa del desarrollo <https://doi.org/10.1016/j.apenergy.2025.125556>

³ https://solar.minenergia.cl/downloads/Modulo_Evaporacion_DV.pdf



WORLD
METEOROLOGICAL
ORGANIZATION



ADAPTATION FUND



udp UNIVERSIDAD
DIEGO PORTALES



— ENANDES —

El documento se organiza en tres bloques principales. En primer lugar, se presenta una introducción que define los objetivos del trabajo y contextualiza el estudio de caso en Chile. Luego, se describe el sistema en términos generales, abordando tanto el porcentaje de cobertura como la implementación de un modelo ajustado mediante técnicas de Machine Learning. Finalmente, se desarrolla la estructura modular del software, diferenciando el módulo de entrenamiento, orientado al *backend* y al procesamiento, preparación y entrenamiento de los datos, y el módulo de operación, enfocado en el *front-end* y en la gestión automática de archivos diarios y mensuales, incluyendo los procesos de detección, reconstrucción y ejecución del flujo final. El documento se complementa con anexos que contienen material de apoyo.

Este documento se basa en los archivos *solevapp_utils.py* y *SOLEVAPP_v2_colab.py*, incluidos en los archivos Jupyter Notebook disponibles en la siguiente dirección: <https://gitlab.com/rodrigo.caceres1/enandes-solevapp/>

1. INTRODUCCIÓN

Los sistemas fotovoltaicos flotantes (FPV, por sus siglas en inglés) representan una alternativa frente a las restricciones propias de las instalaciones solares en tierra, al permitir un uso más eficiente del espacio y un mejor desempeño energético. En comparación con los sistemas terrestres, presentan ventajas estructurales y operativas relevantes, especialmente asociadas a la reducción del uso de suelo y al aumento de la producción eléctrica [1].

Entre los principales atributos de los FPV se encuentra el enfriamiento pasivo de los módulos, el cual disminuye su temperatura de operación y contribuye a mejorar la eficiencia de los paneles. Como resultado, los sistemas FPV pueden incrementar la generación de energía en el orden de 1.58–2.00% respecto de sistemas fotovoltaicos terrestres convencionales [2]. Junto con las mejoras energéticas, los FPV aportan beneficios importantes en términos de conservación hídrica, al reducir la evaporación en embalses y cuerpos de agua. Algunos estudios muestran que esta reducción puede alcanzar hasta un 85% bajo determinadas condiciones [3, 4]. En conjunto, estos efectos permiten disminuir las pérdidas de agua y generar condiciones favorables para una reducción del costo nivelado de la energía (LCOE) a medida que aumenta la capacidad instalada.

En este contexto, la correcta estimación de la evaporación es un aspecto central en la evaluación de sistemas FPV. Para este propósito, se han desarrollado distintos modelos empíricos y de base física orientados a la estimación de la evapotranspiración, estrechamente vinculada a la evaporación. Entre los modelos empíricos más utilizados se encuentran los de Penman [5], Penman–Monteith [6], Turc [7] y Hargreaves [8], los cuales consideran variables como la temperatura, la humedad, la velocidad del viento y la radiación solar. Estos modelos evidencian que el desempeño de los sistemas FPV depende fuertemente de las condiciones geográficas y climáticas, las cuales afectan tanto la producción energética como la reducción de la evaporación. Entre estos modelos, el método de Penman–Monteith es el comúnmente utilizado para la estimación de la evapotranspiración potencial.

1.1. Objetivos del documento

El presente documento tiene como propósito servir como informe técnico y manual de referencia de la herramienta SOLEVAPP, documentando su versión actualizada, los ajustes metodológicos incorporados y su arquitectura computacional. En él se describen los fundamentos físicos y de aprendizaje automático utilizados para la estimación de la evaporación en cuerpos de agua con sistemas fotovoltaicos flotantes, así como los procedimientos de calibración, validación experimental y regionalización climática. Asimismo, se establecen los requisitos de datos, los flujos de procesamiento y las condiciones de uso de la herramienta, con el objetivo de asegurar su correcta aplicación, reproducibilidad y utilización operativa en contextos técnicos e institucionales.



WORLD
METEOROLOGICAL
ORGANIZATION



ADAPTATION FUND



udp UNIVERSIDAD
DIEGO PORTALES



— ENANDES —

Este manual de usuario está dirigido a una audiencia amplia e interdisciplinaria vinculada a la planificación hídrica y energética, el desarrollo de energías renovables, el análisis hidrológico y la toma de decisiones basada en información climática. SOLEVAPP¹ ha sido concebido como una herramienta aplicada, con una base científica sólida, orientada tanto a usuarios técnicos como a actores institucionales. El público objetivo incluye profesionales de los Servicios Meteorológicos e Hidrológicos Nacionales, organismos de energía y agua, analistas de políticas públicas, investigadores y especialistas en modelación energía-agua. El manual entrega los fundamentos metodológicos, los requisitos de datos y las pautas necesarias para interpretar las estimaciones de evaporación y los escenarios con sistemas fotovoltaicos flotantes, facilitando su uso en evaluación técnica, planificación, formulación de políticas y estudios académicos.

1.2. Estudio de caso en Chile

El rápido desarrollo de la energía solar fotovoltaica ha abierto oportunidades para soluciones como los sistemas fotovoltaicos flotantes, que permiten generar energía sin competir por suelo y, simultáneamente, reducir la evaporación desde cuerpos de agua.

En este contexto, Chile fue seleccionado como caso de estudio en el marco del proyecto ENANDES para responder a necesidades identificadas por la Dirección Meteorológica de Chile (DMC) y el Ministerio de Energía respecto de la evaluación de esta tecnología y su posible aplicación en el país. El caso chileno ofrece condiciones especialmente interesantes para el desarrollo y evaluación de herramientas como SOLEVAPP, ya que su extensión territorial permite analizar una amplia diversidad de climas, desde condiciones áridas y semiáridas hasta una amplia variedad de climas templados, junto con distintos niveles de irradiación solar.

A ello se suma la convergencia de factores estructurales como un elevado nivel de estrés hídrico y una fuerte presencia de infraestructuras de generación hidroeléctrica y consumo agrícola, lo que refuerza la relevancia de estudiar soluciones que integren generación renovable y gestión del recurso hídrico. La disponibilidad de bases de datos meteorológicas públicas de alta resolución, junto con una infraestructura institucional consolidada, hacen de Chile un entorno adecuado para evaluar el desempeño del modelo en condiciones contrastantes. En este marco, SOLEVAPP utiliza como fuente principal de información meteorológica el Explorador Solar del Ministerio de Energía de Chile, desde el cual accede, vía API, a series de radiación global horizontal, temperatura del aire, humedad relativa y velocidad del viento, disponibles en resolución horaria o diaria. La versión más reciente de la herramienta SOLEVAPP, presentada en este documento, incorpora observaciones de evaporación de bandeja provistas por la Dirección Meteorológica de Chile para fines de ajuste y validación. En este caso, el principal desafío es disponer de series temporales coherentes que coincidan en el mismo año y hora para variables clave como evaporación de bandeja, radiación instantánea, velocidad del viento,

humedad relativa y temperatura ambiente. A partir de los registros disponibles, se seleccionaron ocho estaciones distribuidas a lo largo del territorio nacional, desde el norte hasta el sur, abarcando cerca de 2000 km, tal como se muestra en la Fig. 1, todas con información completa de evaporación para un año continuo.

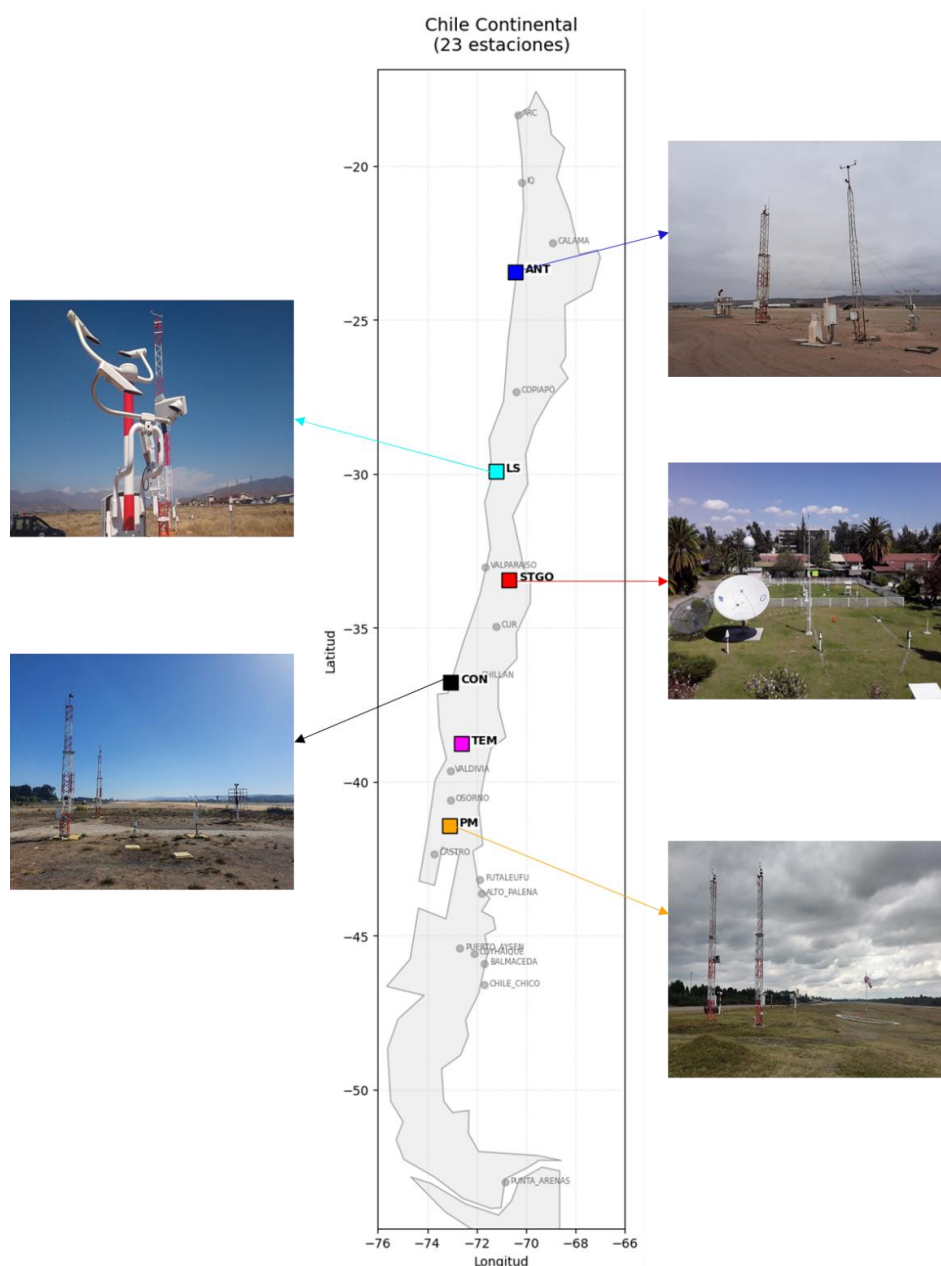


Figura 1. Estaciones meteorológicas seleccionadas para la validación experimental de evaporación. ANT: Antofagasta; LS: La Serena; STGO: Santiago; CUR: Curicó; CON: Concepción; TEM: Temuco; PM: Puerto Montt.

2. DESCRIPCIÓN GENERAL DEL SISTEMA

SOLEVAPP es una plataforma computacional orientada a la estimación de tasas de evaporación en cuerpos de agua, que incorpora la cobertura parcial o total de la superficie libre mediante paneles fotovoltaicos flotantes. Actualmente, la plataforma se encuentra implementada en el Explorador Solar de Chile (Fig. 2)¹ y utiliza el modelo de Penman-Monteith para la predicción de la evaporación. En su primera etapa, el funcionamiento es binario, considerando exclusivamente escenarios de 0% y 100% de cobertura con paneles fotovoltaicos. Para mayor detalle sobre el modelo implementado en la fase inicial de SOLEVAPP, se recomienda revisar el trabajo de Cáceres et al. [9] disponible en la nota al pie².

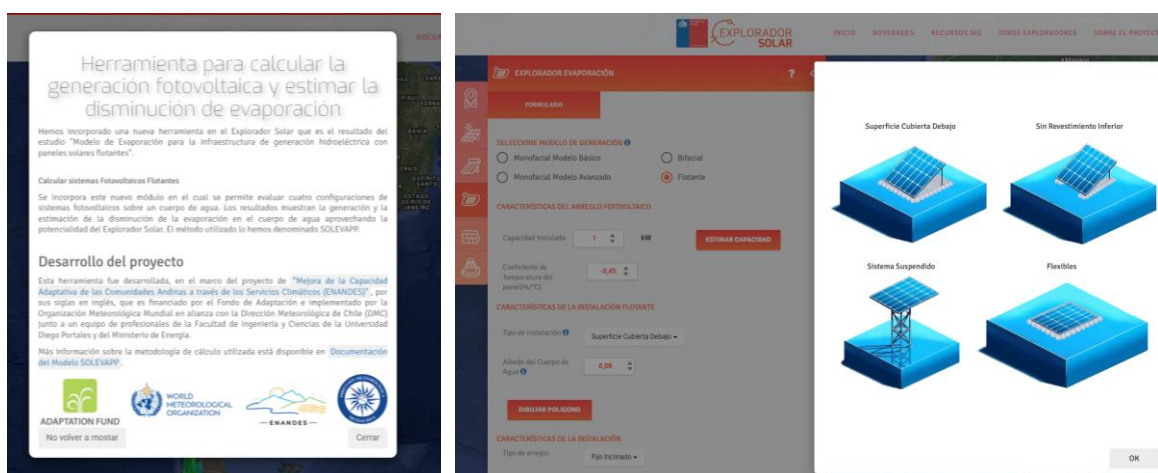


Figura 2: Herramienta SOLEVAPP implementada en explorador solar del ministerio de energía de Chile

La segunda etapa del proyecto apunta a expandir y fortalecer el módulo actualmente disponible en el Explorador, incorporando el modelado de coberturas parciales con paneles fotovoltaicos y sumando información experimental medida en distintas localidades de Chile para ajustar el modelo teórico inicialmente utilizado. Aunque robusto, estos modelos teóricos pueden presentar sesgos sistemáticos cuando se aplican sin calibración local en climas extremos, como zonas áridas (como en la zona norte de Chile) o con alta advección, ya que se desarrollan bajo condiciones ideales. A ello se suma que, en su objetivo de generalizar, recurren a correlaciones para estimar coeficientes de pérdida o corrección.

Esta segunda etapa busca avanzar en aspectos operativos y metodológicos que mejoren la aplicabilidad de la plataforma. Actualmente, SOLEVAPP está diseñada para trabajar con datos horarios o diarios continuos, por lo que resulta necesario ajustarla para enfrentar conjuntos de datos discontinuos o segmentados. Asimismo, esta etapa busca complementar el enfoque físico con herramientas capaces de capturar variaciones locales que la teoría no siempre reproduce con precisión, mediante la integración de técnicas basadas en datos, como modelos de aprendizaje automático.

Dentro de este tipo de herramientas, es habitual recurrir a modelos flexibles, como las redes neuronales, capaces de representar relaciones no lineales complejas. Sin embargo, al no incorporar restricciones físicas, su comportamiento puede volverse incoherente cuando las condiciones de entrada se alejan de aquellas consideradas durante el entrenamiento. En este contexto, la segunda etapa propone utilizar la ecuación física como estructura principal, asegurando estabilidad y coherencia, y emplear algoritmos de aprendizaje automático únicamente para modelar el residual entre los valores teóricos y los observados. Este enfoque híbrido busca mejorar la precisión local sin perder consistencia física ni generalidad.

A continuación, se presenta un breve resumen de SOLEVAPP, seguido de una descripción de las modificaciones incorporadas en esta etapa, las cuales se concentran en los ajustes realizados a la función *evap_estimate_API()*, responsable de la estimación de la evaporación en SOLEVAPP. Este capítulo tiene como propósito presentar los fundamentos técnicos y teóricos de la herramienta desarrollada. La descripción del uso directo se encuentra disponible en el capítulo 3.

2.1. Implementación inicial de SOLEVAPP

En esta sección se presenta el funcionamiento general de SOLEVAPP en su versión inicial. Los fundamentos científicos y la formulación detallada del modelo se describen en la publicación asociada². Aquí se explica, de forma general, qué información utiliza la herramienta, cómo procesa los datos y qué resultados entrega al usuario.

SOLEVAPP es el módulo del Explorador Solar encargado de estimar la evaporación en cuerpos de agua, con o sin sistemas fotovoltaicos flotantes. El módulo utiliza la información meteorológica del Explorador y la configuración de paneles flotantes escogida (ver Fig. 2) para estimar, en base horaria, la evaporación potencial de agua bajo dichas condiciones. Con esto, el módulo busca facilitar la comparación entre el agua evaporada desde una superficie libre y una cubierta con paneles fotovoltaicos.

En su versión inicial, la información meteorológica que usa SOLEVAPP corresponde a datos horarios de radiación global horizontal, temperatura ambiente, humedad relativa, velocidad del viento, latitud y altitud. Estos datos se complementan con parámetros definidos por el usuario, tales como el tipo de arreglo fotovoltaico flotante y el albedo de la superficie del agua (capacidad de la superficie para reflejar la radiación solar). Para apoyar la correcta selección por parte del usuario, el módulo incluye información visual sobre los tipos de arreglo fotovoltaico flotante (Fig. 2), así como valores de referencia para orientar la estimación del albedo de la superficie del agua.

Al ejecutarse, SOLEVAPP sigue una secuencia de operación compuesta por tres etapas: Primero se procesan los datos meteorológicos con el objetivo de revisar la compatibilidad de las unidades requeridas por el modelo físico/matemático que sustenta a SOLEVAPP. Así, en caso de ser necesario, se ajustan y convierten las unidades de las distintas variables meteorológicas del módulo. En segundo lugar se procede con el cálculo de la radiación neta



WORLD
METEOROLOGICAL
ORGANIZATION



ADAPTATION FUND



udp UNIVERSIDAD
DIEGO PORTALES



— ENANDES —

sobre la superficie del agua, considerando la cobertura del sistema fotovoltaico flotante seleccionado. Finalmente, se procede con la estimación de la evaporación a partir de los datos procesados y el balance energético asociado a la selección del sistema fotovoltaico flotante.

Luego de su ejecución, la versión inicial de SOLEVAPP entrega la estimación de evaporación en base horaria. Así, SOLEVAPP busca que esta salida de información pueda emplearse para analizar tendencias temporales y estimar diferencias entre una superficie libre y una superficie intervenida con una determinada configuración fotovoltaica flotante.

2.2. Porcentaje de cobertura

En su segunda versión, SOLEVAPP incorpora el parámetro β , destinado a representar cuerpos de agua parcialmente cubiertos por paneles fotovoltaicos flotantes. Así, este parámetro expresa la fracción de superficie ocupada por el sistema flotante y permite extender la herramienta desde el caso original (cobertura total) hacia configuraciones cercanas a condiciones reales de operación.

2.2.1. Definición del porcentaje de cobertura

El parámetro β representa la fracción de superficie cubierta por paneles fotovoltaicos flotantes. Su interpretación es directa:

- $\beta = 0.0$: superficie libre
- $\beta = 1.0$: cobertura total (comportamiento del modelo inicial)
- $0.0 < \beta < 1.0$: condiciones de cobertura parcial

De esta forma, el usuario puede analizar no solo los casos extremos, sino también situaciones intermedias representativas de proyectos reales.

2.2.2. Cálculo de las dos contribuciones a la evaporación

Una vez definido el nivel de cobertura, se procede con el cálculo de la evaporación total considerando:

- la fracción de superficie libre, donde el agua se evapora sin la presencia de paneles flotantes
- la fracción de superficie cubierta, donde la evaporación está afectada por la presencia de paneles flotantes



WORLD
METEOROLOGICAL
ORGANIZATION



ADAPTATION FUND



udp UNIVERSIDAD
DIEGO PORTALES



— ENANDES —

Esto permite representar el cuerpo de agua como una superficie compuesta, en lugar de asumir que toda el área se encuentra en una sola condición.

2.2.3. Cálculo de la potencial evaporación efectiva

La evaporación total se calcula como un promedio espacial ponderado entre la evaporación de la superficie libre y la evaporación de la superficie cubierta por paneles. La expresión utilizada es la siguiente:

$$E_{\beta} = (1 - \beta)E_{free} + \beta E_{panel} \quad (1)$$

donde E_{free} corresponde a la evaporación de la superficie libre y E_{panel} a la evaporación en la zona cubierta por paneles flotantes. Este enfoque permite mantener la consistencia física de la herramienta y, al mismo tiempo, introducir una representación más flexible del grado de ocupación de la superficie.

La incorporación de β fue diseñada para mantener compatibilidad con la versión base de SOLEVAPP. Si el usuario no define explícitamente este parámetro, la herramienta conserva el comportamiento original y asume cobertura total. Cuando β es ingresado, el sistema verifica que su valor se mantenga dentro del rango:

$$0.0 \leq \beta \leq 1.0$$

Si el valor se encuentra fuera de ese intervalo, se genera un error explícito. De manera complementaria, si durante el cálculo se obtiene una evaporación negativa, esta se ajusta a cero, manteniendo el mismo criterio operativo empleado en la implementación previa.

Una vez validado, el parámetro β se incorpora al flujo de cálculo de SOLEVAPP sin modificar la lógica general de uso de la herramienta. En términos prácticos, el usuario puede seguir utilizando el mismo esquema de análisis descrito en la Sección 2.1, pero ahora incluyendo explícitamente el porcentaje de cobertura para representar escenarios parciales.

Esta actualización mejora la utilidad de SOLEVAPP, ya que permite comparar cómo cambia la evaporación a medida que aumenta o disminuye la superficie ocupada por paneles fotovoltaicos flotantes.

2.2.4. Generación de resultados

Como resultado, la herramienta entrega una estimación de evaporación que incorpora el efecto de cobertura parcial. A modo de ejemplo, la Fig. 3 muestra la reducción acumulada de evaporación para dos escenarios de cobertura parcial: a) 40 % y b) 10 %, expresada en l/m^2 . Estos resultados permiten visualizar de manera directa cómo una mayor cobertura incrementa el ahorro acumulado respecto del caso sin paneles.

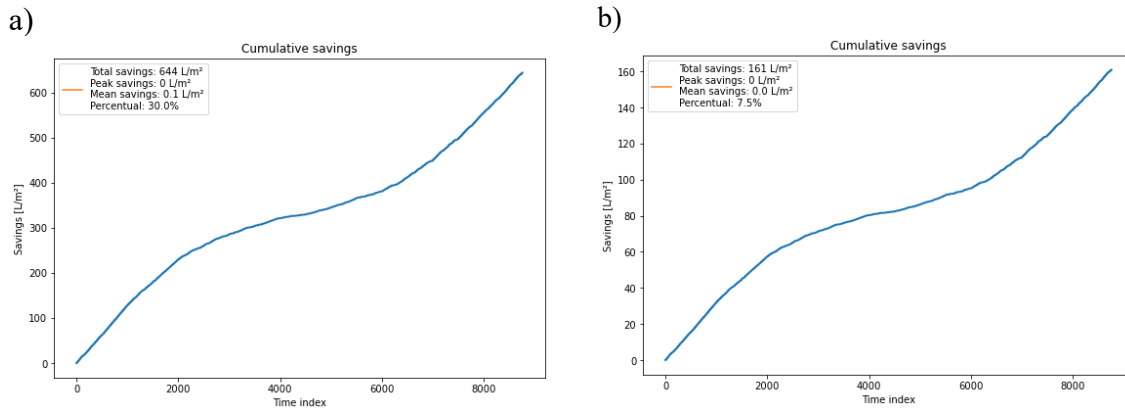


Figura 3: Reducción acumulada de evaporación con a) 40% de cobertura y b) 10%. Total savings es el ahorro total por incorporar la cobertura. Peak, mean y porcentual es el máximo ahorro, el ahorro promedio en el año y el porcentaje de ahorro total en el año definido como la razón entre la evaporación acumulada con cobertura y sin cobertura.

2.3. Implementación de modelo ajustado por Machine Learning para SOLEVAPP

En esta segunda etapa, SOLEVAPP incorpora una opción de cálculo ajustado que combina la base física del modelo original con técnicas de aprendizaje automático. Esta actualización permite mejorar la estimación de evaporación sobre superficie libre, reduciendo sesgos asociados a las variables climatológicas mediante el uso de mediciones experimentales de evaporación de bandeja. La selección del enfoque se realiza mediante la variable **modelo**, que admite las opciones “físico” o “ajustado”.

La función responsable de este flujo es `evap_estimate_API()`, que según el valor del argumento “modelo” redirige al cálculo físico tradicional o al cálculo híbrido basado en Machine Learning:

```
evap_estimate_API(df, alturas, tipo, alpha, beta= $\beta$ , modelo="ajustado")
```

Así, el usuario define el enfoque de cálculo que utilizará la herramienta. Cuando se selecciona la opción “físico”, SOLEVAPP reproduce el flujo de cálculo tradicional de la implementación original (etapa inicial). Cuando se selecciona la opción “ajustado”, la herramienta activa un esquema híbrido que conserva la base física del modelo, pero incorpora una corrección mediante Machine Learning para la evaporación de la superficie libre.

El enfoque híbrido (modelo ajustado) consiste en los siguientes 5 pasos:

2.3.1. Paso 1: Cálculo físico de evaporación en superficie libre

Primero, SOLEVAPP realiza un cálculo físico base de la evaporación sobre superficie libre utilizando el enfoque Penman–Monteith. Este cálculo entrega una primera estimación de

evaporación y, además, permite derivar variables atmosféricas complementarias que serán utilizadas posteriormente por el modelo. Con esto, se busca asegurar que la implementación híbrida conserve coherencia con la lógica física general del sistema.

2.3.2. Paso 2: Construcción de variables de entrada

A partir de los datos meteorológicos disponibles y del cálculo físico inicial, la herramienta construye las variables de entrada requeridas por el modelo de Machine Learning. Entre estas variables se incluyen indicadores asociados a la energía térmica, variables asociadas a humedad y viento, y representaciones de la estacionalidad.

En la implementación actual, estas variables incluyen, entre otras, GHI diario, déficit de presión de vapor (VPD), humedad mínima y máxima, viento diurno y nocturno, y la fase estacional representada mediante funciones seno y coseno del día del año.

2.3.3. Paso 3: Asignación automática a un clúster climático

Una vez construidas las variables de entrada, el caso analizado se asigna automáticamente a un clúster climático. Cada clúster dispone de un modelo de Machine Learning independiente, entrenado con características meteorológicas (de radiación y atmosféricas) propias de condiciones climáticas similares.

Esta asignación permite que la corrección aplicada por el modelo ajustado sea consistente con el contexto climático del sitio evaluado, evitando el uso de una única corrección general para todos los casos.

2.3.4. Paso 4: Predicción ajustada de evaporación potencial en superficie libre y cubierta

Con el clúster ya definido, el modelo de Machine Learning correspondiente estima la evaporación ajustada para la superficie libre. Esta predicción se apoya en mediciones experimentales de evaporación de bandeja y busca corregir la estimación física inicial cuando existen sesgos asociados a condiciones meteorológicas específicas.

En términos operativos, esto significa que el modelo ajustado no reemplaza completamente la base física del cálculo, sino que mejora la representación de la evaporación libre a partir de patrones aprendidos desde datos observados.

En paralelo al ajuste de la superficie libre, SOLEVAPP mantiene el cálculo físico de la evaporación correspondiente a la superficie cubierta por paneles fotovoltaicos flotantes. Este punto es importante porque el efecto del panel sobre la radiación incidente y el balance energético continúa siendo controlado mediante la formulación física del sistema. Así, la

parte cubierta conserva la lógica original del modelo, mientras que el ajuste por Machine Learning se concentra en mejorar la estimación de la porción libre del agua.

2.3.5. Paso 5: Combinación final mediante β

Finalmente, la evaporación de la superficie libre ajustada y la evaporación de la superficie cubierta se combinan mediante la ecuación de ponderación asociada al parámetro β . De este modo, la física controla el efecto del panel fotovoltaico flotante, mientras que Machine Learning corrige la evaporación correspondiente a la parte libre del cuerpo de agua.

Esto da origen a una arquitectura híbrida, en la que ambas aproximaciones cumplen funciones complementarias dentro del mismo flujo de cálculo. Como salida, la herramienta entrega una estimación de evaporación compatible con el resto del flujo operativo de SOLEVAPP, permitiendo comparar el desempeño del enfoque físico tradicional con el del enfoque ajustado.

Dada su arquitectura híbrida, el modelo ajustado fue validado utilizando tanto observaciones experimentales como simulaciones físicas. Sin embargo, en este manual, el énfasis está puesto en su implementación operativa.

2.4. Visión general de los datos requeridos

La Tabla 1 presenta un resumen general de los datos requeridos tanto para la operación de SOLEVAPP como para la fase actual de calibración del modelo ajustado. En términos prácticos, SOLEVAPP utiliza variables meteorológicas y descriptores básicos de la localidad para estimar la evaporación en condiciones de superficie libre y con paneles fotovoltaicos flotantes. La fase de calibración, en cambio, utiliza esas salidas junto con mediciones experimentales de evaporación y datos de estaciones para entrenar y evaluar el modelo ajustado. El detalle del formato de archivos y ejemplos de carga de datos se presenta en la sección 3.

Tabla 1: Descripción general de datos a utilizar

Etapa	Categoría de datos	Variables o contenido principal	Resolución / formato general	Uso principal
Implementación de SOLEVAPP (fase inicial)	Datos meteorológicos de entrada	Radiación global horizontal (GHI), temperatura ambiente, humedad relativa, velocidad del viento	Series temporales horarias por sitio	Datos climáticos principales para la estimación de evaporación
	Descriptores del sitio	Latitud, longitud, altitud	Datos fijos por localidad o estación	Caracterización geográfica del sitio y apoyo a los cálculos de radiación y evaporación
	Parámetros de configuración	Tipo de sistema fotovoltaico flotante (Superficie cubierta debajo, Sin revestimiento inferior, Sistema suspendido, Flexibles), albedo del agua	Parámetros definidos por el usuario	Representación de la configuración seleccionada del sistema flotante
Implementación de SOLEVAPP (segunda fase)	Parámetro de cobertura	(β), fracción de superficie cubierta por paneles	Valor escalar entre 0 y 1	Representación de escenarios de cobertura parcial o total
	Parámetro de selección de modelo	Opción “físico” u opción “ajustado”	Parámetro definido por el usuario	Selección entre el modelo físico tradicional y el modelo híbrido ajustado
Calibración (segunda fase)	Base meteorológica histórica con resultados SOLEVAPP	Variables meteorológicas horarias y estimaciones base de evaporación generadas por SOLEVAPP para las localidades de calibración	Archivo tabular horario	Construcción de predictores y generación de la base física inicial del ajuste
	Mediciones experimentales de evaporación	Registros observados de evaporación de bandeja	Archivo tabular asociado a estaciones de medición	Referencia experimental para calibración y validación del modelo ajustado
	Datos de estaciones	Código, nombre de estación, localidad, coordenadas, altitud	Archivo tabular de datos	Identificación, georreferenciación y vínculo entre observaciones y simulaciones
	Variables derivadas para Machine Learning	GHI diario, déficit de presión de vapor (VPD), humedad mínima y máxima, viento diurno y nocturno, indicadores estacionales	Variables construidas a partir de datos meteorológicos y resultados físicos	Variables de entrada para el modelo de aprendizaje automático
	Clasificación climática	Asignación de cada caso a un clúster climático	Clasificación interna del flujo de calibración	Selección del modelo de Machine Learning más apropiado para cada condición climática

3. ESTRUCTURA MODULAR DEL SOFTWARE DESARROLLADO

Con el fin de facilitar su uso y documentación, el código de SOLEVAPP se organizó de forma modular en un Notebook Maestro, en línea con recomendaciones de la Organización Meteorológica Mundial (WMO). Este documento presenta la estructura del notebook siguiendo su secuencia de ejecución e indica, en cada etapa, su función, los datos de entrada requeridos y los resultados generados. Su implementación está pensada para entornos de trabajo como Google Colab.

El notebook utiliza funciones contenidas en los archivos `solevapp_utils.py` y `solevapp_v1_colab.py`, disponibles en el repositorio:

<https://gitlab.com/rodrigo.caceres1/enandes-solevapp>

3.1. Estructura del flujo de trabajo

Archivo Fuente: `SOLEVAPP_maestro_v2.ipynb`

El notebook se organiza en dos partes claramente diferenciadas según el perfil del usuario:

Parte	Pasos	Dirigido a:	Tiempo estimado
A — Operación	0 al 6	Todo usuario que quiera estimar evaporación en un nuevo sitio	10–20 min
B — Entrenamiento	T1 al T3	Equipo técnico que necesite reentrenar con nuevos datos de estaciones	30–120 min

La Parte A es el flujo principal y el que se documenta con mayor detalle en esta sección. La Parte B se describe en la Sección 3.4 de este documento.

El repositorio oficial del proyecto contiene todos los archivos necesarios para ejecutar el Notebook Maestro:

Archivo / Carpeta	Descripción	Requerido para
SOLEVAPP_maestro_v2.ipynb	Notebook Maestro y punto de entrada principal	Parte A y B
solevapp_utils.py	Funciones extraídas de los módulos originales	Parte A y B
SOLEVAPP_v1_colab.py	Modelo físico de evaporación Penman-Monteith	Parte A
models/modelo_cluster_N.pkl	Modelos ML entrenados por clúster ($N = 0, 1, 2, 3$)	Parte A (modelo ajustado)
models/scaler_X_cluster_N.pkl	Escaladores de variables de entrada por clúster	Parte A (modelo ajustado)
models/scaler_y_cluster_N.pkl	Escaladores del objetivo por clúster	Parte A (modelo ajustado)
models/features_list.txt	Lista de variables de entrada requeridas por los modelos	Parte A (modelo ajustado)
models/enrutador_knn.pkl	Clasificador KNN del enrutador climático	Referencia
models/enrutador_scaler.pkl	Escalador del enrutador climático	Referencia
data/ejemplos/	Archivos de ejemplo en los tres formatos (horario, diario, mensual)	Referencia

3.2. Flujo de operación: Parte A

La Parte A permite estimar la reducción de evaporación producida por la instalación de paneles FPV sobre un cuerpo de agua. El flujo completo se ejecuta en seis pasos secuenciales cuya lógica es la siguiente:





3.2.1. Configuración inicial (Paso 0)

El Paso 0 prepara el entorno de ejecución. Debe ejecutarse una sola vez al abrir el notebook y no requiere ninguna edición por parte del usuario.

Se compone de cuatro subceldas:

- 0A — Detección de entorno: identifica automáticamente si el notebook se ejecuta en Google Colab, Jupyter Notebook local, JupyterLab o VS Code, y configura los helpers de carga de archivos en consecuencia.
- 0B — Instalación de dependencias: instala los paquetes Python requeridos (lightgbm, scikit-learn, pandas, numpy, matplotlib, seaborn, ipywidgets, joblib, requests, scipy).
- 0C — Descarga desde el repositorio: descarga automáticamente los archivos de código (solevapp_utils.py, SOLEVAPP_v1_colab.py) y todos los modelos entrenados (.pkl) desde el repositorio GitLab del proyecto.
- 0D — Importación de funciones: importa todas las funciones necesarias desde solevapp_utils.py y SOLEVAPP_v1_colab.py.

Importante: Si el repositorio no está disponible o la conexión falla, los archivos deben colocarse manualmente en el mismo directorio que el notebook antes de ejecutar el Paso 0D.

3.2.2. Parámetros del sitio (Paso 1)

El usuario debe editar esta celda con las coordenadas y características físicas del embalse a estudiar. Son los únicos parámetros que requieren edición manual en el flujo de operación.

Parámetro	Variable	Unidad	Valor por defecto	Descripción
Latitud	LAT	° decimal	-38.766	Negativo = hemisferio sur
Longitud	LON	° decimal	-72.637	Negativo = oeste
Altitud	ALT	msnm	96.0	Metros sobre el nivel del mar

En el Paso 6 el usuario puede modificar el valor de albedo (fracción de radiación solar que refleja la superficie del cuerpo de agua) para distintos tipos de superficie al configurar los parámetros de la simulación.



WORLD
METEOROLOGICAL
ORGANIZATION



ADAPTATION FUND



udp UNIVERSIDAD
DIEGO PORTALES



— ENANDES —

```
ALBEDOS = {
  "Cuerpo de agua (0.08)": 0.08,
  "Asfalto seco (0.12)": 0.12,
  "Asfalto mojado (0.18)": 0.18,
  "Césped (0.18)": 0.18,
  "Césped nuevo (0.26)": 0.26,
  "Césped seco (0.33)": 0.33,
  "Arena desierto (0.40)": 0.40,
  "Nieve (0.98)": 0.98,
}
```

3.2.3. Carga y preparación de datos (Paso 2)

Este paso carga el archivo de datos meteorológicos del sitio, detecta su formato, verifica que las columnas requeridas estén presentes y, si los datos no son horarios, los convierte automáticamente a resolución horaria.

Formatos de archivo aceptados

La herramienta acepta tres resoluciones temporales, detectadas automáticamente a partir de la columna clave del archivo:

Formato	Columna clave	Columnas obligatorias	Columnas opcionales
Horario	Fecha+Hora (UTC)	ghi_S1, temp_S1, rh_S1, vel_S1	evap_S1, E_PM_S1, Mes, dia_del_anio
Diario	date	ghi, t_mean, rh_mean, wind_mean	t_min, t_max, evap, Mes, dia_del_anio
Mensual	month	ghi, t_mean, rh_mean, wind_mean	t_min, t_max, evap

Los nombres de columna deben coincidir exactamente con los indicados en la tabla anterior. Si el archivo utiliza nombres distintos, la herramienta se detiene y muestra un mensaje de error con instrucciones para renombrar las columnas tanto en Excel como en Python. Ver Anexo 1 del manual para ejemplos completos de cada formato.

Conversión de datos diarios y mensuales

Cuando el archivo de entrada es diario o mensual, la herramienta genera automáticamente una serie horaria sintética mediante interpolación física. Este preprocesamiento es necesario porque el modelo físico SOLEVAPP, los modelos de aprendizaje automático y el enrutador climático fueron desarrollados y entrenados con datos horarios.

La reconstrucción aplica los siguientes supuestos por variable:

Variable	Método de reconstrucción	Supuesto físico
Radiación (GHI)	Curva sinusoidal solar ajustada por día del año	Máximo al mediodía solar, cero antes del amanecer y después del atardecer
Temperatura	Función sinusoidal entre T_min y T_max	Mínimo ~06:00h, máximo ~15:00h
Humedad relativa	Comportamiento inverso a la temperatura	Máximo en la madrugada, mínimo en la tarde
Velocidad del viento	Patrón diurno creciente hacia la tarde	Calmas nocturnas, máximo en horas de mayor convección

Ver Anexo 3 del manual para el detalle metodológico de cada variable.

3.2.4. Verificación y conversión de unidades (Paso 3)

Tras cargar y convertir los datos, el usuario debe verificar que las variables estén en las unidades correctas. La herramienta presenta una interfaz interactiva con cuatro selectores, uno por variable.

Variable	Unidad requerida	Conversiones disponibles
Radiación (GHI)	MJ/m ² /h	W/m ² → MJ/m ² /h (÷ 3.6)
Temperatura	°C	°F → °C ((T-32) × 5/9)
Humedad relativa	Fracción (0–1)	Porcentaje % → fracción (÷ 100)
Velocidad del viento	m/s a 2 m	km/h → m/s (÷ 3.6) m/s a 10 m → 2 m (× 0.75)

El usuario selecciona la unidad actual de sus datos y presiona **Aplicar y continuar**, tal como se muestra a continuación:

3.2.6. Clasificación climática del sitio, enrutador (Paso 5)

El enrutador climático analiza el perfil meteorológico del sitio y determina automáticamente a qué zona climática pertenece. Esta clasificación determina cuál de los cuatro modelos de aprendizaje automático se utilizará en el Paso 6.

Cómo funciona la clasificación

El algoritmo calcula el promedio de tres variables representativas del sitio (promedios reales de 24 horas sobre todo el período de datos disponible):

Radiación solar global horizontal (W/m^2)

- Humedad relativa (%)
- Velocidad del viento (m/s)

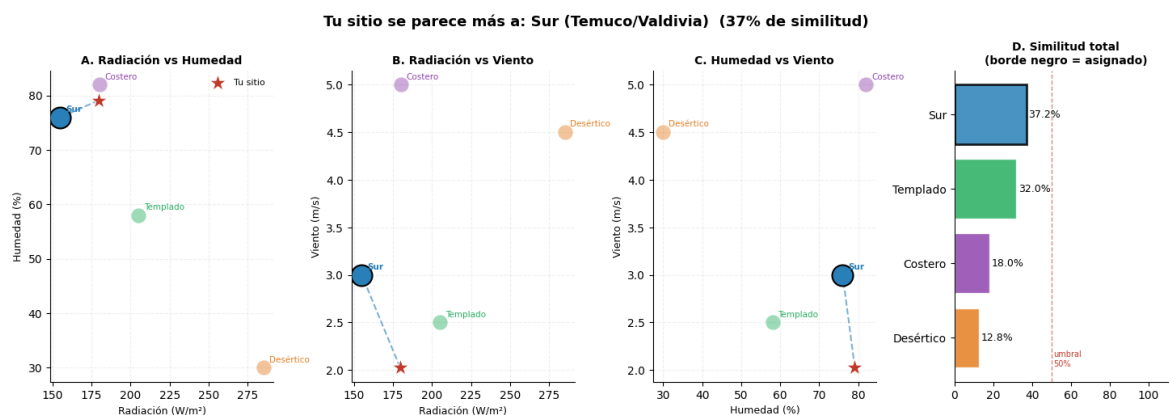
Estas tres variables se normalizan para que tengan el mismo rango de valores, contruidos a partir de estaciones meteorológicas chilenas. Luego se calcula la distancia del sitio a cada arquetipo. El arquetipo de menor distancia determina el clúster asignado.

Clúster	Zona climática	Radiación ref.	Humedad ref.	Viento ref.
0	Árido desértico	285 W/m^2	30%	4.5 m/s
1	Mediterráneo	205 W/m^2	58%	2.5 m/s
2	Oceánico templado	155 W/m^2	76%	3.0 m/s
3	Costero húmedo	180 W/m^2	82%	5.0 m/s

Nota: los valores de referencia de radiación están expresados en W/m^2 promedio 24h, que es la unidad que usa el enrutador internamente para la clasificación. Los datos de entrada al modelo se expresan en $\text{MJ/m}^2/\text{h}$ y son convertidos automáticamente antes de la clasificación.

Interpretación de la visualización

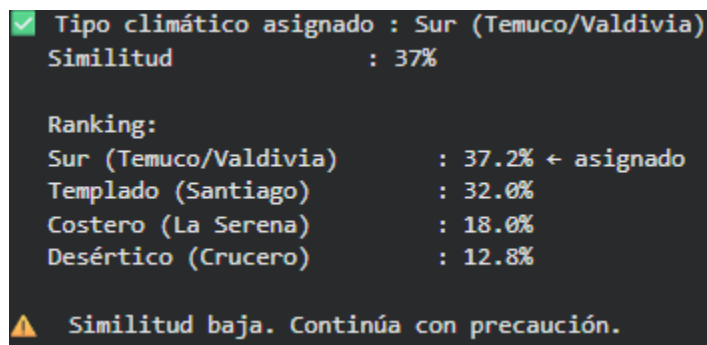
La herramienta genera tres gráficos de dispersión en paralelo, uno por cada par de variables (Radiación vs Humedad, Radiación vs Viento, Humedad vs Viento). En cada gráfico se muestran los cuatro arquetipos climáticos como puntos coloreados, y el sitio del usuario como una estrella roja. Una línea punteada conecta el sitio con el arquetipo asignado, tal como se muestra a continuación:



Este diseño permite al usuario verificar visualmente la clasificación en los tres ejes de decisión simultáneamente, evitando la ambigüedad que ocurriría al mostrar solo dos variables cuando las tres influyen en la clasificación.

Advertencia de clasificación de baja confianza

Cuando el porcentaje de similitud del arquetipo asignado es inferior al 50%, la herramienta muestra una advertencia indicando que el sitio se encuentra en una zona de transición entre arquetipos climáticos, tal como se muestra en la figura:



En este caso los resultados del modelo ajustado deben interpretarse con mayor precaución, ya que el modelo fue entrenado con datos de condiciones climáticas más definidas.

3.2.7. Simulación (Paso 6)

El Paso 6 es el resultado central del flujo de operación. El usuario configura los parámetros del proyecto FPV y la herramienta calcula la evaporación estimada con y sin paneles para el período cubierto por los datos.

Parámetros de la simulación

Parámetro	Opciones	Descripción
-----------	----------	-------------

Modelo	físico / ajustado	Selecciona el enfoque de cálculo (ver 3.2.2)
Tecnología FPV	free / float-A / float-B / float-C / float-D	Tipo de instalación según Bontempo-Scavo et al.
Albedo superficie	Cuerpo de agua (0.08) ... Nieve (0.98)	Albedo de la superficie del embalse
Cobertura (%)	0 al 100% en pasos de 5%	Fracción del embalse cubierta por paneles FPV

Modelos de cálculo disponibles

La herramienta ofrece dos enfoques de cálculo seleccionables mediante el toggle Modelo:

Modelo físico: utiliza exclusivamente el modelo de Penman-Monteith implementado en SOLEVAPP_v1_colab.py. Calcula la evaporación a partir de las variables termodinámicas del aire (temperatura, humedad, radiación, viento) y la geometría del sistema FPV. No requiere los modelos entrenados (.pkl).

Modelo ajustado: combina la estimación física de Penman-Monteith con una corrección aditiva aprendida por el modelo de aprendizaje automático del clúster asignado. Esto es:

$$\text{Evaporación final} = E_{\text{PM}} + \text{Residuo_ML}$$

SOLEVAPP_v1 (física pura) modelo_cluster_N.pkl (corrige lo que Penman no captura)

El residuo ML fue entrenado con mediciones reales de evaporación en bandeja distribuidas a lo largo del territorio chileno. Por lo tanto, el modelo aprende las desviaciones sistemáticas entre las estimaciones de Penman-Monteith y las observaciones en distintas zonas climáticas, vinculadas a efectos de microclima, neblina costera o condiciones de aridez extrema que el modelo físico no logra capturar completamente.

Cobertura parcial del embalse

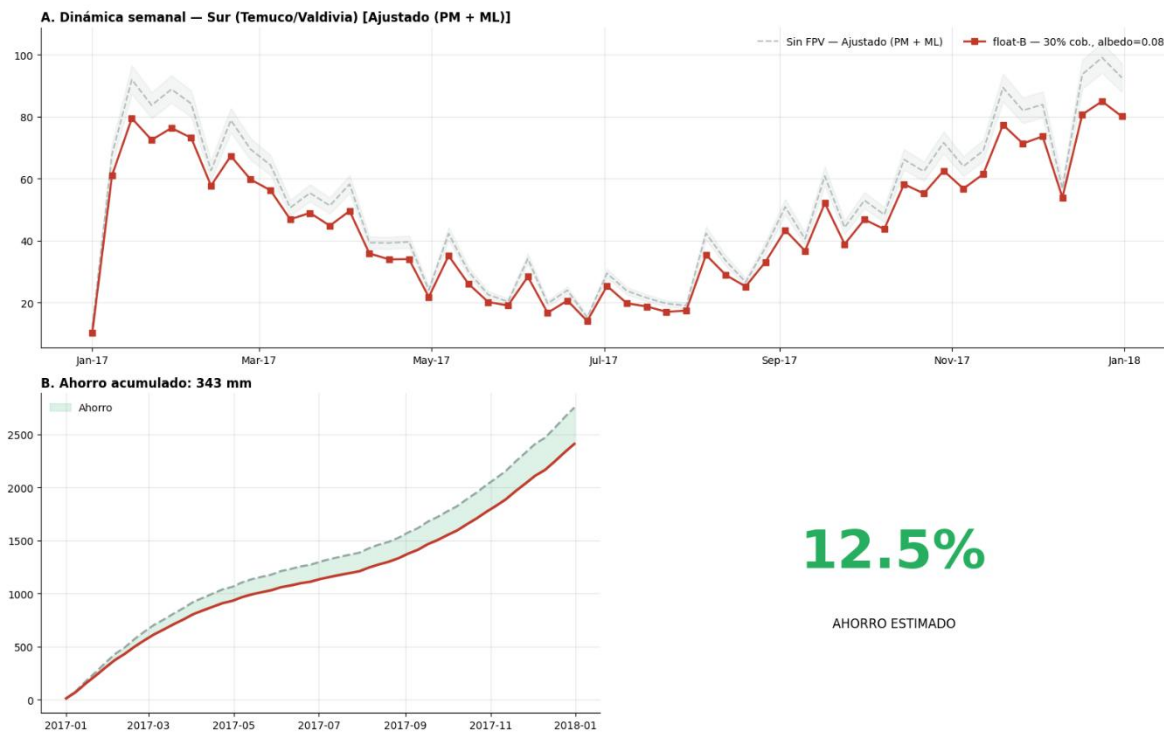
La cobertura parcial se modela como una ponderación espacial entre la evaporación de la superficie libre y la evaporación bajo los paneles:

$$\text{Evap_proyecto} = \text{Evap_base} \times (1 - \text{cobertura}) + \text{Evap_FPV} \times \text{cobertura}$$

De esta manera, un embalse con 30% de cobertura evaporará como superficie libre en el 70% restante, y como superficie bajo paneles en el 30% cubierto.

Resultados generados

La simulación genera un panel de tres gráficos. Esto es:



donde,

- Panel A — Dinámica semanal: serie temporal de evaporación semanal para la línea base (sin FPV) y el proyecto FPV. Si el archivo incluye datos de evaporación observada (columna `evap_S1` o similar), también se muestra la serie observada.
- Panel B — Evaporación acumulada: curvas acumuladas de la línea base y el proyecto FPV, con el área de ahorro resaltada en verde.
- Panel C — Validación o KPI: si hay datos observados, muestra un gráfico de dispersión Real vs Simulado con métricas KGE, R^2 y Bias. Si no hay datos observados, muestra directamente el porcentaje de ahorro estimado.

Junto a los gráficos, la herramienta imprime un resumen numérico con la evaporación total sin FPV, la evaporación total con FPV, el agua conservada en milímetros y el porcentaje de reducción. Si hay datos observados disponibles, también se reporta el KGE del modelo ajustado.

3.3.Reentrenamiento del modelo (uso avanzado): Parte B

Esta sección no es necesaria para el uso normal de SOLEVAPP. Solo es relevante para equipos técnicos que necesiten calibrar el modelo con datos de nuevas estaciones meteorológicas o para nuevas regiones geográficas.

La Parte B del Notebook Maestro replica el pipeline completo del Módulo de Entrenamiento original, utilizando las mismas funciones de `solevapp_utils.py`. Se organiza en tres pasos los cuales ejecutan el flujo de trabajo de ciencia de datos (*pipeline*), desde la ingesta de datos crudos hasta la serialización de los modelos predictivos, tal como muestra la Fig. 3. Este cuaderno es el "laboratorio" del sistema. Aquí no se realizan simulaciones rápidas, sino que se construyen y calibran los modelos que luego usará la herramienta principal.

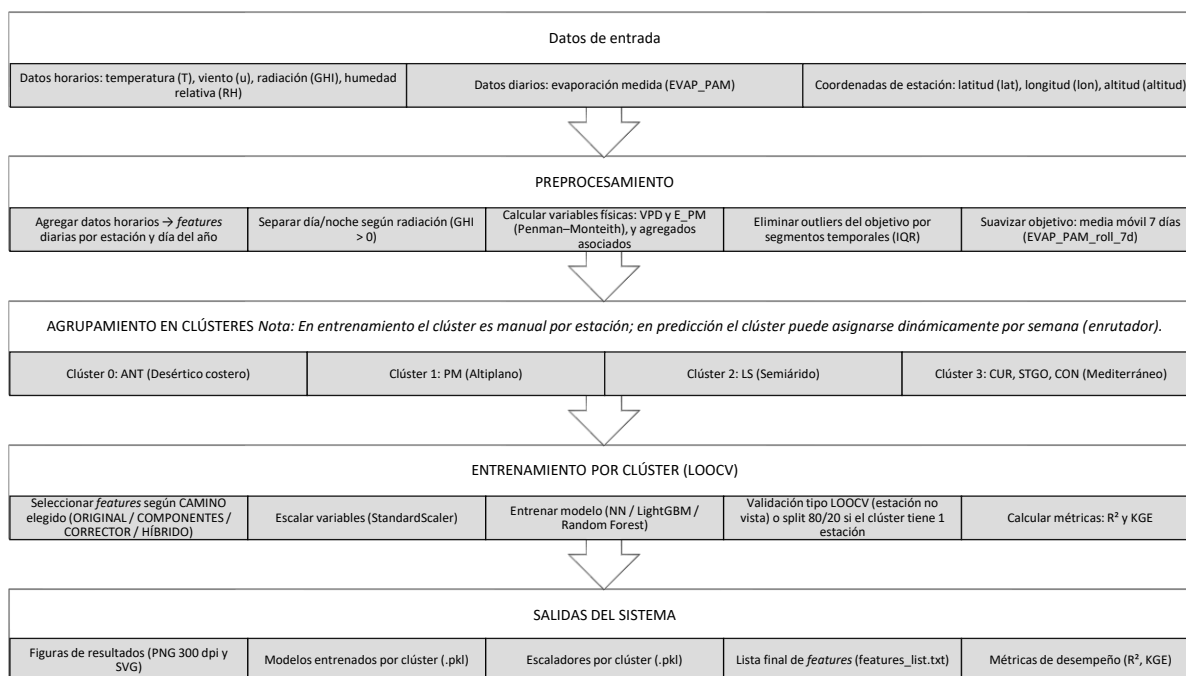


Figura 3: Módulo de entrenamiento desarrollado

3.3.1. Configuración del entrenamiento (Paso T1)

El usuario define los archivos de entrada, la asignación de estaciones a clústeres, el modelo a entrenar y los parámetros de limpieza de datos. Los parámetros configurables son:

- **MODELO_ELEGIDO**: algoritmo de aprendizaje automático a utilizar. Opciones: Neural Network (red neuronal MLP, por defecto), LightGBM o Random Forest.
- **CAMINO_MODELO**: conjunto de variables de entrada al modelo. Opciones: ORIGINAL (variables meteorológicas clásicas), COMPONENTES (variables físicas termodinámicas), CORRECTOR (E_{PM} + variables originales, recomendado) o HÍBRIDO.



WORLD
METEOROLOGICAL
ORGANIZATION



ADAPTATION FUND



udp UNIVERSIDAD
DIEGO PORTALES



— ENANDES —

- **OPTIMIZAR:** si se activa la búsqueda de hiperparámetros mediante RandomizedSearchCV. Valor `s` activa la optimización (más lento); `n` usa los hiperparámetros por defecto.
- **CLUSTER_MAP_MANUAL:** diccionario que asigna cada estación meteorológica a su clúster climático.

La **Sección 2.4** presenta un resumen general de los datos requeridos para SOLEVAPP y para la fase actual de calibración. Esta subsección describe de manera específica los archivos necesarios para ejecutar la etapa de entrenamiento del modelo ajustado.

Para que el modelo pueda aprender, necesita datos provenientes de registros históricos. El primer paso del código consiste en solicitar tres archivos fundamentales; si falta alguno de ellos, el proceso no puede ejecutarse. Estos archivos deben subirse a la sesión o estar en la misma carpeta que el código maestro, todos en formato `.txt`:

- meteo_ajuste.txt:** contiene los datos meteorológicos horarios de las localidades en estudio junto a las estimaciones de evaporación entregados por la herramienta SOLEVAPP. Es decir, incluye los resultados obtenidos en la primera etapa del proyecto, aplicados específicamente a las localidades que disponen de datos experimentales.
- meteo_observado.txt:** corresponde a las mediciones en terreno, es decir, la evaporación real medida en bandeja.
- Est_principales.txt:** contiene la ubicación geográfica de las estaciones, con las coordenadas y la altitud de cada una.

Explicación del Código: El script define una lista de verificación para indicar la información cargada.

```

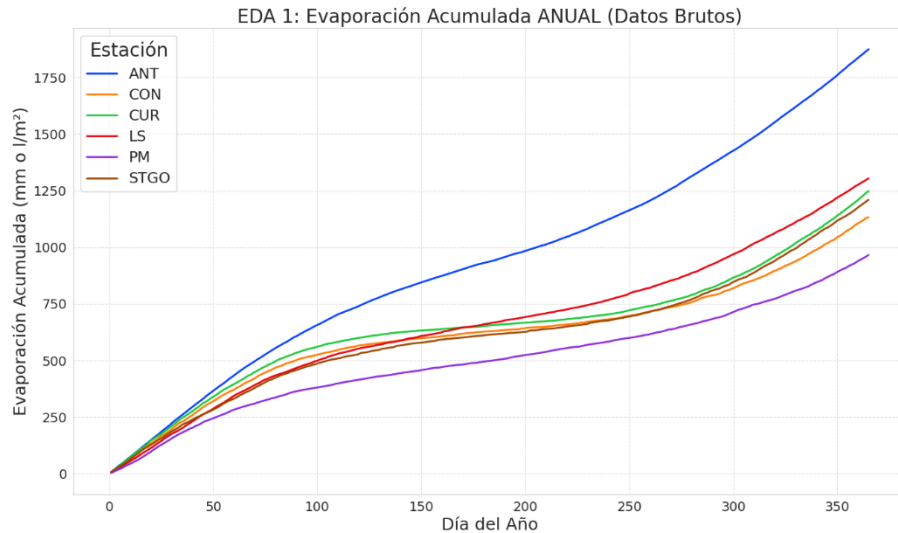
✓ Parámetros de estilo configurados.
Configuración de entrenamiento:
Modelo      : Neural Network
Camino      : CORRECTOR
Optimizar   : s
Clústeres   : {0: ['ANT'], 1: ['PM'], 2: ['LS'], 3: ['CUR', 'STGO', 'CON']}
Figuras     → Resultados_Entrenamiento/Figuras_CORRECTOR_Neural Network
Modelos     → Resultados_Entrenamiento/Modelos_CORRECTOR_Neural Network
  
```

3.3.2. Carga y preparación de datos (Paso T2)

Carga los tres archivos de entrenamiento requeridos y ejecuta el pipeline de preprocesamiento:

1. Carga de datos horarios y diarios con `cargar_y_preparar_datos()`.
2. Generación de features diarios físicos con `crear_features_diarias_fisicas()`.
3. Merge maestro entre features y evaporación observada (EVAP_PAM).
4. Incorporación de coordenadas geográficas desde `Est_principales.txt`.

5. Generación del gráfico diagnóstico Plot 01 (evaporación acumulada bruta por estación). Esto es:



El módulo de entrenamiento ejecuta todo el proceso de construcción del modelo a partir de datos meteorológicos observados. Comienza cargando los datos de entrada, que incluyen series horarias de temperatura, velocidad del viento y radiación, junto con datos diarios de evaporación promedio y las coordenadas geográficas de cada estación (latitud, longitud y altitud). Luego aplica un preprocesamiento que integra y organiza la información para el modelado. En esta etapa se agregan características diarias, se distingue entre condiciones de día y noche según radiación, y se calculan variables derivadas como el déficit de presión de vapor (VPD), la evaporación teórica según Penman-Monteith y la energía térmica disponible. También se corrigen anomalías eliminando los datos atípicos (outliers) que se encuentran fuera de rangos de los intercuartiles, utilizando el método IQR (Interquartile Range), en esta etapa también se suaviza la serie objetivo mediante una media móvil de 7 días.

El detalle del formato de archivos puede encontrarse en el Anexo 1. Las tablas, a continuación, muestran un ejemplo del formato de los datos a cargar:

Tabla 2: Muestra de datos del archivo horario "meteo_ajuste_final_v2.txt" El usuario puede generar su propio archivo con sus datos y modelos

Fecha+Hora	(UTC)	ghi_S1	evap_S1	temp_S1	rh_S1	vel_S1	Mes	dia_del_anio	E_PM_S1	R_n_S1	delta_S1	gamma_S1	e_s_S1	e_a_S1	localidad
01-01-2017	0:00	0	0.014	20.054	0.94218	1.045455587	1	1	0.022621075	0.069677088	0.145163789	0.067311793	2.346109	2.210456978	ANT
01-01-2017	1:00	0	0.022	19.656	0.85132	1.78420023	1	1	0.038520555	0.075982738	0.142066173	0.067286025	2.288949728	1.948628682	ANT
01-01-2017	2:00	0	0.022	19.586	0.86812	2.381132284	1	1	0.039905774	0.075127464	0.141527194	0.067281494	2.279023556	1.978465929	ANT
01-01-2017	3:00	0	0.022	19.01	0.90964	2.024951832	1	1	0.030364489	0.073978958	0.137157508	0.067244241	2.198764582	2.000084214	ANT
01-01-2017	4:00	0	0.022	20.17	0.90804	2.445442644	1	1	0.032779167	0.071413136	0.146077293	0.067319307	2.363001641	2.14570001	ANT

Tabla 3: Muestra de datos del archivo horario "meteo_consolidado_final_def.txt" El usuario puede generar su propio archivo con sus datos medidos de forma experimental.

dia_del_anio	nombre_estacion	latitud	longitud	EVAP_PAM	SOLEVAPP_EXPLORADOR	SOLEVAP_expl_Calculada	localidad
1	Antofagasta	-23.45028	-70.44056	8	0	0	ANT
2	Antofagasta	-23.45028	-70.44056	7	0	0	ANT
3	Antofagasta	-23.45028	-70.44056	6.2	0	0	ANT
4	Antofagasta	-23.45028	-70.44056	6.3	0	0	ANT
5	Antofagasta	-23.45028	-70.44056	7	0	0	ANT

Tabla 4: Muestra de datos del archivo horario "Est_principales.txt" El usuario puede generar su propio archivo con sus datos de estaciones

cod	localidad	nom	lat	lon	cod_est	altitud	est_largo
180005	ARC	DMC	-18.35555	-70.33889	litoral	51	Arica
200006	IQ	DMC	-20.53972	-70.16944	48	Iquique	
230001	ANT	DMC	-23.45028	-70.44056	litoral	139	Antofagasta
270009	COPIAPÒ	DMC	-27.345833	-70.407777	Copiapò		
290004	LS	DMC	-29.9175	-71.20333	secano	146	La Serena
330002	VALPARAISO	DMC	-33.02278	-71.64389	Valparaiso		
330020	STGO	DMC	-33.445	-70.67778	valle	520	Santiago
330031	JUANFERNÁNDEZ	DMC	-33.63583	-78.83028	Juan Fernández		
340031	CUR	DMC	-34.96639	-71.22028	valle	220	Curico
360011	CHILLAN	DMC	-36.58722	-72.03389	Chillan		
360019	CON	DMC	-36.77833	-73.05083	litoral	8	Concepcion
380013	TEM	DMC	-38.76778	-72.62694	Temuco		
390006	VALDIVIA	DMC	-39.65055	-73.06778	Valdivia		
400009	OSORNO	DMC	-40.605	-73.05083	Osorno		
410005	PM	DMC	-41.435	-73.08472	valle	90	Puerto Montt
420014	CASTRO	DMC	-42.3467	-73.71500	Castro		
430002	FUTALEUFU	DMC	-43.18889	-71.86417	Futaleufu		
430004	ALTOPALENA	DMC	-43.61167	-71.81333	Alto Palena		
450001	PUERTOAYSEN	DMC	-45.39944	-72.67778	Puerto Aysen		
450004	COYHAIQUE	DMC	-45.59389	-72.10167	Coyhaique		
450005	BALMACEDA	DMC	-45.91833	-71.67778	Balmaceda		
460001	CHILECHICO	DMC	-46.58083	-71.69472	Chile Chico		
520006	PUNTAARENAS	DMC	-53.00333	-70.84722	Punta Arenas		
270001	RAPANUI	DMC	-27.16055	-109.42361	Rapa Nui		
220002	CALAMA	DMC	-22.49528	-68.915	Calama		

3.3.3. Entrenamiento y guardado de modelos (Paso T3)

Ejecuta el pipeline completo de entrenamiento y genera todos los artefactos necesarios para el módulo de operación:

6. Limpieza de outliers del objetivo con limpiar_datos_objetivo() (método IQR por segmentos temporales).
7. Suavizado de la serie objetivo con media móvil de 7 días.
8. Selección de features según el CAMINO_MODELO configurado.
9. Asignación de clústeres según CLUSTER_MAP_MANUAL.
10. Validación mediante LOOCV (Leave-One-Out Cross-Validation) para clústeres con múltiples estaciones, o split 80/20 para clústeres con una sola estación.
11. Generación de los Plots 02 al 09 de diagnóstico (correlaciones, series de tiempo, métricas R^2 y KGE).

12. Entrenamiento de modelos finales sobre el conjunto completo de datos por clúster.
13. Guardado de artefactos: `modelo_cluster_N.pkl`, `scaler_X_cluster_N.pkl`, `scaler_y_cluster_N.pkl` y `features_list.txt`.

Esta parte del código genera 8 visualizaciones asociadas al entrenamiento. El ejemplo de ellas se encuentra en el anexo.

Los modelos generados en la Parte B quedan almacenados en la carpeta *Resultados_Entrenamiento* de la sesión tal como muestra la figura

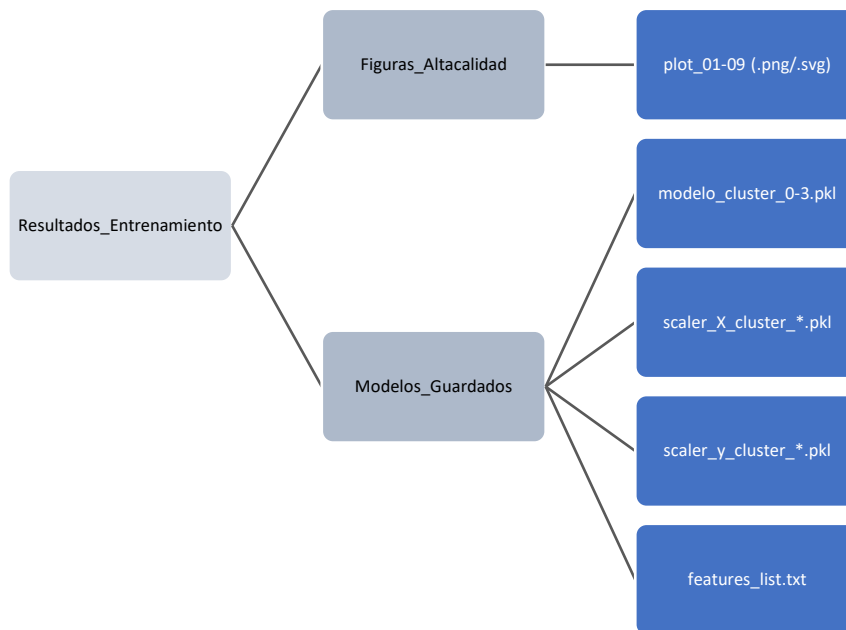


Figura 4. Productos generados por el módulo de entrenamiento.

donde,

<code>modelo_cluster_X.pkl</code>	Contiene el modelo entrenado para el clúster X.
<code>scaler_X_cluster_X.pkl</code>	Transformación utilizada para estandarizar las variables de entrada.
<code>scaler_y_cluster_X.pkl</code>	Transformación utilizada para estandarizar el objetivo del modelo.
<code>features_list.txt</code>	Lista de variables empleadas en el modelo final.

Estos archivos permiten cargar y utilizar los modelos en cualquier entorno Python para realizar predicciones futuras. La descripción de las figuras (*plot_01-09*) se encuentra en el Anexo 2.

4. RECOMENDACIONES TÉCNICAS

La presente versión de SOLEVAPP representa un avance importante respecto de la implementación original, al incorporar la modelación de coberturas parciales y un esquema híbrido que combina un enfoque físico con técnicas de aprendizaje automático basadas en observaciones experimentales. Estas mejoras aumentan la representatividad del modelo para aplicaciones reales y amplían su flexibilidad operativa. En este contexto, es importante tener presente que el cálculo de la evaporación se estructura en dos componentes: la fracción de superficie libre, no cubierta por paneles, y la fracción de superficie cubierta por el sistema fotovoltaico flotante. Dado que ambas responden a condiciones físicas distintas, cada una requiere procesos de ajuste y validación específicos. A continuación, se detallan algunos de los aspectos a considerar para su transferencia a otras regiones.

1. En primer lugar, se recomienda ampliar la base de validación experimental para la fracción de superficie libre. Actualmente, el modelo ajustado incorpora observaciones de evaporación de bandeja dentro del territorio chileno, caracterizado por climas secos y templados. Sin embargo, su robustez puede incrementarse mediante el entrenamiento con datos de estaciones ubicadas en climas tropicales y fríos. Esto permitiría fortalecer la calibración del modelo ajustado para esta componente y mejorar la estabilidad de los clústeres climáticos utilizados en el flujo de Machine Learning. Así, al transferir la herramienta a regiones con condiciones climáticas distintas de las chilenas, se recomienda reentrenar el modelo y redefinir dichos clústeres. En caso de no contar con datos para este reentrenamiento, se recomienda comenzar utilizando el modelo físico y dejar el modelo ajustado para una fase posterior de calibración.
2. En segundo lugar, se recomienda ampliar la base de validación experimental para la fracción de superficie cubierta por paneles. Actualmente, una de las principales limitaciones es la escasez de estudios experimentales que cuantifiquen cómo distintas configuraciones de sistemas fotovoltaicos flotantes afectan la evaporación del agua bajo la superficie cubierta. Esta limitación restringe la posibilidad de ajustar con mayor precisión esa componente del cálculo. Por ello, sería especialmente valioso disponer de datos experimentales para configuraciones como sistemas con revestimiento inferior, sin revestimiento inferior, suspendidos, flexibles, entre otras posibles variantes. Contar con esta información permitiría fortalecer la calibración del modelo y mejorar la representación de la evaporación bajo paneles en distintas condiciones de diseño.
3. En tercer lugar, se debe considerar la resolución temporal de los datos meteorológicos disponibles, ya que esta afecta la representatividad de las estimaciones de evaporación. SOLEVAPP incorpora una lógica de reconstrucción para operar con series diarias o mensuales cuando no se dispone de datos horarios, lo que introduce un nivel adicional de incertidumbre. Por ello, una línea de mejora futura consiste en cuantificar explícitamente la incertidumbre asociada a estas aproximaciones, de modo que el usuario pueda distinguir entre resultados obtenidos a partir de datos horarios medidos y resultados derivados de reconstrucciones sintéticas. En este sentido, al transferir la

herramienta a otras regiones, será importante considerar que la calidad y resolución temporal de variables como la radiación global horizontal, la temperatura ambiente, la humedad relativa y la velocidad del viento influirán directamente en la confiabilidad de las estimaciones.

En conjunto, estos elementos evidencian que el desarrollo futuro de SOLEVAPP debe orientarse a fortalecer su base experimental, ampliar la representación de configuraciones fotovoltaicas flotantes y cuantificar de forma más explícita la incertidumbre asociada a la disponibilidad y resolución de los datos meteorológicos. Avanzar en estas líneas permitirá mejorar la robustez del modelo y facilitar su aplicación confiable en contextos climáticos y operacionales distintos de aquellos para los cuales fue inicialmente ajustado.

5. Guía de Formatos de Archivos

En su etapa actual, el código admite tres tipos de archivos de entrada: horario, diario y mensual. El formato se identifica automáticamente a partir de las columnas presentes. Esto es,

Columna presente	Se detecta como:	Se convierte a:
Fecha+Hora (UTC)	HORARIO	(no se convierte)
date	DIARIO	→ Horario (24h/día)
month	MENSUAL	→ Diario → Horario

Los archivos deben cargarse considerando los siguientes criterios de formato:

- **Separador:** Tabulador (\t) — Guardar como .txt
- **Decimales:** Punto (.) — Ejemplo: 25.65 y no 25,65
- **Encoding:** UTF-8

Unidades de medida⁴:

Variable	Unidad CORRECTA	Unidad INCORRECTA
Radiación (GHI)	MJ/m ²	W/m ² , kWh/m ²
Humedad Relativa	Fracción (0-1)	Porcentaje (0-100)
Viento	m/s (a 2m de altura)	km/h, nudos
Temperatura	°C	°F, K
Evaporación	mm	cm, m

⁴ Conversiones útiles:

- **GHI:** 1 kWh/m² = 3.6 MJ/m²
- **GHI:** W/m² promedio × 0.0036 = MJ/m² (por hora)
- **Humedad:** % ÷ 100 = fracción
- **Viento:** km/h ÷ 3.6 = m/s

5.1.Formato horario (Hourly)

El archivo de datos cargados debe tener la siguiente estructura:

Fecha+Hora (UTC)	ghi_S1	temp_S1	rh_S1	vel_S1	Mes	evap_S1
01-01-2017 0:00 0	15.5	0.85	1.2	1		0.018
01-01-2017 1:00 0	14.8	0.87	1.1	1		0.015
01-01-2017 2:00 0	14.2	0.89	0.9	1		0.012
...						
01-01-2017 12:00	3.25	22.5	0.55	2.8	1	0.42
01-01-2017 13:00	3.10	23.1	0.52	3.1	1	0.45

El código revisa la columna “Fecha+Hora (UTC)”, cuya información debe ingresarse en el formato mostrado en la siguiente tabla descriptiva:

Columna	Descripción	Unidades	Ejemplo
Fecha+Hora (UTC)	Fecha y hora	DD-MM-YYYY H:MM	01-01-2017 14:00
ghi_S1	Radiación solar global horizontal	MJ/m ² (por hora)	2.45
temp_S1	Temperatura del aire	°C	18.5
rh_S1	Humedad relativa	Fracción (0-1)	0.65
vel_S1	Velocidad del viento a 2m	m/s	2.15
Mes	Número del mes (1-12)		1

Columnas Opcionales (para validación):

Columna	Descripción	Unidades
evap_S1	Evaporación medida (bandeja)	mm/hora
E_PM_S1	Evaporación Penman-Monteith	mm/hora
dia_del_anio	Día del año (1-365)	

5.2.Formato diario (daily)

El archivo de datos cargados debe tener la siguiente estructura:

date	ghi	evap	t_mean	t_min	t_max	rh_mean	wind_mean	Mes	dia_del_anio	
01-01-2017		25.65	4.12	16.2	9.2	20.5	0.72	2.15	1	1
02-01-2017		27.32	4.58	17.1	10.1	21.3	0.68	1.89	1	2
03-01-2017		22.18	3.45	15.8	9.8	19.2	0.78	2.45	1	3
04-01-2017		28.91	4.89	18.2	11.2	22.8	0.62	1.72	1	4

El código revisa la columna “date”, cuya información debe ingresarse en el formato mostrado en la siguiente tabla descriptiva:

Columna	Descripción	Unidades	Ejemplo
date	Fecha	DD-MM-YYYY	01-01-2017
ghi	Radiación solar diaria total	MJ/m ² /día	25.6
t_mean	Temperatura media diaria	°C	17.2
rh_mean	Humedad relativa media (0-1)		0.68
wind_mean	Velocidad viento media	m/s	2.1

Columnas Opcionales (recomendadas):

Columna	Descripción	Unidades
t_min	Temperatura mínima	°C
t_max	Temperatura máxima	°C
evap	Evaporación medida (bandeja)	mm/día
Mes	Número del mes (1-12)	
dia_del_anio	Día del año (1-365)	

5.3.Formato mensual (monthly)

El archivo de datos cargados debe tener la siguiente estructura:

month	ghi	evap	t_mean	t_min	t_max	rh_mean	wind_mean
1	25.8	4.25	17.2	10.5	22.1	0.68	2.05
2	23.5	3.85	16.8	10.2	21.5	0.70	2.15
3	18.2	2.95	14.5	8.5	18.8	0.75	2.35
4	12.8	1.85	11.2	5.8	15.2	0.82	2.55
5	8.5	1.15	8.5	3.5	12.2	0.88	2.85
6	6.2	0.78	6.2	1.8	9.5	0.92	3.05
7	6.8	0.85	5.8	1.2	9.2	0.91	2.95
8	9.5	1.25	7.5	2.5	11.2	0.87	2.75
9	14.2	2.15	10.2	4.8	14.5	0.80	2.45
10	19.5	3.25	13.5	7.2	18.2	0.74	2.25
11	23.8	4.05	15.8	9.2	20.5	0.69	2.08
12	26.5	4.45	17.5	10.8	22.5	0.66	1.98

El código revisa la columna “month”, cuya información debe ingresarse en el formato mostrado en la siguiente tabla descriptiva:

Columna	Descripción	Unidades	Ejemplo
month	Número del mes	1-12	1
ghi	Radiación solar promedio diaria	MJ/m ² /día	25.8
t_mean	Temperatura media mensual	°C	17.2
rh_mean	Humedad relativa media	Fracción (0-1)	0.68
wind_mean	Velocidad viento media	m/s	2.05

Columnas Opcionales (recomendadas):

Columna	Descripción	Unidades
t_min	Temperatura mínima promedio	°C
t_max	Temperatura máxima promedio	°C
evap	Evaporación promedio diaria	mm/día

Validación opcional:

- Si el usuario tiene datos de evaporación medida (evap o evap_S1), puede incluirlos para validar el modelo
- Si el usuario NO tiene datos de validación, desmarca ☐ "¿Usar datos de validación?" en el simulador.

5.4.Errores comunes y soluciones

• Separador incorrecto

✗ INCORRECTO (coma o punto y coma):

date,ghi,t_mean,rh_mean

01-01-2017,25.65,16.2,0.72

✓ CORRECTO (tabulador):

date ghi t_mean rh_mean

01-01-2017 25.65 16.2 0.72

Sugerencia: Guardar siempre como "Texto delimitado por tabulaciones" (.txt)

• Decimal con coma (formato español)

✗ INCORRECTO:

25,65 16,2 0,72

✓ CORRECTO:

25.65 16.2 0.72

Sugerencia: Buscar y reemplazar , por . en los valores numéricos

- **Humedad relativa en porcentaje**

✗ INCORRECTO (0-100):

rh_mean
72
68
78

✓ CORRECTO (0-1):

rh_mean
0.72
0.68
0.78

Sugerencia: Dividir valores por 100

- **Radiación en unidades incorrectas**

✗ INCORRECTO (W/m² instantáneo):

ghi
850
920
780

✓ CORRECTO (MJ/m² acumulado):

ghi
3.06
3.31
2.81

Sugerencia horaria: W/m² × 0.0036 = MJ/m²

Sugerencia diaria: kWh/m² × 3.6 = MJ/m²

- **Formato de fecha incorrecto**

✗ INCORRECTO:

2017-01-01 (YYYY-MM-DD)
01/01/2017 (con barras)
1-1-2017 (sin ceros)

✓ CORRECTO:

01-01-2017 (DD-MM-YYYY con guiones)

Sugerencia: Reformatear fechas a DD-MM-YYYY

- **Nombre de columnas incorrecto**

✗ INCORRECTO:

fecha radiacion temperatura humedad
Date GHI Temp RH

✓ CORRECTO:

date ghi t_mean rh_mean

Sugerencia: Renombrar columnas exactamente como indica la guía (minúsculas, sin tildes)

- **Datos faltantes sin marcar**

✗ INCORRECTO (celda vacía o texto):

25.65 16.2
27.32 N/A 0.68

✓ CORRECTO (NaN o valor numérico):

25.65 16.2 NaN
27.32 17.1 0.68

Sugerencia: Reemplazar vacíos por NaN o interpolar valores

- **Viento a altura incorrecta**

✗ INCORRECTO (10m de altura - típico de estaciones):

wind_mean
4.2

✓ CORRECTO (2m de altura):

wind_mean
3.0

Sugerencia: Aplicar corrección logarítmica:

$$u_2 = u_{10} \cdot \left(\frac{4.87}{\ln(67.8 \times 10 - 5.42)} \right) \approx u_{10} \times 0.75 \quad (\text{A.2})$$

- **Archivo con *encoding* incorrecto**

✗ INCORRECTO:

- Caracteres raros: ã±, ã©, â€
- Error al cargar archivo

✓ CORRECTO:

- Encoding UTF-8

Sugerencia: Guardar archivo con encoding UTF-8

- **Mezcla de formatos en un archivo**

✗ INCORRECTO:

date ghi Fecha+Hora (UTC)
01-01-2017 25.65 01-01-2017 12:00

✓ CORRECTO:

Elegir UN solo formato (date O Fecha+Hora, no ambos)

Sugerencia: Usar solo las columnas del formato elegido



✓ **CHECKLIST SUGERIDO ANTES DE USAR DATOS PROPIOS**

- ✓ ☐ Separador es tabulador (no coma ni punto y coma)
- ✓ ☐ Decimales con punto (no coma)
- ✓ ☐ Humedad relativa entre 0 y 1
- ✓ ☐ Radiación en MJ/m²
- ✓ ☐ Viento en m/s a 2m de altura
- ✓ ☐ Fecha en formato DD-MM-YYYY
- ✓ ☐ Nombres de columnas exactos (minúsculas)
- ✓ ☐ Sin filas vacías al inicio o final
- ✓ ☐ Encoding UTF-8

6. Figuras generadas por el módulo de entrenamiento

Al ejecutar la celda principal del pipeline, se genera automáticamente la estructura descrita dentro de la carpeta `Resultados_ModeloEntrenamiento`. El proceso produce nueve figuras en alta calidad, cada una con un propósito específico dentro del análisis exploratorio y de la validación del modelo. Todas se exportan en formato PNG (300 dpi) y en SVG vectorial, adecuado para publicación.

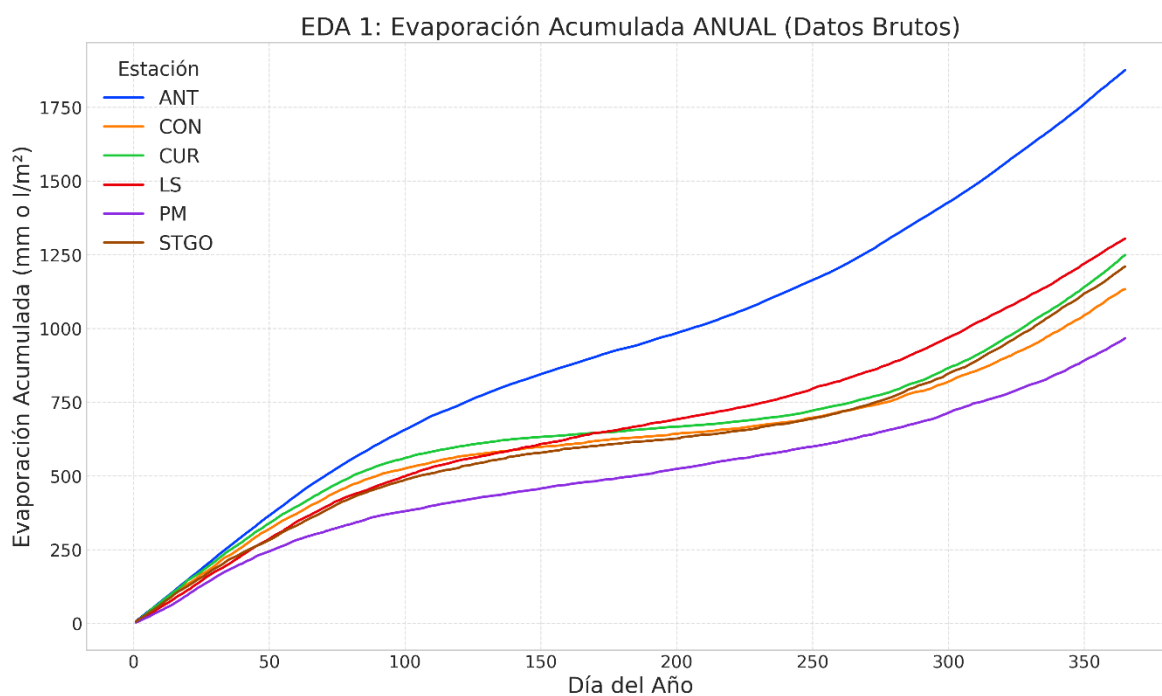
6.1. Evaporación Acumulada Anual (Datos Brutos)

Archivo: `plot_01_acumulado_bruto.png/svg`

Función: `plot_01_acumulado_bruto(df, path_base)`

Contenido: muestra la evaporación acumulada día a día a partir de datos sin procesar ni suavizar.

Utilidad: ofrece una visión general de la magnitud del fenómeno, permite identificar estaciones con mayor evaporación anual y sirve para detectar duplicados y errores importantes.



6.2. Correlación entre Features y Objetivo

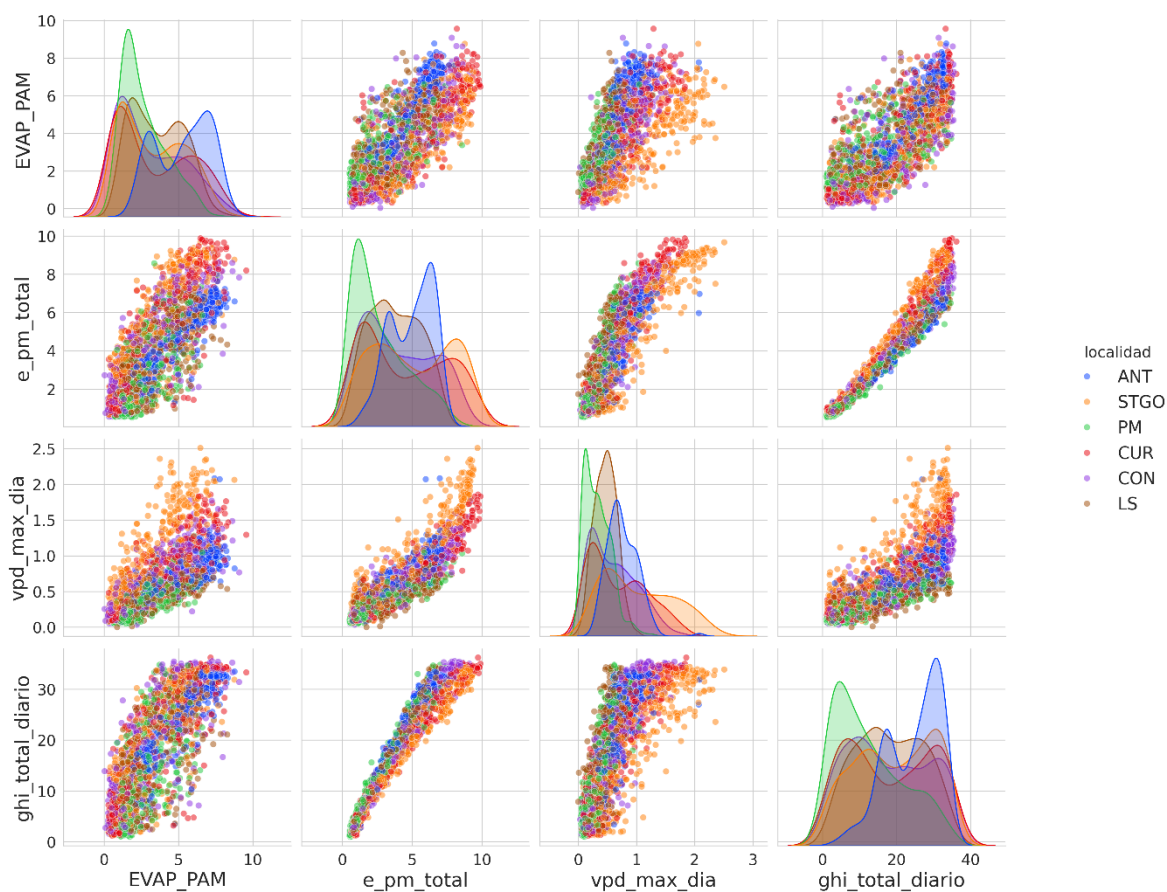
Archivo: `plot_02_correlacion_features.png/svg`

Función: `plot_02_correlacion_features(df_limpio, path_base)`

Contenido: par de gráficos que muestran correlaciones entre variables como GHI, VPD, radiación y humedad, junto con la evaporación diaria EVAP_PAM.

Utilidad: identifica relaciones lineales y no lineales, justifica la selección de características según el CAMINO_MODELO y permite detectar redundancias.

EDA 2: Correlación entre Features Clave y Objetivo (Datos Diarios Limpios)



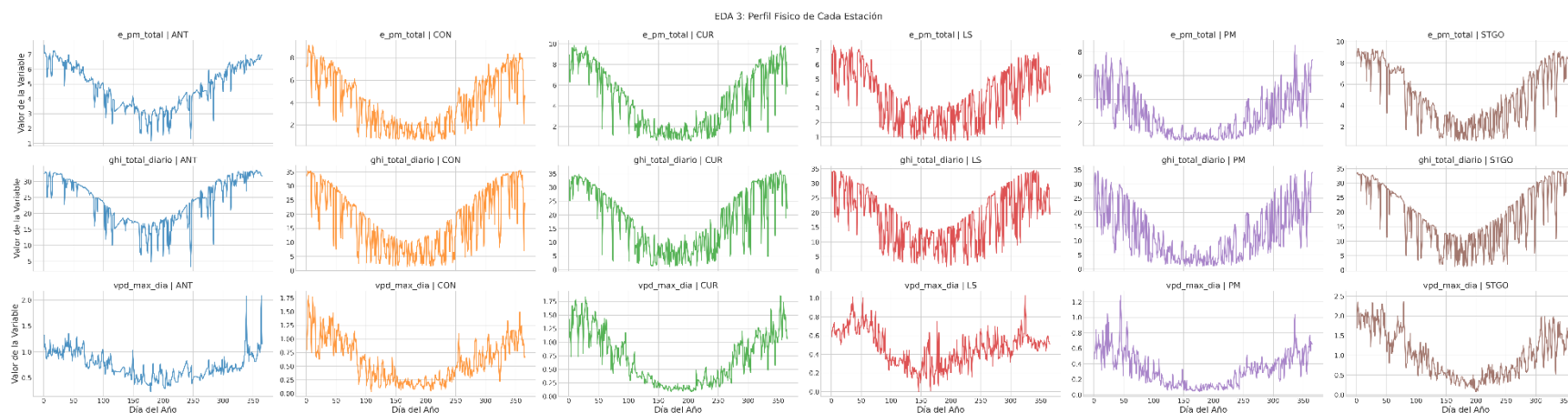
6.3.Series de Tiempo por Estación

Archivo: `plot_03_series_tiempo.png/svg`

Función: `plot_03_series_tiempo_features(df, path_base)`

Contenido: series anuales de variables relevantes separadas por estación.

Utilidad: sustenta el agrupamiento por clúster, permite identificar estacionalidades y revisar la coherencia de los datos.



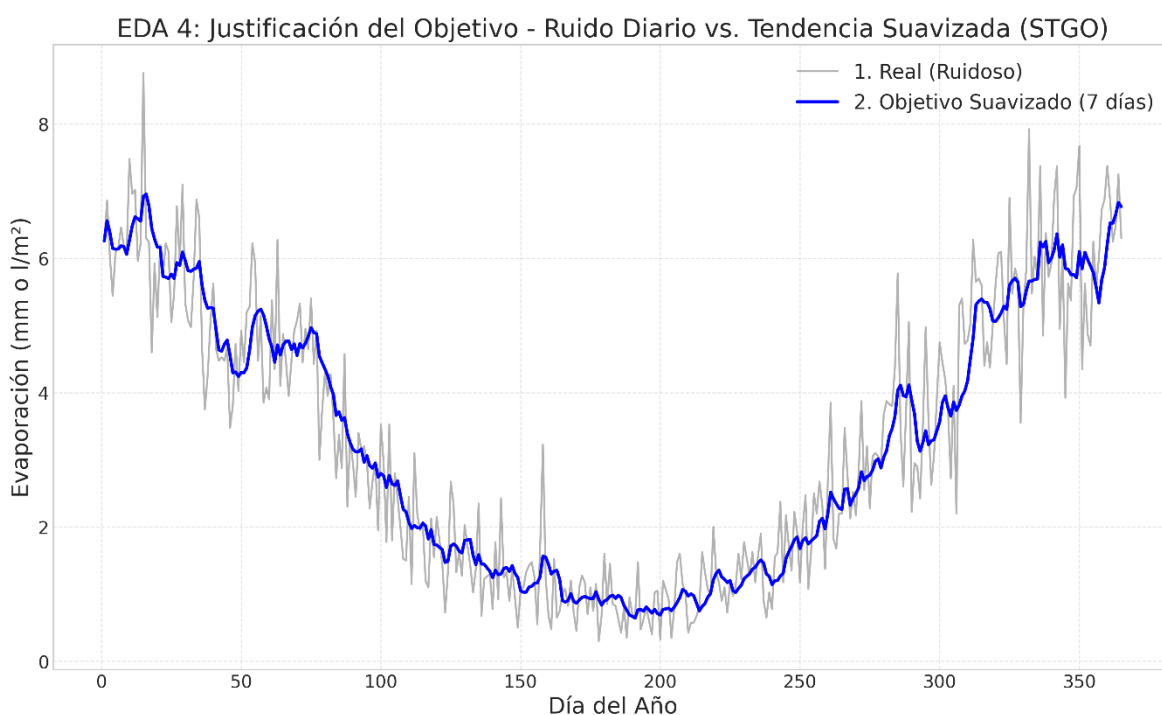
6.4. Justificación del Suavizado del Objetivo

Archivo: plot_04_justificacion_suavizado.png/svg

Función: plot_04_justificacion_suavizado(df, path_base)

Contenido: comparación entre la evaporación diaria real (ruidosa) y la media móvil de 7 días.

Utilidad: demuestra la necesidad de suavizar la señal para evitar que el modelo aprenda ruido meteorológico puntual.



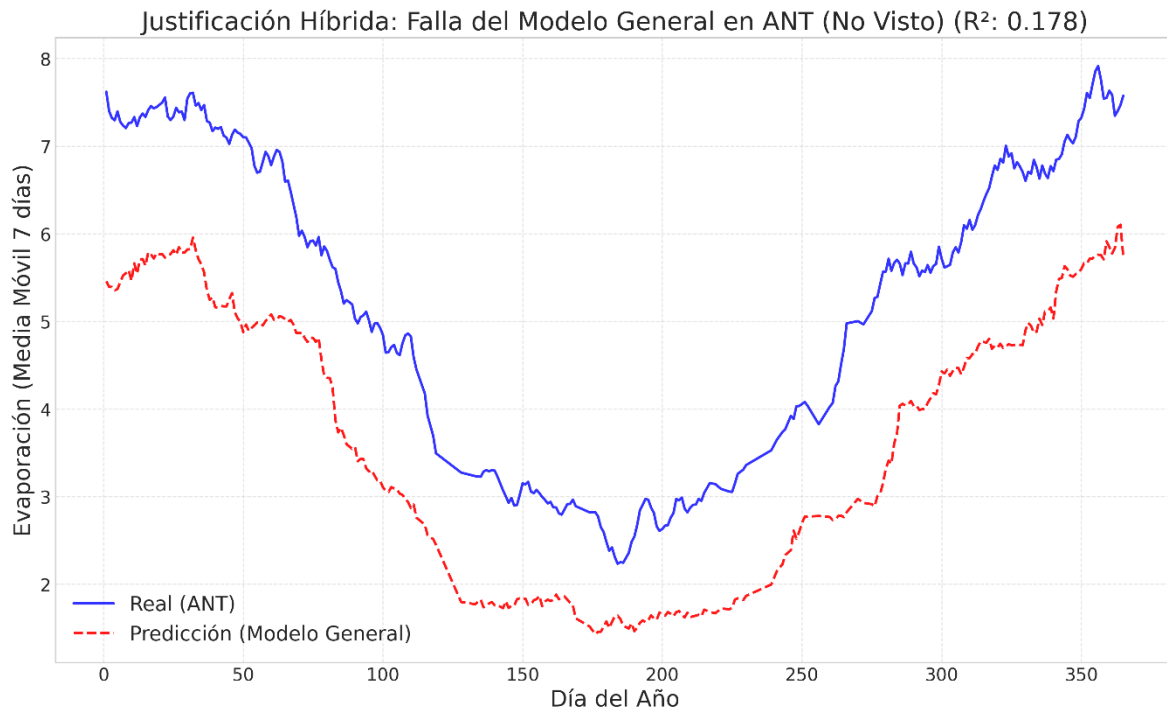
6.5.Falla del Modelo General en ANT

Archivo: plot_05_falla_modelo_general_ANT.png/svg

Función: plot_05_falla_modelo_general_ANT(df, features, path_base, target)

Contenido: comparación entre un modelo global y los datos reales de Antofagasta.

Utilidad: muestra por qué no es posible entrenar un solo modelo para todas las estaciones y motiva el enfoque por clúster y el uso del modelo híbrido.



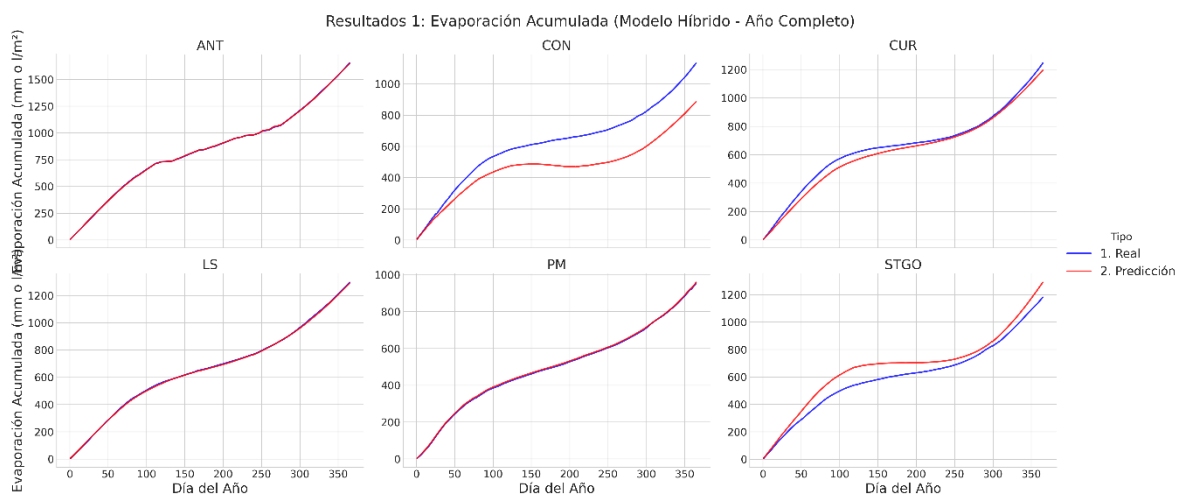
6.6. Evaporación Acumulada (Real vs ML)

Archivo: plot_06_resultados_acumulado.png/svg

Función: plot_06_resultados_acumulado(df_plot, path_base)

Contenido: comparación entre evaporación real acumulada (línea azul) y predicción ML acumulada (línea roja).

Utilidad: muestra tendencias y errores acumulados, y permite identificar estaciones bien modeladas o con dificultades.



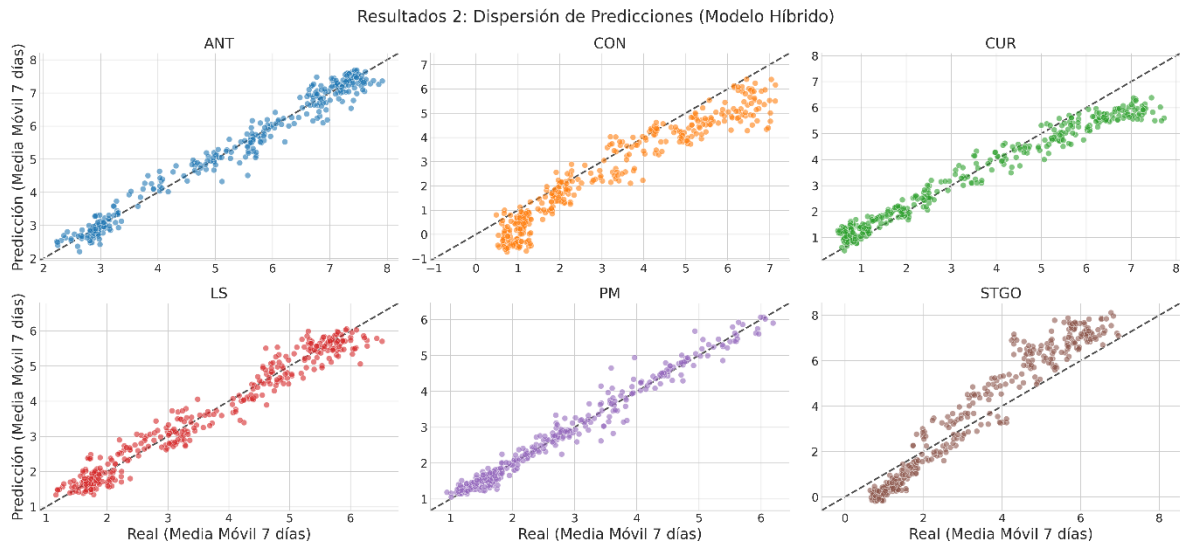
6.7.Dispersión Real vs Predicho

Archivo: `plot_07_resultados_dispersion.png/svg`

Función: `plot_07_resultados_dispersion(df_plot, path_base)`

Contenido: gráfico de dispersión con línea 1:1

Utilidad: permite evaluar el sesgo del modelo y es un soporte visual útil para presentaciones.



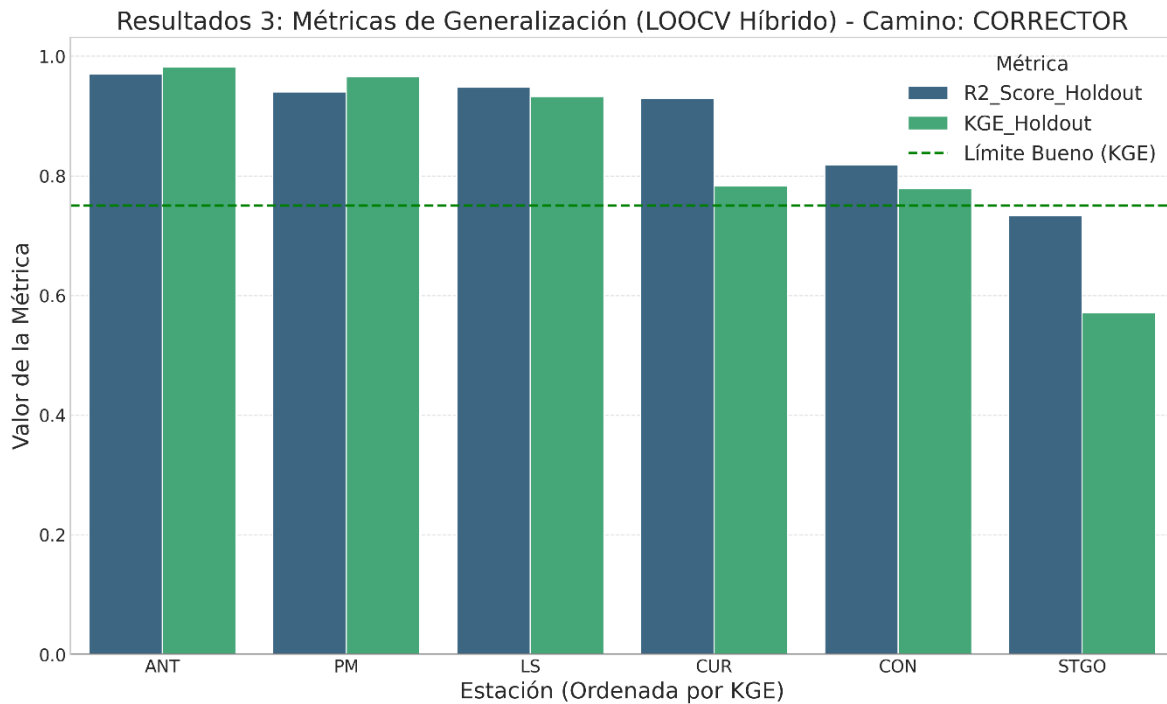
6.8.Métricas (R^2 y KGE) por Estación

Archivo: plot_08_resultados_metricas.png/svg

Función: plot_08_resultados_metricas(df_summary, path_base)

Contenido: gráfico de barras por estación, con línea verde que indica el umbral de buen desempeño $KGE \geq 0.75$.

Utilidad: permite comparar el desempeño entre estaciones y documentar los resultados del modelo.



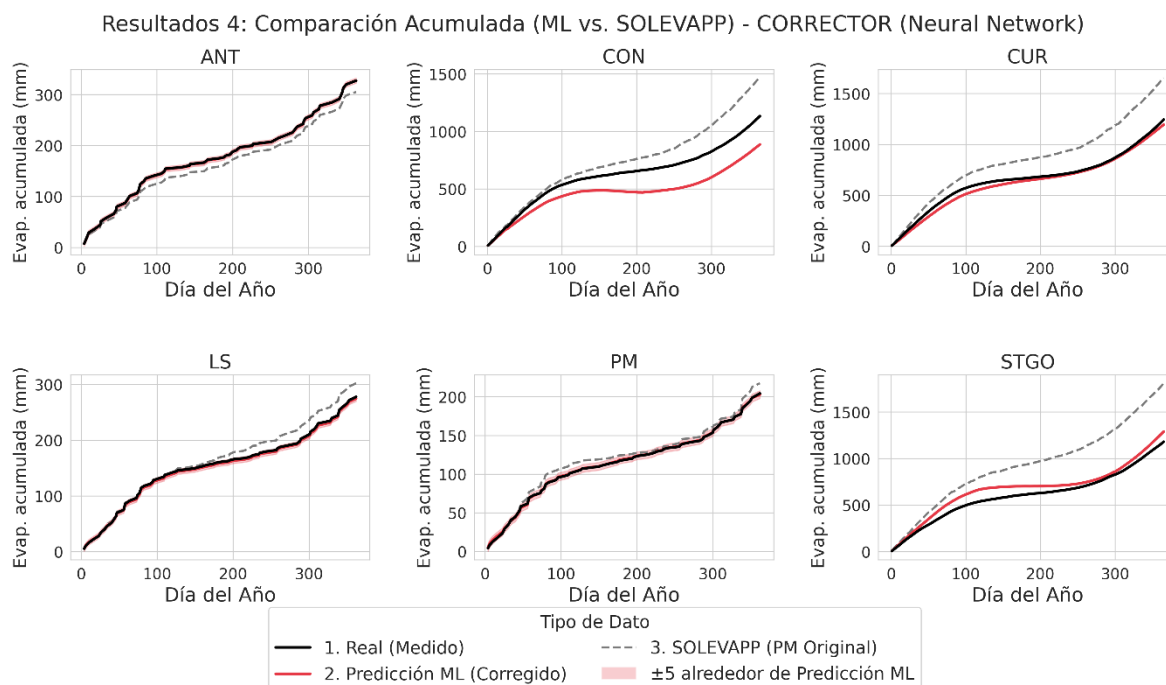
6.9. Comparación Acumulada Real vs ML vs PM

Archivo: plot_09_comparacion_acumulada.png/svg

Función: plot_09_comparacion_acumulada(df_plot, path_base)

Contenido: evaporación real (negro), predicción ML (rojo, con banda de incertidumbre de ± 5 unidades) y estimación Penman–Monteith (gris punteado).

Utilidad: ilustra las mejoras del modelo de machine learning respecto al método físico, siendo una de las figuras más relevantes para reportes.



6.10. Material de apoyo para validación del entrenamiento

El notebook incluye visualizaciones adicionales que permiten verificar la calidad del agrupamiento y detectar oportunidades de mejora. Dichas visualizaciones se muestran a continuación:

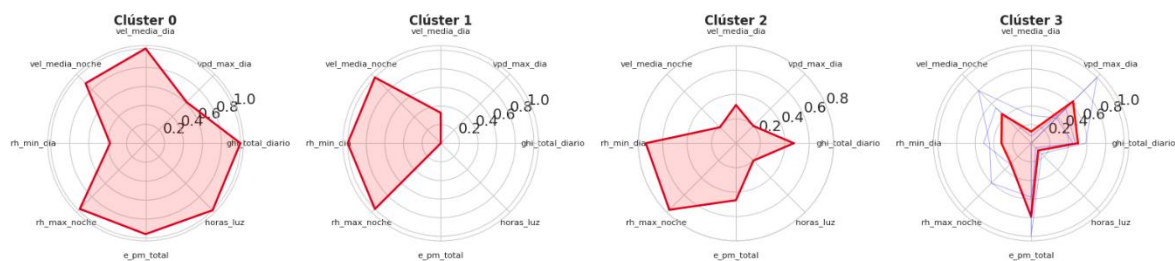


Figure A 1: Radar plot

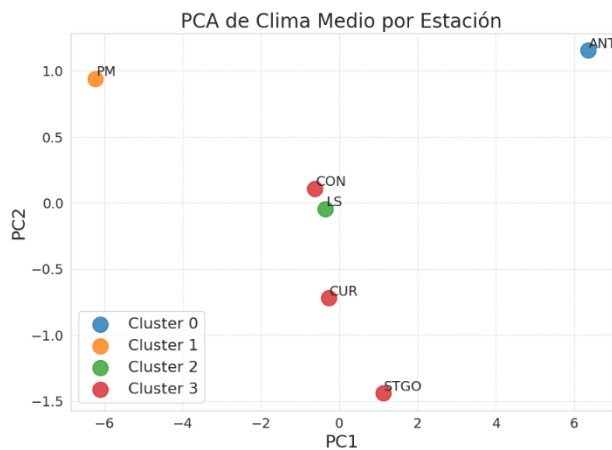


Figure A 2: Análisis PCA por clúster

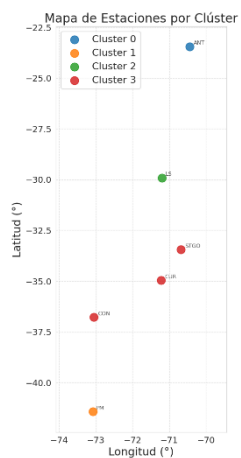


Figure A 3 Mapa geográfico de clúster

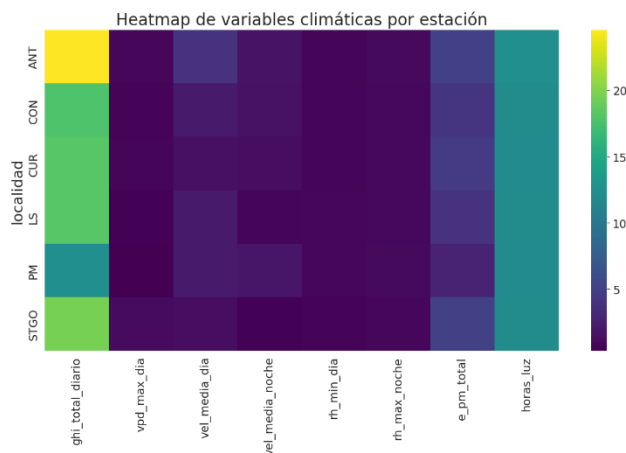


Figure A 4: Heat map de variabilidad de variables climáticas por localidad

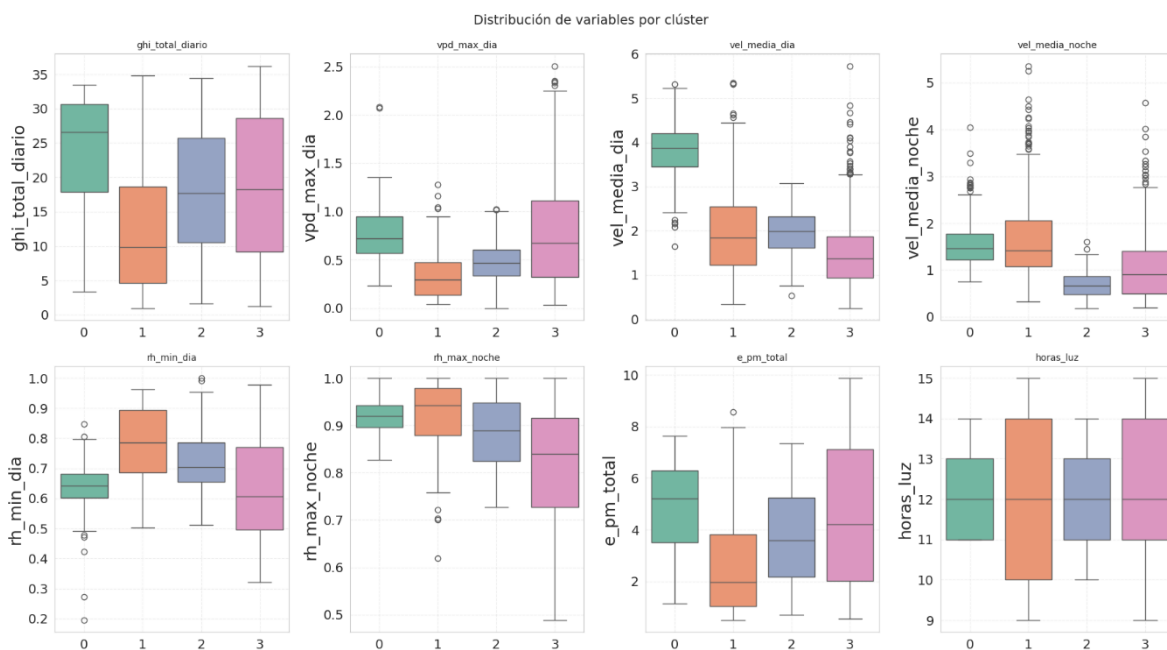


Figure A 5: Box plot de variables climatológicas por cluster

7. Procedimiento de Reconstrucción de Datos Diarios y mensuales

7.1. Detección automática del tipo de archivo

Al cargar los datos, la herramienta identifica automáticamente su formato y los clasifica como:

- HORARIO: se utilizan directamente.
- DIARIO: requieren reconstrucción de 24 valores por día.
- MENSUAL: requieren generar primero una serie diaria sintética y luego una serie horaria.

La reconstrucción es necesaria porque el modelo físico SOLEVAPP funciona exclusivamente con datos horarios, los modelos de inteligencia artificial fueron entrenados utilizando características que se derivan también de información horaria y, además, el enrutador climático depende de indicadores que solo pueden calcularse correctamente a partir de series horarias, como la radiación acumulada (GHI).

7.2. Reconstrucción desde datos diarios

Variable	Cómo se reconstruye	Comportamiento asumido	Objetivo dentro del flujo
Radiación global horizontal (GHI)	Se distribuye a lo largo del día mediante una curva solar horaria	La radiación aumenta desde el amanecer, alcanza un máximo cercano al mediodía solar y disminuye hacia el atardecer. Para cada fecha se estiman las horas de amanecer y atardecer	Representar la disponibilidad energética horaria necesaria para los cálculos físicos y para la generación de variables derivadas
Temperatura del aire	Se modela mediante una función sinusoidal suavizada	La temperatura aumenta desde la mañana, alcanza un máximo durante la tarde y desciende durante la noche, respetando los valores mínimos y máximos diarios observados	Reproducir el ciclo térmico diario usado en la estimación horaria de evaporación
Humedad relativa	Se reconstruye con un patrón horario inversamente relacionado con la temperatura	La humedad relativa presenta valores más altos en la madrugada y más bajos durante las horas de mayor temperatura	Mantener coherencia atmosférica entre temperatura y humedad para el cálculo posterior de evaporación

Velocidad del viento	Se distribuye con un perfil horario simple, consistente con el comportamiento diario típico	El viento suele presentar valores menores durante la mañana y un aumento progresivo hacia la tarde, patrón frecuente en climas mediterráneos y semiáridos	Generar una señal horaria plausible para los cálculos físicos y para la construcción de características del modelo ajustado
----------------------	---	---	---

Con estas variables reconstruidas, la herramienta estima también una versión horaria de la evaporación Penman–Monteith. Aunque esta señal es aproximada, conserva la dinámica horaria necesaria para las etapas posteriores del flujo. El resultado final es una serie de 24 valores por día con coherencia física suficiente para aplicar el enrutador climático, generar las características diarias requeridas por el modelo ajustado e integrar los cálculos de evaporación con o sin paneles flotantes.

7.3. Reconstrucción desde datos mensuales

Cuando los datos están en formato mensual, el desafío es mayor porque solo existen doce valores al año. Dado que muchos catastros hidrometeorológicos utilizan este formato, la herramienta permite trabajar también con este tipo de información.

El método considera que cada valor mensual corresponde al promedio climatológico del mes. A partir de este valor, la herramienta genera una serie diaria sintética mediante:

$$X_d = \bar{X}_{mes} (1 + s e) \quad (3)$$

donde: $\bar{X}_{mes}$ es el promedio mensual,
 $e: N(0,1)$ es ruido aleatorio normal,
 s es un coeficiente de variabilidad calibrado para cada variable.

Los rangos de variabilidad utilizados son:

Variable	Unidad	Variabilidad (σ)
Radiación	MJ/m ² /día	± 20%
Viento	m/s	± 25%
Evaporación	mm/día	± 15%
Temperatura media	°C	± 2°C

Además, se aplican límites físicos para evitar valores imposibles. Por ejemplo:

- humedad relativa entre 0.30 y 0.95,
- viento entre 0.5 m/s y el doble del promedio mensual,
- radiación entre 0.3 y 1.5 veces el valor mensual.

7.4. Radiación Solar Diaria

El código asume que la radiación varía proporcionalmente alrededor del promedio, usando una distribución normal (campana de Gauss).

$$GHI_d = \overline{GHI}_{mes} \cdot (1 + 0.2 \cdot N(0,1)) \quad (A.4)$$

donde $N(0,1)$ es un número aleatorio de una distribución normal estándar (media 0, desviación 1). Además, el código fuerza a que el resultado nunca sea menor al 30% ni mayor al 150% del promedio mensual para evitar generar outliers.

7.5. Temperatura Media Diaria

La temperatura varía de forma aditiva respecto al promedio mensual (más alta o más baja).

$$T_{mean,d} = \bar{T}_{mes} + 2.0 \cdot N(0,1) \quad (A.5)$$

donde el factor 2.0 indica que la desviación estándar típica de la temperatura diaria es de 2°C respecto al promedio del mes.

7.5.1. Amplitud Térmica Diaria

Para calcular las mínimas y máximas, el código varía el rango térmico usando una distribución uniforme.

$$T_{rango,d} = (\bar{T}_{max} - \bar{T}_{min}) \cdot (0.8 + 0.4 \cdot U(0,1)) \quad (A.6)$$

donde $U(0,1)$ es un número aleatorio entre 0 y 1. Con estos valores se calculan

$$\begin{aligned} T_{min,d} &= T_{mean,d} - \frac{T_{rango,d}}{2} \\ T_{max,d} &= T_{mean,d} + \frac{T_{rango,d}}{2} \end{aligned} \quad (A.7)$$

7.6. Velocidad del Viento y Humedad

Siguen la misma lógica proporcional que la radiación, con distintos coeficientes de variabilidad o ruido

$$\begin{aligned} u_{2,d} &= \bar{u}_{2,mes} \cdot (1 + 0.25 \cdot N(0,1)) \\ RH_d &= \overline{RH}_{mes} + 0.1 \cdot N(0,1) \end{aligned} \quad (A.8)$$

Los datos mensuales, ya convertidos a diario, son procesados en la función *daily_to_hourly()*. El objetivo de esta función es reconstruir un día de datos.

7.7. Radiación

Para pasar de un total diario de radiación a una curva horaria, el código simula la trayectoria del sol. Primero, calcula a qué hora sale y se pone el sol basándose en el día del año (J).

$$\begin{aligned} H_{salida} &\approx 6 + 2 \cdot \cos\left(\frac{2\pi \cdot J}{365}\right) \\ H_{puesta} &\approx 18 - 2 \cdot \cos\left(\frac{2\pi \cdot J}{365}\right) \end{aligned} \quad (\text{A.9})$$

En invierno ($J \approx 180$), el día es corto (ej. 08:00 a 16:00), mientras que en verano ($J \approx 1$ o 365), el día es largo (ej. 06:00 a 20:00). Luego, la radiación instantánea a la hora h sigue una curva sinusoidal modificada, que es cero durante la noche.

$$f_{solar}(h) = \begin{cases} \sin\left(\pi \cdot \frac{h - H_{salida}}{H_{puesta} - H_{salida}}\right)^{1.2} & \text{si } H_{salida} \leq h \leq H_{puesta} \\ 0 & \text{si es noche} \end{cases} \quad (\text{A.10})$$

Finalmente, se reparte el total diario proporcionalmente:

$$GHI_h = GHI_d \cdot \frac{f_{solar}(h)}{\sum_{t=0}^{23} f_{solar}(t)} \quad (\text{A.11})$$

7.8. Temperatura

El modelo asume un comportamiento sinusoidal.

$$T_h = T_{media} + \frac{T_{max} - T_{min}}{2} \cdot \sin\left(\frac{2\pi \cdot (h-9)}{24}\right) \quad (\text{A.12})$$

donde la amplitud es la mitad de la diferencia entre máxima y mínima, la fase se desplaza hacia las 3 PM ($h-9$).

7.9. Humedad Relativa

La humedad relativa se comporta de forma similar a la temperatura.

$$RH_h = RH_{media} + A_{rh} \cdot \cos\left(\frac{2\pi \cdot (h-4)}{24}\right) \quad (\text{A.13})$$

El código aplica un límite físico para que $0.2 \leq RH_h \leq 0.99$.

7.10. Velocidad del Viento

Se asume una variación del 30% alrededor del promedio.

$$u_{2,h} = u_{2,media} + (0.3 \cdot u_{2,media}) \cdot \sin\left(\frac{2\pi \cdot (h-12)}{24}\right) \quad (\text{A.14})$$

8. Funciones del entrenamiento

8.1. Preprocesamiento: De Horas a Días

Aunque la frecuencia de registro de los datos meteorológicos es horaria, la evaporación se modela a escala diaria para reducir el ruido asociado a procesos instantáneos. En la Fig. A6 se observa este efecto para Santiago: la línea gris corresponde a la señal real, altamente variable, mientras que la línea azul muestra la serie de tiempo suavizada (objetivo) mediante una media móvil de 7 días, revelando la tendencia subyacente. Por lo tanto, la herramienta ajusta estos datos para extraer una señal diaria más estable.

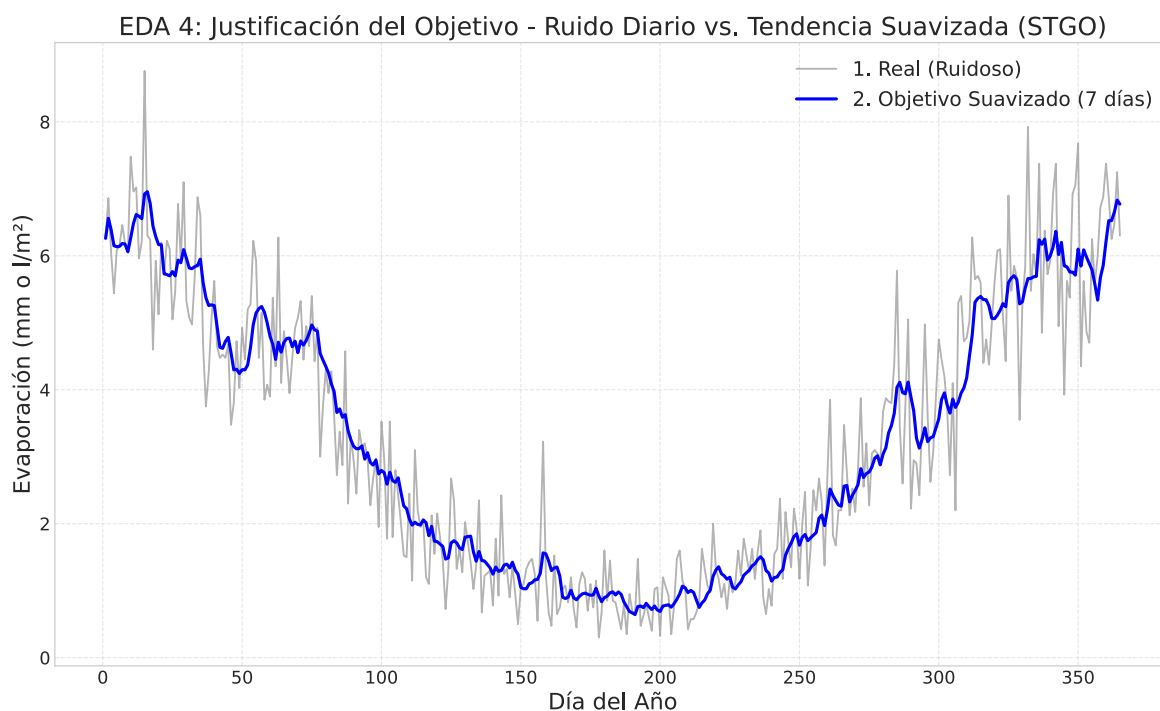


Figura A 6. Comparación entre la señal diaria original y la evaporación suavizada (7 días) en Santiago.

Función Clave: `crear_features_diarias_fisicas(df)`

La lógica detrás de la función se aplica en tres pasos:

- i. Suma la radiación que llega a la superficie horizontal durante todo el día (GHI), obteniendo así la energía diaria expresada en MJ/(m²·día).

$$R_{s,dia} = \sum_{h=0}^{23} GHI_h \quad \left[\frac{MJ}{m^2 \cdot dia} \right] \quad (15)$$

- ii. La función identifica las horas en que la radiación supera los 10 W/m², lo que indica condiciones de día. Con ese filtro toma los valores de velocidad de viento registrados en esas horas y calcula su promedio. De esta manera obtiene la velocidad del viento diurno, que es el que más influye en la evaporación.

```
# Filtra solo las horas donde hay sol (radiación > 10)
viento_dia = g['vel_S1'][g['ghi_S1'] > 10].mean()
```

- iii. Calcula el déficit de presión de vapor (VPD) a partir de las condiciones de humedad (RH) y temperatura (T), obteniendo así un indicador del potencial de evaporación en ese momento.

$$VPD = e_s \left(1 - \frac{RH}{100} \right) \quad (16)$$

donde

$$e_s = 0.6108 \cdot \exp \left(\frac{17.27 \cdot T}{T + 237.3} \right) \quad (17)$$

8.2. Agrupamiento en Clústeres: Enrutador Climático

Debido a la fuerte variabilidad climática a lo largo de Chile, un único modelo global no logra captar adecuadamente las diferencias locales, altamente sitio específicas. Para resolver esto, se implementó una estrategia de regionalización automática basada en *Clustering No Supervisado* [11, 12].

Algoritmo K-Means: K-Means agrupa las estaciones según atributos que no cambian día a día, como la latitud, la altitud (en metros sobre el nivel del mar, msnm) y los promedios anuales de temperatura y humedad. Esto permite formar clústeres que representan bien las condiciones climáticas de cada zona [13, 14].

El algoritmo divide el territorio en cuatro zonas o “clústeres expertos” (por ejemplo, Desierto Costero, Valle Central, Transición Sur, Sur Húmedo). El modelo `KMeans` entrenado se almacena como `router_climatico.pkl`, lo que permite clasificar futuras estaciones según su similitud con los centroides definidos.

8.3. Entrenamiento por Clúster

Para cada clúster identificado se entrena un modelo de regresión independiente. Todos los modelos aprenden el residuo entre la evaporación observada ($E_{pan(observada)}$) y la estimada por Penman-Monteith ($E_{PM(teórica)}$).

$$\delta = E_{pan(observada)} - E_{PM(teórica)} \quad (18)$$

Esta técnica asegura que la red neuronal se enfoque exclusivamente en corregir aquello que el modelo físico no logra capturar, reduciendo la varianza de la predicción y mejorando la capacidad de generalización.

```
# Definición del Objetivo de Aprendizaje
# Si Penman es perfecto, el target es 0.
# Si Penman subestima, el target es positivo.
y = df_cluster['Evap_PAM'] - df_cluster['E_PM_S1']
```

La herramienta permite elegir entre tres algoritmos mediante el parámetro `MODELO_ELEGIDO`:

Modelo	Descripción
Neural Network	Perceptrón Multicapa (MLP). Por defecto.
LightGBM	Gradient Boosting. Rápido y robusto.
Random Forest	Ensamble de árboles. Interpretable.

Optimización de Hiperparámetros:

Si `OPTIMIZAR = 's'`, el sistema busca la mejor configuración mediante `RandomizedSearchCV`, evaluando 10 combinaciones aleatorias:

Modelo	Hiperparámetros optimizados
Neural Network	hidden_layer_sizes: (100,), (100,50), (100,50,25) / alpha: 0.0001, 0.001, 0.01 / learning_rate_init: 0.001, 0.01
LightGBM	n_estimators, max_depth, learning_rate
Random Forest	n_estimators, max_depth, min_samples_split

Configuración por defecto (Neural Network):

Parámetro	Valor	Descripción
hidden_layer_sizes	(100, 50)	Dos capas ocultas para capturar relaciones no lineales
activation	relu	Evita problemas de desvanecimiento de gradiente
solver	adam	Optimizador eficiente
max_iter	5000	Iteraciones suficientes para convergencia

El código itera automáticamente por cada clúster identificado y ejecuta los siguientes pasos:

- Filtrado de Datos:** Selecciona solo las estaciones que pertenecen al clima actual.
- Escalado (Normalización):** Las redes neuronales son sensibles a la magnitud de los números. Se usa `StandardScaler` para poner todas las variables (temperatura, radiación, viento) en una misma escala comparable.
- Optimización (opcional):** Busca mejores hiperparámetros si `OPTIMIZAR = 's'`.
- Entrenamiento:** Ajusta el modelo con los datos escalados.
- Guardado:** Exporta modelo y escaladores a archivos `.pkl`.

Código de Entrenamiento (Nota: El código mostrado es una versión simplificada. La implementación completa incluye optimización automática de hiperparámetros cuando `OPTIMIZAR = 's'`):

```
def optimizar(X_train, y_train):
    if OPTIMIZAR != "s": return {}
    if MODELO_ELEGIDO == "Neural Network":
        param_dist = {"hidden_layer_sizes":[(100,50),(100,50,25),(100,,)], "alpha":uniform(0.0001,0.01)}
    elif MODELO_ELEGIDO == "LightGBM":
        param_dist = {"n_estimators":randint(100,500),"learning_rate":uniform(0.01,0.2),"num_leaves":randint(20,50)}
    else:
        param_dist = {"n_estimators":randint(100,500),"max_depth":randint(10,25),"min_samples_leaf":randint(1,5)}
    sc = StandardScaler()
    search = RandomizedSearchCV(crear_modelo(), param_dist, n_iter=10, cv=3, n_jobs=-1, random_state=42, scoring="r2")
    search.fit(sc.fit_transform(X_train), y_train)
    return search.best_params_

for cluster_id in cluster_ids:
    print(f"\n→ Clúster {cluster_id}...")
    df_cl = master_df_limpio[master_df_limpio["Cluster"] == cluster_id]
    estaciones = sorted(df_cl["localidad"].unique())

    folds = [(estaciones[0], df_cl.sample(frac=0.8,random_state=42), df_cl.sample(frac=0.2,random_state=42))]
    if len(estaciones)==1:
        else [(e, df_cl[df_cl["localidad"]!=e], df_cl[df_cl["localidad"]==e]) for e in estaciones])

    for est, df_train, df_test in folds:
        X_tr, y_tr = df_train[features_finales], df_train[target]
        X_te, y_te = df_test[features_finales], df_test[target]
        hp = optimizar(X_tr, y_tr)
        sc_X, sc_y = StandardScaler(), StandardScaler()
        X_tr_sc = sc_X.fit_transform(X_tr)
        y_tr_sc = sc_y.fit_transform(y_tr.values.reshape(-1,1)).ravel()
        m = crear_modelo(hp); m.fit(X_tr_sc, y_tr_sc)
        preds = sc_y.inverse_transform(m.predict(sc_X.transform(X_te)).reshape(-1,1)).ravel()
        df_pred = pd.DataFrame({"y_real":y_te,"y_pred":preds,"localidad":est,
                               "dia_del_anio":df_test["dia_del_anio"],"e_pm_total":df_test["e_pm_total"]})
        all_metrics_list.append(df_pred); all_plotting_list.append(df_pred)
        print(f"    {est}: R²={r2_score(y_te,preds):.3f}, KGE={calcular_kge(y_te,preds):.3f}")
```

Material de apoyo para validación del entrenamiento

El notebook incluye visualizaciones adicionales que permiten verificar la calidad del agrupamiento y detectar oportunidades de mejora. Estas incluyen: radar plots por clúster, análisis PCA, mapa geográfico y boxplots comparativos. También están disponibles en el Anexo 2.

8.4. Construcción del Enrutador Climático

Esta etapa construye el Enrutador Climático, el componente que clasifica automáticamente cualquier nuevo conjunto de datos meteorológicos dentro de los clústeres definidos durante el entrenamiento. Su función es indicar qué modelo debe utilizarse al trabajar con datos nuevos. Para lograrlo, se implementa un clasificador KNN ($k = 1$), que compara el perfil climático del nuevo punto con los perfiles ya conocidos y lo asigna al clúster más parecido. Con esta asignación, la herramienta selecciona directamente el modelo entrenado del clúster correspondiente.

Esta etapa no puede ejecutarse de manera independiente, requiere que el entrenamiento haya finalizado y que los siguientes objetos estén disponibles en memoria:

Variable	Función
<code>master_df_limpio</code>	Base diaria limpia y suavizada
<code>features_finales_modelo</code>	Lista final de variables utilizadas en el modelo
<code>mapa_localidad_a_cluster</code>	Diccionario que asigna cada estación al clúster correspondiente
<code>MODEL_FOLDER</code>	Carpeta donde se almacenan los modelos del clúster (<i>.pkl</i>).

El proceso comienza seleccionando un conjunto de variables meteorológicas básicas (`features_router_candidatas`), como radiación, VPD y velocidad del viento. Con estas variables se construye un perfil climático para cada estación, entendido como una representación numérica que resume sus características meteorológicas según los cálculos realizados en esta etapa. Posteriormente, estos perfiles se comparan entre sí para identificar cuáles estaciones presentan condiciones climáticas similares y, con esa información, construir el enrutador.

Esta etapa genera tres archivos:

<code>enrutador_knn.pkl</code>	Contiene el modelo matemático que determina a qué clúster pertenece un clima nuevo.
<code>enrutador_scaler.pkl</code>	Contiene la transformación necesaria para normalizar los datos antes de pasarlos al enrutador
<code>enrutador_centroides.csv</code>	Contiene el clima promedio de cada estación, lo que permite comparar las características de cada clúster, entender cómo se define cada clima y verificar que el enrutador esté asignando correctamente cada estación.

Referencias

- [1] L. C. DOBRE, A. Țurcanu, and A. Crăciunescu, "Floating Photovoltaic Power Plants," in *2021 12th International Symposium on Advanced Topics in Electrical Engineering (ATEE)*, 2021: IEEE, pp. 1–4.
- [2] L. Liu, Q. Wang, H. Lin, H. Li, and Q. Sun, "Power generation efficiency and prospects of floating photovoltaic systems," *Energy Procedia*, vol. 105, pp. 1136–1142, 2017.
- [3] E. Muñoz-Cerón, J. C. Osorio-Aravena, F. J. Rodríguez-Segura, M. Frolova, and A. Ruano-Quesada, "Floating photovoltaics systems on water irrigation ponds: Technical potential and multi-benefits analysis," *Energy*, vol. 271, p. 127039, 2023.
- [4] R. S. Spencer, J. Macknick, A. Aznar, A. Warren, and M. O. Reese, "Floating photovoltaic systems: assessing the technical potential of photovoltaic systems on man-made water bodies in the continental United States," *Environmental science & technology*, vol. 53, no. 3, pp. 1680–1689, 2018.
- [5] H. L. Penman, "Natural evaporation from open water, bare soil and grass," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 193, no. 1032, pp. 120–145, 1948.
- [6] J. L. Monteith, "Evaporation and environment," in *Symposia of the society for experimental biology*, 1965, vol. 19: Cambridge University Press (CUP) Cambridge, pp. 205–234.
- [7] L. Turc, "Evaluation des besoins en eau d'irrigation, évapotranspiration potentielle," *Ann. Agron.*, vol. 12, pp. 13–49, 1961.
- [8] G. H. Hargreaves and Z. A. Samani, "Reference crop evapotranspiration from temperature," *Applied engineering in agriculture*, vol. 1, no. 2, pp. 96–99, 1985.
- [9] R. C. Gonzalez *et al.*, "Exploring climate-driven performance of floating photovoltaic systems: Energy production enhancement and evaporation reduction," *Applied Energy*, vol. 386, p. 125556, 2025.
- [10] F. Bontempo Scavo, G. M. Tina, A. Gagliano, and S. Nižetić, "An assessment study of evaporation rate models on a water basin with floating photovoltaic plants," *International journal of energy research*, vol. 45, no. 1, pp. 167–188, 2021.
- [11] Á. Arroyo, Á. Herrero, V. Tricio, and E. Corchado, "Analysis of meteorological conditions in Spain by means of clustering techniques," *Journal of Applied Logic*, vol. 24, pp. 76–89, 2017.
- [12] S. Borzooei, G. H. Miranda, S. Abolfathi, G. Scibilia, L. Meucci, and M. C. Zanetti, "Application of unsupervised learning and process simulation for energy optimization of a WWTP under various weather conditions," *Water Science and Technology*, vol. 81, no. 8, pp. 1541–1551, 2020.
- [13] M. Eghtesadifard, P. Afkhami, and A. Bazyar, "An integrated approach to the selection of municipal solid waste landfills through GIS, K-Means and multi-criteria decision analysis," *Environmental research*, vol. 185, p. 109348, 2020.
- [14] R. A. Caceres Gonzalez and M. C. Hatzell, "Prioritizing the best potential regions for brine concentration systems in the usa using gis and multicriteria decision analysis," *Environmental Science & Technology*, vol. 57, no. 46, pp. 17863–17875, 2022.