



# Evaluación de los productos de evapotranspiración disponibles en Climate Engine y del algoritmo Support Vector Machine Regression con datos NASA Power

## Evaluation of evapotranspiration products available from Climate Engine and the Support Vector Machine Regression model with NASA Power data.

 María Florencia Degano<sup>1,2\*</sup> 

 Raúl Eduardo Rivas<sup>2,3</sup> 

\* Contacto

Recibido: 27 de febrero de 2023 • Aceptado: 24 de abril de 2023 • Publicado: 28 de julio de 2023

### Resumen

El manejo hidrológico de un país depende, en gran medida, del conocimiento de las cuencas existentes, del potencial de estas y de la manera de gestionar adecuadamente los excedentes hídricos. En este sentido es de vital importancia el estudio y análisis de la evapotranspiración (ET) de referencia ( $ET_0$ ), real y potencial (ETp). Por lo que se torna indispensable evaluar el comportamiento de los diferentes productos de ET que se encuentran disponibles para su uso de manera libre. En este sentido, el objetivo principal de este trabajo es analizar los datos de los modelos existentes en la plataforma *Climate Engine* (*TerraClimate*, ERA 5, MERRA-2 y MOD16A2), que posee datos a diferentes escalas temporales y espaciales. Además, evaluar el algoritmo *Support Vector Machine Regression* (SVR) de inteligencia artificial, aplicado con parámetros obtenidos de NASA Power y con datos locales registrados en la Región Pampeana Argentina (RPA). En general, el error cuadrático medio (RMSE) varía entre 0,5 y 1,2 mm d<sup>-1</sup> y los valores del índice de eficiencia de Nash-Sutcliffe (NSE) entre 0,6 y 0,9 (para  $ET_0$ ); entre 0,4 y 0,7 para ET real y entre 0,6 y 0,9 para ETp. Asimismo, queda demostrado que el modelo más propicio para el cálculo de  $ET_0$  como ET real, es el SVR, mientras que para ETp es ERA 5.

**Palabras clave:** inteligencia artificial, productos de reanálisis, productos de satélite.

1. Consejo Nacional de Investigaciones Científicas y Técnicas

2. Instituto de Hidrología de Llanuras, Tandil, Buenos Aires, Argentina

3. Comisión de Investigaciones Científicas

## Abstract

The hydrological country management depends on the knowledge of the existing basins, their potential and to adequately manage water surpluses. In this sense, the study and analysis of reference ( $ET_0$ ), actual and potential ( $ET_p$ ) evapotranspiration (ET) is of vital importance. Therefore, it is essential to evaluate the behavior of the different ET products that are freely available for use. Therefore, the main work objective is to analyze the data of the existing models in the *Climate Engine* platform (*TerraClimate*, ERA 5, MERRA-2 and MOD16A2), which has data at different temporal and spatial scales. In addition, to evaluate the artificial intelligence Support Vector Machine Regression (SVR) algorithm, determined with parameters obtained from NASA Power, with local data in the Argentinean Pampean Region (APR). In general, the Root Mean Square Error (RMSE) varies between 0.5 and 1.2 mm d<sup>-1</sup> and Nash-Sutcliffe Efficiency Index (NSE) values between 0.6 and 0.9 ( $ET_0$ ); between 0.4 and 0.7 for actual ET and between 0.6 and 0.9 for  $ET_p$  were obtained. Likewise, it is demonstrated that the most favorable model for the calculation of  $ET_0$  as real ET is SVR, while for  $ET_p$  it is ERA 5.

**Keywords:** artificial intelligence, reanalysis products, satellite products.

## Introducción

El manejo hidrológico de un país depende, en gran medida, del conocimiento de las cuencas existentes, del potencial de éstas y de la manera de gestionar adecuadamente los excedentes hídricos. A partir su estudio es posible desarrollar un manejo adecuado del agua disponible, siendo necesario registrar la información no solo de las estaciones meteorológicas, sino también, validar los datos obtenidos de sensores remotos con datos *in situ* a modo de lograr medidas consistentes a fin de monitorear y/o evaluar las variables hidrológicas.

Siguiendo este eje, es posible analizar adecuadamente la Evapotranspiración (ET) para gestionar y planificar el manejo hídrico de las cuencas de llanura (Faramiñán, *et al.*, 2021a), en particular, en la Región Pampeana Argentina (RPA), dado que, alrededor del 85% del agua que precipita es devuelta a la atmósfera mediante este proceso (Weinzettel y Usunoff, 2001).

La ET es considerada un término de relevancia para el análisis de disponibilidad hídrica en diferentes tipos de suelo y clima de la RPA, en este sentido, el uso de las tecnologías y la información proveniente de satélite, reanálisis y algoritmos de inteligencia artificial es de vital importancia para el análisis hídrico de la región, en particular, en las zonas donde son escasos los datos de estaciones locales o de campo.

Actualmente, existen diversas plataformas que proveen datos de ET, calculada con información proveniente de distintos satélites y fuentes de reanálisis, como es el caso de *Climate Engine* (<http://ClimateEngine.org>). *Climate Engine* es una aplicación web de acceso libre, que opera en la nube, empleando *Google Earth Engine*, para procesar, visualizar, descargar y compartir conjuntos de datos climáticos y de teledetección en tiempo real. Posee una recopilación de productos climáticos, meteorológicos y de teledetección de diferentes variables hidrológicas, poniéndolos al servicio de la comunidad para estudios aplicados a la agricultura, finanzas y economía, cadena de producción y consumo, servicios profesionales, infraestructura y transporte, monitoreo agrí-

cola, estrés de cultivos y sequías, calor y viento extremo, peligro de inundación, capa de nieve, potencia solar, pronósticos subestacionales a estacionales, incendios, superficie del agua, en tiempo (casi) real (Huntington *et al.*, 2017).

La capacidad de esta aplicación de acceder a archivos de datos bajo demanda genera grandes oportunidades para la supervisión avanzada de los recursos naturales y la comprensión de los procesos; tal es el caso de la ET real, ET de referencia ( $ET_0$ -pasto corto); ETr-alfalfa (que no se tuvo en cuenta en este análisis) así como de la ET potencial ( $ET_p$ ) para distintas zonas del planeta, obtenida con diferentes algoritmos para su cálculo (Huntington *et al.*, 2017).

Por otro lado, el uso de las nuevas tecnologías en el estudio, determinación y aplicación de variables hidrológicas, han tomado gran importancia en el último tiempo, como es el caso de los algoritmos de inteligencia artificial, en particular *Support Vector Machine Regression* (SVR), que ha demostrado ser adecuado para el uso y análisis de la ET en zonas donde no existen o hay escasez de datos de terreno, tomando como entrada información de la base de datos de NASA Power (Faramiñán *et al.*, 2021b).

El algoritmo SVR permite obtener resultados aplicados a grandes conjuntos de datos complejos con ruido, es decir con variables inherentes al modelo que para otros algoritmos aumentan la posibilidad de error en los resultados (Chia *et al.*, 2020a; Cristianini y Shawe-Taylor, 2000; Liakos, *et al.*, 2018; Shrestha y Shukla, 2015). Debido a que la ET depende de un gran número de variables hidrometeorológicas, se lo define como un proceso complejo y no lineal, por lo que, en los últimos años se ha incrementado el uso de este algoritmo para estimarla (Chia *et al.*, 2020b; Fan *et al.*, 2018; Juárez *et al.*, 2018)

En este contexto, la ET puede analizarse a través de diferentes estimaciones de productos (obtenidos con medidas indirectas) que responden a un tipo de dato determinado. Por lo que, el objetivo principal de este trabajo es analizar los productos existentes de  $ET_0$ , ET real y  $ET_p$  en la plataforma *Climate Engine* a diferentes resoluciones espaciotempo-

rales y evaluar el algoritmo SVR (con parámetros obtenidos de NASA Power). Los resultados son comparados con datos de campo provenientes de la Oficina de Riesgo Agropecuario (ORA) con la finalidad de conocer la mejor herramienta de monitoreo y gestión de las cuencas hidrológicas de la RPA.

### Zona de estudio y metodología

#### - Estaciones hidrometeorológicas en la región Pampeana Argentina.

Para analizar el proceso de ET mediante medidas indirectas, es necesario contar con datos de campo. En este análisis, se tienen en cuenta siete estaciones distribuidas en la RPA por presentar características edafoclimáticas representativas del proceso analizado para un área aproximada

de 520.000 km<sup>2</sup> (Degano *et al.*, 2021b) (Figura 1). La RPA se encuentra en el centro-este de la República Argentina y constituye el más importante ecosistema de praderas del país. Según Pereyra (2003), la RPA se subdivide en 12 ecorregiones: Planicies poligenéticas de Entre Ríos (D25), Pampa pedemontana cordobesa (D26), Pampa arenosa (D27), Pampa endorreica (D28), Delta del Paraná (D29), Pampa Ondulada (D30), Pampa Deprimida (D31), Planicies litorales pampeanas (D32), Sierras Septentrionales Bonaerenses (D33), Pampa Interserrana (D34), Sierras Australes Bonaerenses (D35) y Planicies poligenéticas de La Pampa (D36).

El clima de la región es subhúmedo/húmedo (centro-este) a semiárido (hacia el oeste). Las precipitaciones anuales varían entre 1200 mm en el Noreste y 600 mm en el Suroeste. En lo que respecta a la ETp, se encuentra en un rango entre 900 y 1100 mm anuales. Por su parte, la tempe-

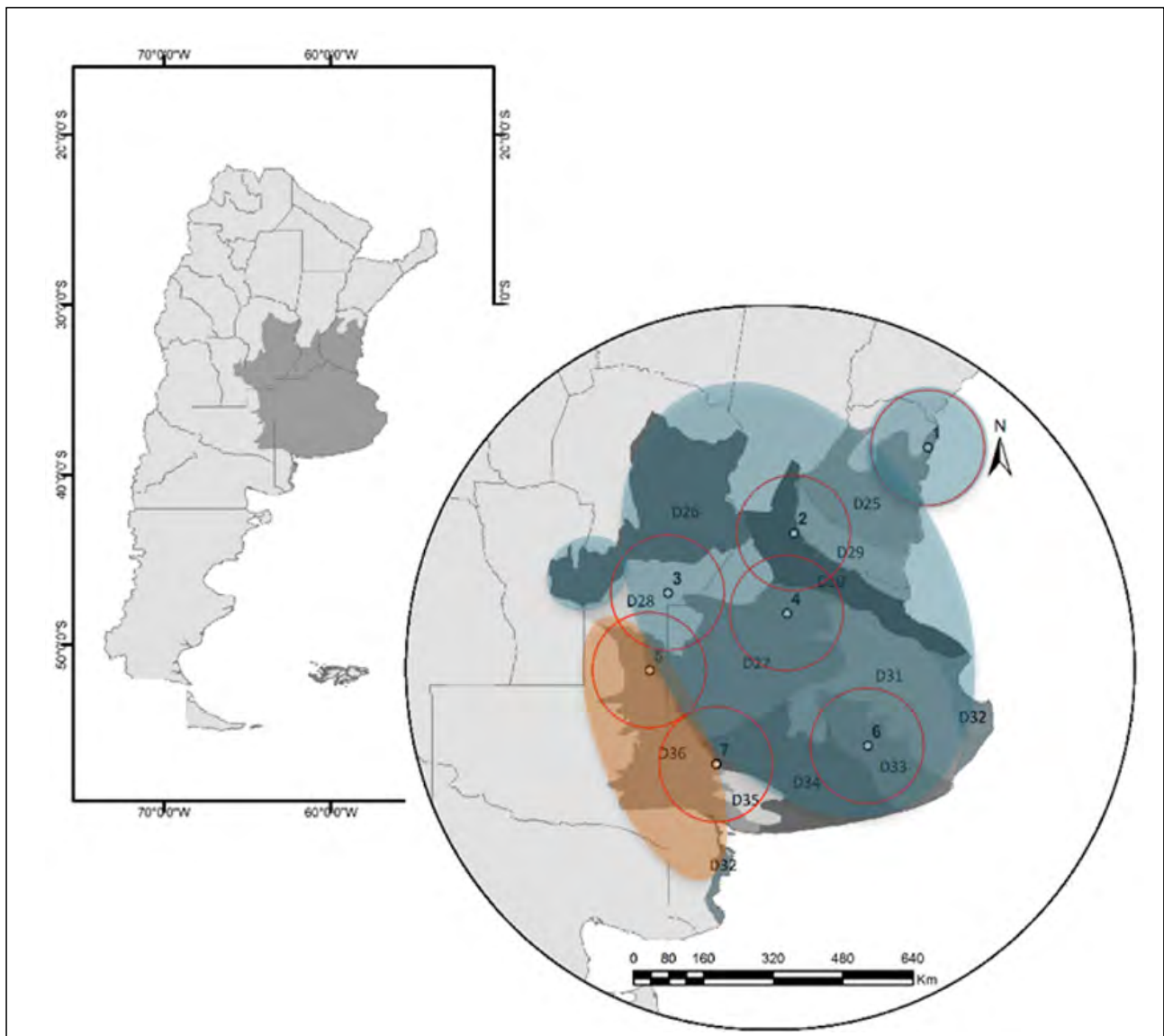


Figura 1: Ubicación de la RPA en la República Argentina; Ecorregiones (D25-D36); Distribución espacial del área de influencia de las estaciones meteorológicas (círculo rojo); Zona sombreada naranja: Región semiárida; Zona sombreada celeste: Región subhúmeda/húmeda. Estaciones: 1. Concordia; 2. Rosario; 3. Laboulaye; 4. Junín; 5. General Pico; 6. Tandil; 7. Pigüé (Modificado de Degano *et al.*, 2021a)

ratura media anual máxima es de 16°C (en el Sur) y de 18°C (en el Norte); la mínima media anual se encuentra entre los 6°C y 14°C. Las temperaturas decrecen en sentido Norte-Sur, con una amplitud de 15°C en el Oeste y 12°C en el Este.

Los datos de campo utilizados fueron proporcionados por la Oficina de Riesgo Agropecuario (ORA), que obtienen los valores de ETO con datos de entrada diarios medidos in situ por el Servicio Meteorológico Nacional (SMN) de Argentina (precipitación, temperatura máxima y mínima, humedad relativa, velocidad de viento, heliofanía o nubosidad) y se basan en el método de la *Food and Agriculture Organization of the United Nations* (FAO)- Penman-Monteith (PM) (Allen *et al.*, 1998, Monteith y Unsworth, 1990). De esta se deriva la ETp, que se toma como sinónimo de ET<sub>0</sub>, debido a que el análisis se realiza en estaciones que están ubicadas en sectores con pasto corto y características de referencia.

Respecto de los datos de ET real, se obtienen despejando este término de la ecuación de balance hidrológico, de resolución temporal diaria, donde se toman los ingresos y egresos de agua del sistema (Allen, *et al.*, 1998; Basualdo, 2011):

**- Productos de reanálisis y satélite (Climate Engine)**

En este trabajo se analizaron los productos TerraClimate (Abatzoglou *et al.*, 2018), ERA 5 (Copernicus Climate Change Service, 2017), MERRA 2 (Gelaro *et al.*, 2017) y MOD16A2 (Mu *et al.*, 2013; Running *et al.*, 2017) (Tabla 1). Los productos mencionados ponen a disposición estimaciones de distintos tipos de ET: ET<sub>0</sub>, ET real y ETp calculada con el método de Hargreaves (ETp<sub>Hargreaves</sub>). Las estimaciones de la ETO derivan del modelo Penman-Monteith (Allen *et al.*, 1998), asumiendo una superficie de pasto corto (12 cm). La determinación de la ET real es en base a la ETO. Por último, la ETp<sub>Hargreaves</sub> (Hargreaves y Samani, 1985), que requiere datos de temperatura y de radiación solar incidente (Rs) y se evalúa a partir de la radiación solar extraterrestre.

El producto *TerraClimate* es un conjunto de datos de reanálisis, analizando el clima a una resolución temporal mensual, mediante la ecuación de balance hídrico a nivel global. Combina datos climatológicos de alta resolución espacial del conjunto WorldClim (<https://www.worldclim.org/>), con una resolución espacial menor, con datos de CRU

Ts4.0 (*Climatic Research Unit - https://data.ceda.ac.uk/badc/cru/data/cru\_ts/*) variables en el tiempo; e información de reanálisis JRA55 (*Japanese 55-year Reanalysis - https://jra.kishou.go.jp/JRA-55/index\_en.html*). La ET se calcula a partir de un balance hídrico unidimensional simplificado basado en un enfoque Thornthwaite-Mather modificado, que utiliza precipitaciones, ETp (Penman-Monteith), la humedad del suelo y el almacenamiento de agua del manto de nieve.

En general, además de los productos de ET, se encuentran datos mensuales de temperatura, precipitación, velocidad del viento, presión de vapor, radiación solar, humedad de suelo, escorrentía, índices de sequía y nieve. Estos datos proporcionan información importante para los estudios ecológicos e hidrológicos a escala mundial que requieren alta resolución espacial y datos sobre el clima y el balance hídrico variables en el tiempo (Abatzoglou *et al.*, 2018).

Por su parte, ERA 5 proviene del *European Center for Medium-Range Weather Forecasts* (ECMWF), combina datos del modelo con observaciones de 24 km de resolución espacial en todo el mundo, en un conjunto de datos completo y consistente. ERA5 reemplaza a su predecesor, denominado ERA-Interim (Castañeda Calatayud, 2020). Este producto, proporciona valores agregados para cada día para distintos parámetros de reanálisis climático: temperatura, precipitación, ET, viento, condiciones atmosféricas e índices de sequía. Los datos se encuentran disponibles desde 1979 a la actualidad.

Por otro lado, el producto de reanálisis, MERRA 2 (*Modern-Era Retrospective analysis for Research and Applications version 2*), reemplaza el conjunto de datos MERRA original. Es producido por la *Global Modeling and Assimilation Office* (GMAO) de la NASA. En particular, dentro de la aplicación *Climate Engine*, presenta productos relacionados a temperatura máxima y mínima, precipitaciones, ET, índices de sequía y fuego, con una resolución espacial de 50 km y temporal diaria.

En relación con el producto de satélite, MOD16A2 es producido con una resolución temporal de 8 días y con píxeles de 500 m. Se basa en la lógica de la ecuación de Penman-Monteith, que incluye datos de entrada de reanálisis meteorológico diario junto con productos de datos de detec-

**Tabla 1: Características de los productos de ET utilizados. \*Basado en el trabajo Degano *et al.* (2021b). \*\*El producto MOD16A2 presenta datos de ETp que, en este caso, se toma como sinónimo de ET<sub>0</sub>.**

Producto	Resolución espacial [km]	Resolución temporal	Período analizado	Dato	Base de datos
TerraClimate	4	Mensual	2009 - 2013	ET <sub>0</sub> ET real ETp <sub>Hargreaves</sub>	Reanálisis
ERA 5	24	Diario	2009 - 2013	ETp <sub>Hargreaves</sub>	Reanálisis
MERRA-2	50	Diario	2009 - 2013	ET <sub>0</sub> ETp <sub>Hargreaves</sub>	Reanálisis
MOD16A2*	0,5	8 días	2012 - 2014	ETp** ET real	Satélite

ción remota MODIS (propiedades de la vegetación, albedo, cobertura terrestre, entre otras). Los valores de píxel para las dos capas de evapotranspiración (ET real y ETp) y para el flujo de calor latente ( $LE$  y  $LE_p$ ) son la suma y el promedio de ocho días (Mu *et al.*, 2013; Running *et al.*, 2017).

#### - Obtención de los datos de la aplicación Climate Engine

A partir de los datos disponibles en la plataforma *Climate Engine* (<https://app.climateengine.com/climateEngine>), se obtienen los valores de las diferentes variables, según se encuentren disponibles en cada producto, y, se analizan, siguiendo los aportes realizados por Degano *et al.* (2021a) y Faramiñán *et al.* (2021b), los cuales brindan información necesaria para la valoración de modelos satelitales aplicados a zonas de llanura.

En este punto se determina de manera rápida y sencilla el proceso para la obtención de los datos en diferentes puntos del territorio utilizando *Climate Engine*. En la solapa "Make Graph" se determina la región de la cual se van a obtener los valores de ET, seleccionando un máximo de 5 puntos para representar geográficamente en el mapa a través de las coordenadas (latitud-longitud).

En el apartado "Variable" se determina el producto que se requiere y la variable, así como el período de tiempo del que se requieren los datos. Se genera la serie de datos ("get time series") y posteriormente se descarga la tabla con la información requerida, lo que facilita la manipulación y procesamiento de los datos.

#### - Algoritmo Support Vector Machine Regression y datos de entrada NASA Power

SVR es un modelo de regresión lineal mediante el cual, se puede definir el valor del error que se acepta en el modelo y encontrar una línea apropiada (hiperplano) para hacer el ajuste de los datos (Chia *et al.*, 2020b). Su función es minimizar los coeficientes y determinar un error absoluto menor o igual que un margen específico (error máximo), que se puede ajustar para obtener la precisión deseada.

Puede suceder que no todas las observaciones se encuentren dentro del margen de error, debido a la no linealidad de las variables explicativas (en particular, en el proceso de ET), por lo que, se debe ajustar o corregir para cualquier valor que esté fuera del error máximo y agregar las desviaciones a la función objetivo (Faramiñán, 2022).

Para el desarrollo de este algoritmo, como entrada, se tomaron variables provenientes de la base de datos NASA Power (POWER Release-8 <https://power.larc.nasa.gov/>). Estas se basan en el modelo de asimilación de *Goddard Global Modeling and Assimilation Office* (CMMS) y provienen de MERRA2, de SRB (*Surface Radiation Budget*) y de FLASH Flux. Poseen una resolución espacial de 50 km y temporal diaria.

El algoritmo SVR, se llevó a cabo según la metodología

descrita en Faramiñán *et al.* (2021a). Toma como entrada las variables principales que determinan el proceso de ET (Temperatura, Humedad Relativa, Presión Superficial, Radiación de onda corta y larga entrante y Velocidad de Viento), entre el año 1984 y 2021 (denominadas variables explicativas) para relacionarlas con datos de ET real y ETp medidas *in situ* (de la ORA) en la RPA (variables respuesta). Para desarrollar el procedimiento, se tuvieron en cuenta las librerías de Python: Pandas, scikit-learn y NumPy y, en este caso, se utilizó el 80% de los datos para el entrenamiento del algoritmo y el 20% para hacer el testeado de los datos.

#### - Análisis estadístico

Para evaluar los conjuntos de datos tanto de los productos de ET de reanálisis, como de satélite e inteligencia artificial, se utilizaron métricas estadísticas relacionadas al tipo de estudio. Por lo que, se evaluaron la raíz del error cuadrático medio (RMSE), el error absoluto medio (MAE), el índice de eficiencia de Nash-Sutcliffe (NSE), el coeficiente de determinación ( $R^2$ ), la pendiente (a) y la ordenada del origen (b), la Mediana (Me) y la desviación estándar robusta (RSD)

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (E_i - O_i)^2}{N}} \quad (1)$$

$$MAE = \frac{|\sum_{i=1}^N (E_i - O_i)|}{N} \quad (2)$$

$$NSE = 1 - \frac{\sum_{i=1}^N (O_i - E_i)^2}{\sum_{i=1}^N (O_i - \bar{O})^2} \quad (3)$$

$$R^2 = \left( \frac{\sum_{i=1}^N (O_i - \bar{O})(E_i - \bar{E})}{(\sum_{i=1}^N (O_i - \bar{O})^2 \sum_{i=1}^N (E_i - \bar{E})^2)^{0,5}} \right)^2 \quad (4)$$

$$RSD = 1,483 Me (|(E_i - O_i) - Me(E_i - O_i)|) \quad (5)$$

donde  $E_i$  son los valores estimados,  $O_i$ , los valores observados y  $N$  es el número de observaciones. Respecto del NSE, se trata de un indicador que se utiliza a menudo en hidrología, debido a que es muy flexible para ser aplicado a distintos tipos de modelos matemáticos, aunque se deben tener en cuenta sus limitaciones, como lo es la sensibilidad al sesgo en las predicciones del modelo y el posible efecto de los valores atípicos de la serie de datos. Este índice determina que el modelo es "insatisfactorio", cuando el valor de NSE es menor a 0,65; "aceptable", cuando el NSE es  $\geq 0,65$ ; "bueno" cuando se trata de un valor  $\geq 0,8$  y "muy bueno" cuando el valor es  $\geq 0,9$  (McCuen *et al.*, 2006; Ritter y Muñoz, 2013).

## Resultados y discusión

### - Evapotranspiración de referencia - $ET_0$

Esta variable se encuentra disponible en el producto de *TerraClimate* (los datos de este producto se pasaron a valores diarios  $ET/n$ , siendo  $n$  la cantidad de días del mes) MERRA-2, MOD16A2 (los valores de  $ET$  se llevaron a una resolución diaria, aplicando  $ET/8d$ ) y se pretende modelar con el algoritmo SVR. En la Tabla 2 se muestran los estadísticos obtenidos para los diferentes productos de  $ET_0$ . Cabe aclarar que este análisis se realizó sobre todas las estaciones determinadas en la Figura 1, ubicadas en la RPA, por lo que, en este apartado se estudia la región en general.

Según el análisis de las métricas estadísticas, que muestran para los productos de reanálisis un error menor en lo que respecta al producto *TerraClimate* en comparación con MERRA-2, en ambos casos la distribución de los datos es consistente, determinada por el valor de  $R^2$ , cuyo valor es cercano a 1 en ambos casos. Por su parte, la Mediana, que determina el error sistemático, arroja valores positivos para ambos productos, lo que indica una sobreestimación. El NSE indica que ambos modelos son “muy buenos” para la determinación de la  $ET_0$ .

Respecto del producto de satélite, MOD16A2, se tomaron los valores obtenidos por Degano *et al.* (2021b). Los errores obtenidos son mayores que para el resto de los productos en general. Asimismo, la distribución de los datos es muy buena, determinada por el  $R^2$ . En general, el producto sobreestima sistemáticamente, determinado por la Mediana positiva. Respecto del NSE, el producto es “aceptable” y con utilidad para conocer la variabilidad espacial y temporal. En este caso, Degano *et al.* (2021ayb) desarrollaron un ajuste lineal para corregir este producto. Tras la corrección, los errores disminuyen significativamente, y los resultados del producto mejoran considerablemente. El RMSE y el MAE se reducen en torno al 85%, y el NSE asciende a 0,9, otorgando al modelo una calificación de “muy bueno” para obtener los valores de la  $ET_0$ .

Por último, el producto obtenido mediante el algoritmo de inteligencia artificial muestra errores menores que en el resto de los productos, con valores de NSE que indican que el modelo es “bueno” para su aplicación en el campo de la  $ET_0$ .

En particular, se analizaron dos estaciones de la RPA para los productos *TerraClimate* y MERRA-2 (Figura 2), la estación General Pico, que se encuentra en la zona semiárida, y

la estación Tandil, ubicada en la región subhúmeda-húmeda. Para el producto *TerraClimate*, el RMSE para General Pico es de  $0,8 \text{ mm d}^{-1}$ , mientras que, para la región de Tandil, disminuye a  $0,51 \text{ mm d}^{-1}$ . En ambos casos, el producto muestra una distribución cercana a la línea 1:1 hacia los valores bajos de  $ET_0$ , sobreestimando a medida que aumenta su valor. Esto indica, que, tanto en zonas húmedas como semiáridas, el producto es aplicable, ya que, los errores asociados se encuentran dentro de lo esperable para productos de reanálisis (Allen *et al.*, 2011).

Por otro lado, se observan diferencias en las comparaciones del producto MERRA-2 con datos de campo, respecto de cada zona. Por un lado, en la zona semiárida, la distribución de la nube de puntos se encuentra más cercana a la línea 1:1 y el RMSE alcanza un valor de  $1 \text{ mm d}^{-1}$ , mientras que, en la zona subhúmeda, el comportamiento es similar al producto *TerraClimate*, mostrando una distribución cercana a la línea 1:1 en valores bajos de  $ET_0$ , y alejándose al aumentar su valor. En esta zona, el error asciende a  $1,3 \text{ mm d}^{-1}$ .

### - Evapotranspiración real – $ET_{real}$

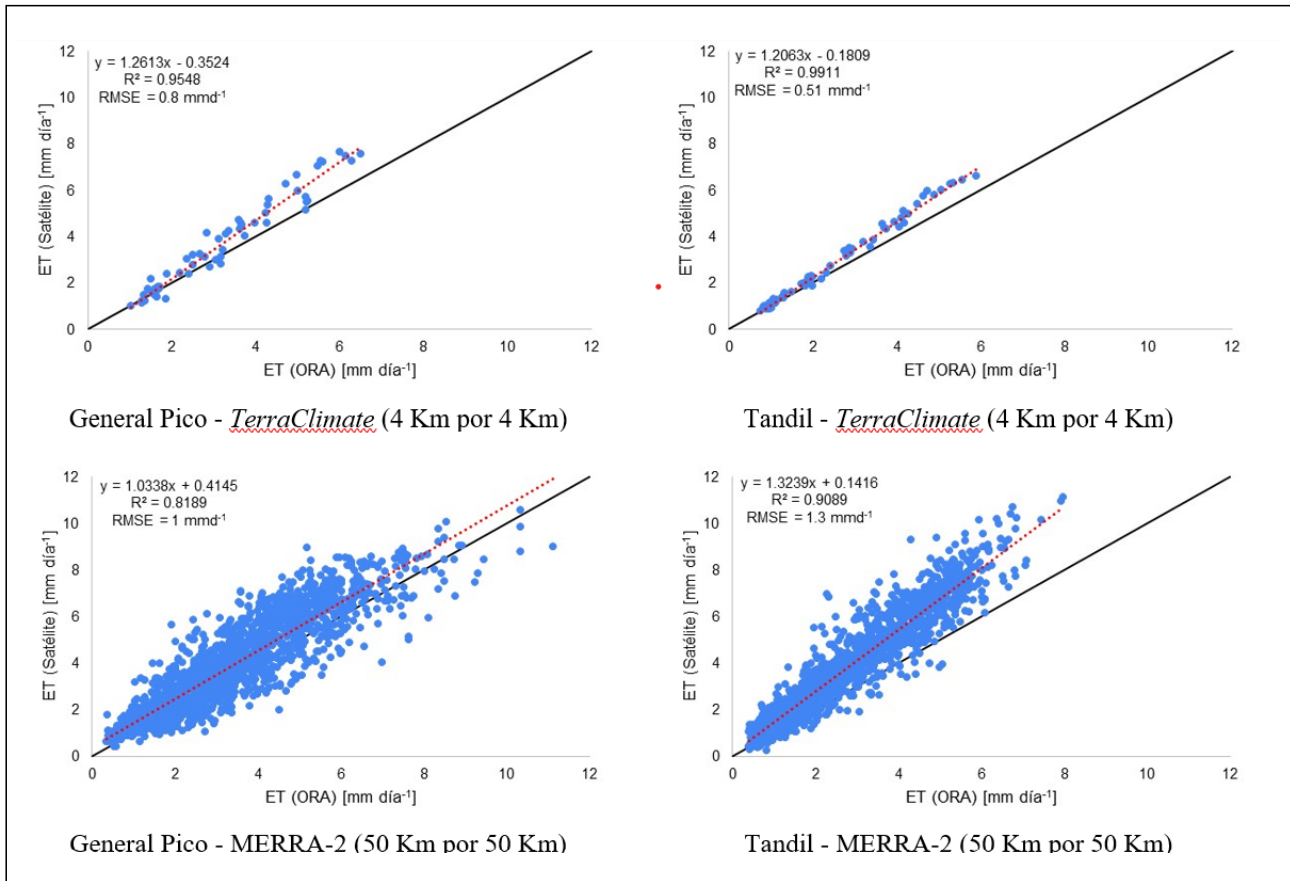
Este parámetro se encuentra disponible en el producto de *TerraClimate*, MOD16A2 y el calculado usando el algoritmo SVR. En general, según los estadísticos obtenidos para  $ET_{real}$  (Tabla 3), el modelo que mejor se ajusta es el obtenido con el algoritmo SVR, indicado por el valor mayor de NSE (que indica que el producto es “aceptable”), mientras que, *TerraClimate* indica valor negativo de NSE, mostrando una performance deficitaria para determinar la  $ET_{real}$ . Por su parte, MOD16A2 tendría la clasificación de “insuficiente”, debido a un NSE de 0,4, aunque el error obtenido se encuentra dentro de lo esperable para los productos de satélite (menor al 20% respecto del valor medio observado - Allen *et al.*, 2011; Degano *et al.*, 2021b). Asimismo, el error determinado es menor para el producto derivado de inteligencia artificial, indicando a éste como el más propicio para analizar la  $ET_{real}$ . Por el contrario, el algoritmo SVR ajustado es propio de la zona y no sería extrapolable a otras regiones del planeta.

### - Evapotranspiración potencial mediante Hargreaves – $ET_{pHargreaves}$

La evapotranspiración de Hargreaves se encuentra disponible en el producto de *TerraClimate*, ERA 5 y MERRA-2. Según los estadísticos obtenidos (Tabla 4), los modelos son “muy buenos” para analizar la  $ET_p$ , ya que el NSE otorga valores mayores a 0,9. Los resultados de los estadísticos indican

Tabla 2: Estadísticos asociados a los productos de  $ET_0$ .

Producto	Número de datos	RMSE [mm d <sup>-1</sup> ]	MAE [mm d <sup>-1</sup> ]	R <sup>2</sup>	Me [mm d <sup>-1</sup> ]	RSD [mm d <sup>-1</sup> ]	NSE
TerraClimate	399	0.5	0	0.97	0.4	0.2	0.89
MERRA-2	12329	1.1	0.1	0.81	0.4	0.8	0.88
MOD16A2	1722	3.2	3	0.85	0.7	0.9	0.57
SVR	97161	0.4	0.3	0.87	0.02	0.5	0.86



**Figura 2: Relación entre los productos TerraClimate y MERRA-2 en las estaciones General Pico (semiárida) y Tandil (subhúmeda). Línea negra: relación 1:1. Línea punteada roja: línea de tendencia**

que se encuentran dentro de los márgenes de error establecidos por Allen *et al.*, (2011) para productos derivados de información satelital. ERA5 y MERRA-2 presentan una sobreestimación observable por la mediana positiva.

Analizando los productos que se encuentran en la plataforma *Climate Engine* para la determinación del parámetro ETp, se observa que, en General Pico, el producto *TerraClimate* subestima los valores de ETp, encontrándose la nube de puntos por debajo de la relación 1:1 y el valor de mediana negativo. Por su parte, el RMSE alcanza un valor de 1,1 mm d<sup>-1</sup> para el área en donde se ubica la estación de referencia. Respecto de la región subhúmeda/húmeda, el producto se acerca a la línea 1:1 cuando la ETp presenta valores bajos, distanciándose levemente al incrementar su valor, mostrando una sobreestimación del producto y valores positivos de mediana. En esta zona los errores disminuyen (0,5 mm d<sup>-1</sup>) (Figura 3).

Respecto del producto ERA 5, los estadísticos indican errores cercanos a 1 mm d<sup>-1</sup> en ambos sitios de estudio, asimismo, la mediana es determinada por valores positivos, indicando una leve sobreestimación de los datos. En cuanto a MERRA-2, en ambas zonas, se determina un error de 1,2 mm d<sup>-1</sup>, mientras que la mediana arroja valores positivos y muy similares, determinando la sobreestimación por parte del modelo (Figura 3).

En general, Ruhoff *et al.* (2022) evaluaron distintos productos de satélite y reanálisis de ET en América del Sur, analizando los ingresos y egresos de agua del sistema. En ese sentido, encontraron que, el producto *TerraClimate* muestra resultados equilibrados en las cuencas del Amazonas, así como en otras cuencas tropicales, mientras que, en Tocantins (Brasil), se observan excesos de agua. Por su parte, el producto MOD16A2 presenta valores de desequilibrio más elevados. Así como en la cuenca del Paraná y San Francisco (Brasil), *TerraClimate* mostró un desequilibrio menor al presentado por MOD16A2. Por su parte, en la zona de Uruguay y el Atlántico Sur, tanto el producto ERA5 como MOD16A2 mostraron déficit de agua disponible. En un estudio a escala global, Lv *et al.* (2020) evaluaron los datos de MERRA-2 a nivel global, llegando a la conclusión de que esta base de datos es comparable con otras para el análisis y estudio de la ET en sus distintos productos.

Por otro lado, respecto del algoritmo SVR y los datos de entrada que se utilizaron, diversos autores analizaron las variables disponibles en la base de datos NASA Power, así como los resultados obtenidos con SVR. Como se indica en Degano *et al.* (2023), Orte *et al.* (2021) analizaron datos de entrada de Rs (NASA Power) con medidas *in situ* de la red de irradiancia solar Saver-Net instalada en el sur de América del Sur. Encontrando una correlación relevante (R<sup>2</sup> de 0,95), y

**Tabla 3: Estadísticos asociados a los productos de ET real.**

Producto	Número de datos	RMSE [mm d <sup>-1</sup> ]	MAE [mm d <sup>-1</sup> ]	R <sup>2</sup>	Me [mm d <sup>-1</sup> ]	RSD [mm d <sup>-1</sup> ]	NSE
TerraClimate	399	1.2	0.9	0.6	0.8	0.5	-1.7
MOD16A2	1722	0.8	0.6	0.6	0.6	0.3	0.4
SVR	97161	0.6	0.4	0.67	0.9	0.5	0.66

**Tabla 3: . Estadísticos asociados a los productos de ETp<sub>Hargreaves</sub>**

Producto	Número de datos	RMSE [mm d <sup>-1</sup> ]	MAE [mm d <sup>-1</sup> ]	R <sup>2</sup>	Me [mm d <sup>-1</sup> ]	RSD [mm d <sup>-1</sup> ]	NSE
TerraClimate	399	1.0	0.8	0.73	-0.8	0.4	0.6
ERA 5	12329	0.8	0.6	0.82	0.4	0.4	0.95
MERRA-2	12329	1.1	0.1	0.81	0.6	0.4	0.88

los errores muestran una diferencia esperable (RMSE de 0,48 kWh m<sup>-2</sup> d<sup>-1</sup>; MAE de 0,63 kWh m<sup>-2</sup> d<sup>-1</sup>). En el análisis realizado por Faramiñán *et al.* (2021b), determinaron ET real con datos NASA Power como entrada al algoritmo SVR, comparándolo con datos de 12 estaciones hidrometeorológicas de la RPA. Obteniendo el R<sup>2</sup> (0,67), RMSE (0,54 mm d<sup>-1</sup>) y MAE (0,4 mm d<sup>-1</sup>). Los resultados muestran la factibilidad de aplicar algoritmos de aprendizaje automático para la obtención de valores de ET real en planicies agrícolas.

Respecto de la plataforma *Climate Engine*, que es una aplicación web, de fácil acceso y con diferentes productos con potencial para el análisis hidroclimático, se concluye que posee productos de calidad apropiada para el estudio y análisis avanzado de las cuencas hidrológicas de la región Pampeana Argentina. Estas comparaciones con estudios realizados en distintos sitios de América del Sur indican la factibilidad y variabilidad esperable para la metodología llevada a cabo en este trabajo.

### Conclusiones

El objetivo principal de este trabajo es analizar los productos de evapotranspiración de referencia (ET<sub>0</sub>), real (ET real) y potencial (ETp), determinados con datos de reanálisis y de satélite, disponibles en la plataforma virtual *Climate Engine*, de fácil acceso para la Región Pampeana Argentina. Además, se evaluó el algoritmo de inteligencia artificial, *Support Vector Machine Regression*, para determinar la ET real y ETp, haciendo uso de los datos de la plataforma NASA Power.

En general, se observa un muy buen comportamiento del conjunto de productos evaluados para su uso en la región de estudio. Dado los errores obtenidos, se puede decir que, en mayor o menor medida, todos pueden ser utilizados para la determinación de la evapotranspiración de referencia, real y potencial. Asimismo, el que presenta mejores resultados (para ET real y ETp) y valores altos del índice de eficiencia de Nash-Sutcliffe es el obtenido mediante inteligencia artificial, teniendo en cuenta que la resolución espacial es de 50 Km por 50 Km, con un bajo nivel de detalle de parcelas.

En particular, para los productos disponibles en la plataforma *Climate Engine*, obtiene mejores resultados el modelo *TerraClimate* para la variable ET<sub>0</sub>, mientras que MOD16A2 muestra errores menores para la ET real, y ERA 5 es más apropiado para determinar la ETp mediante Hargreaves.

### Agradecimientos

Los autores del presente trabajo quieren agradecer a la Secretaría de Ciencia, Arte y Tecnología de la Universidad Nacional del Centro de la provincia de Buenos Aires (UNCPBA) en el marco del programa de Fortalecimiento de la Ciencia y la Tecnología, proyecto Jóvenes Investigadores (03-JO-VIN-711), a la UNCPBA, al Instituto de Hidrología de Llanuras “Dr. Eduardo Jorge Usunoff”, a la Oficina de Riesgo Agropecuario por proporcionar los datos de evapotranspiración de campo, al MSc. Martín Bayala por su colaboración con la confección de las figuras en alta calidad y al Dr. Adán Faramiñán por facilitarnos datos e información acerca del algoritmo SVR. Finalmente, a la alumna de grado, Melina Abdala, por su contribución en la descarga de datos.

### Contribución de autoría

Ambos autores contribuyeron a la concepción y el diseño del estudio. La preparación del material, la recopilación y el análisis de datos fueron realizados por María Florencia Degano. El primer borrador del manuscrito fue escrito por María Florencia Degano. La revisión del manuscrito fue realizada por Raúl Rivas y ambos autores comentaron las versiones anteriores del manuscrito. Ambos autores leyeron y aprobaron el manuscrito final.



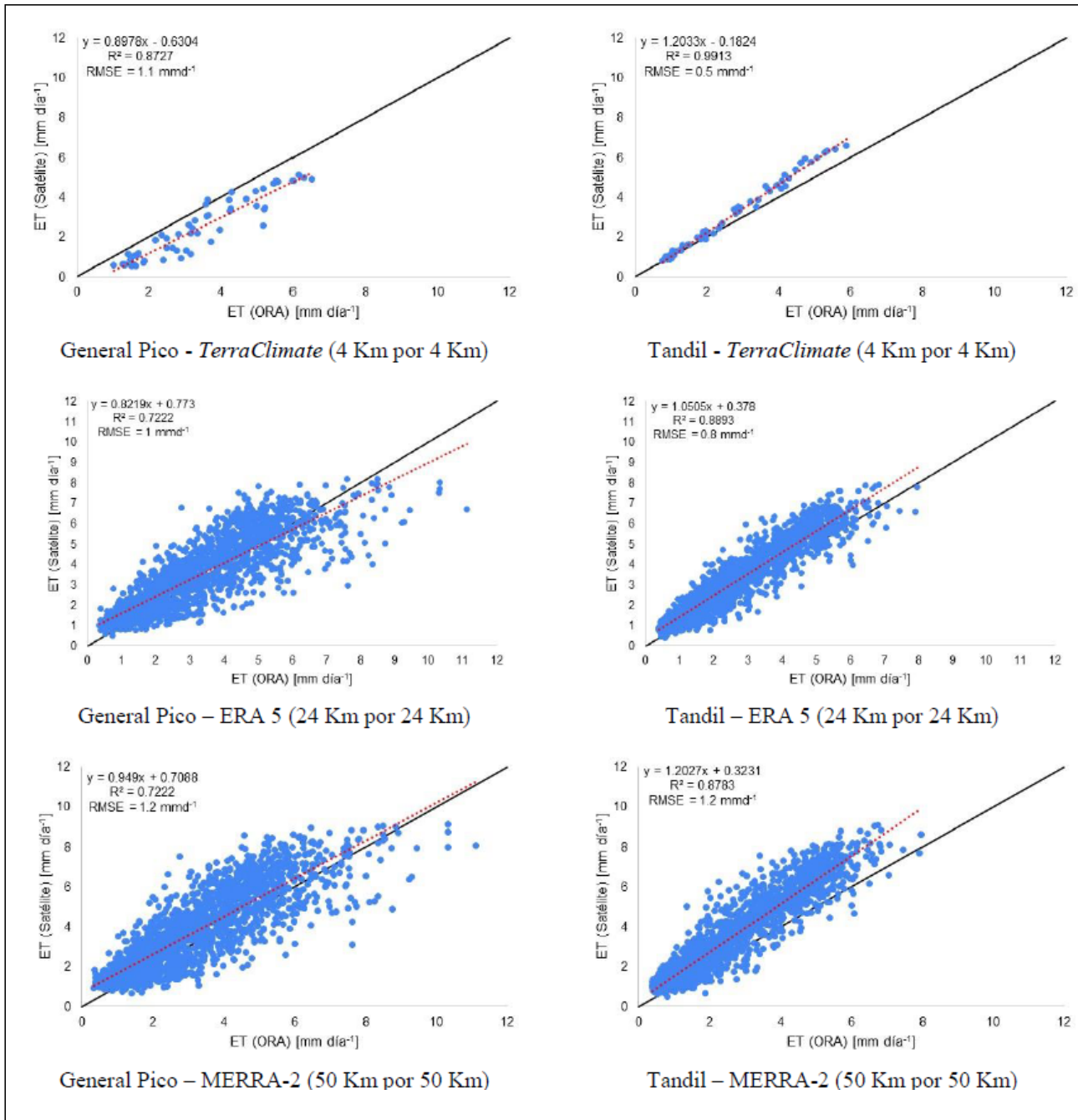


Figura 3: Relación entre los productos de ETp TerraClimate, ERA 5 y MERRA-2 en las estaciones General Pico (semiárida) y Tandil (subhúmeda). Línea negra: relación 1:1. Línea punteada roja: línea de tendencia.

### Declaración de conflictos de intereses

Los Autores declaran que no existe algún conflicto de interés

## Referencias bibliográficas

- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A. y Hegewisch, K. C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958-2015. *Scientific Data*, 5,170191. <https://doi.org/10.1038/sdata.2017.191>
- Allen, R. G., Pereira, L. S., Raes, D. y Smith, M. (1998). *Crop evapotranspiration-Guidelines for computing crop water requirements*. FAO Irrigation and drainage paper N°56. FAO.
- Allen, R., Pereira, L., Howell, T. y Jensen, M. (2011). Evapotranspiration information reporting: I. Factors governing measurement accuracy. *Agricultural Water Management*, 98(6), 899-920. <https://doi.org/10.1016/j.agwat.2010.12.015>
- Basualdo, A. (2011). Balance hídrico como herramienta de decisión.. En S. Occhiuzzi, P. Mercuri y C. Pascale (Coords.), *Herramientas para la evaluación y gestión del riesgo climático en el sector agropecuario* (pp. 55-63). (1ª ed.). Ministerio de Agricultura, Ganadería y Pesca de la Nación.
- Castañeda Calatayud, F. (2020). Análisis del potencial eólico en la Región de Aysén a partir del modelo ERA 5 REANALYSIS. Tesis de grado. Facultad de Ciencias Agronómicas, Universidad de Chile. <https://repositorio.uchile.cl/handle/2250/183486>
- Chia, M. Y., Huang, Y. F., Koo, C. H. y Fung, K. F. (2020a). Recent advances in evapotranspiration estimation using artificial intelligence approaches with a focus on hybridization techniques—a review. *Agronomy*, 10(1), 101. <https://doi.org/10.3390/agronomy10010101>.
- Chia, M. Y., Huang, Y. F. y Koo, C. H. (2020b). Support vector machine enhanced empirical reference evapotranspiration estimation with limited meteorological parameters. *Computers and Electronics in Agriculture*, 175, 105577. <https://doi.org/10.1016/j.compag.2020.105577>
- Copernicus Climate Change Service (C3S). (2017). ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. *Copernicus Climate Change Service Climate Data Store (CDS)*, date of access. <https://cds.climate.copernicus.eu/>
- Cristianini, N. y Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University press.
- Degano, M. F., Rivas, R. E. y Bayala, M. I. (2023). Determinación de la evapotranspiración con datos satelitales y de reanálisis utilizando Google Earth Engine. *Tecnología y Ciencias del Agua*. <https://doi.org/10.24850/j-tyca-15-4-4>
- Degano, M. F., Rivas, R. E., Carmona, F., Faramiñán, A. M. G. y Olivera Rodríguez, P. S. (2021a). Calibración del producto de evapotranspiración potencial “MOD16A2” para la Región Pampeana Argentina. *Boletín Geológico y Minero de España*, 132(1-2), 167-174. <https://doi.org/10.21701/bolgeomin.132.1-2.017>
- Degano, M. F., Rivas, R. E., Carmona, F., Niclòs, R. y Sánchez, J. M. (2021b). Evaluation of the MOD16A2 evapotranspiration product in an agricultural area of Argentina, the Pampas region. *The Egyptian Journal of Remote Sensing and Space Science*, 24(2), 319-328. <https://doi.org/10.1016/j.ejrs.2020.08.004>
- Fan, J., Yue, W., Wu, L., Zhang, F., Cai, H., Wang, X., Lu, X. y Xiang, Y. (2018). Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. *Agricultural and Forest Meteorology*, 263, 225-241. <https://doi.org/10.1016/j.agrformet.2018.08.019>
- Faramiñán A. M. G. (2022). Estudio de la evapotranspiración y su aporte en el desarrollo de modelos por medio de medidas directas e inteligencia artificial. Tesis doctoral. Facultad de Ciencias Exactas, Universidad Nacional del Centro de la provincia de Buenos Aires.
- Faramiñán, A. M. G., Carmona, F., Rivas, R. E., Degano, M. F. y Abadie, P. A. (2021a). Monitoreo de la evapotranspiración real horaria por medio del balance energético y un lisímetro de pesada. *Boletín Geológico y Minero de España*, 132(1-2), 47-56. <http://hdl.handle.net/11336/164282>
- Faramiñán, A. M., Degano, M. F., Carmona, F. y Rodríguez, P. O. (2021 b). Estimation of actual evapotranspiration using NASA-POWER data and Support Vector Machine. En *2021 XIX Workshop on Information Processing and Control (RPIC)* (pp. 1-5). IEEE. <http://dx.doi.org/10.1109/RPIC53795.2021.9648425>
- Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A., Gu, W., .... y Zhao B. (2017). The modern-era retrospective analysis for research and applications, Version 2 (MERRA-2). *Journal of Climate*, 30(13), 5419–5454.

- Hargreaves, G. H. y Samani, Z. A. (1985). Reference crop evapotranspiration from temperature. *Applied Engineering in Agriculture*, 1(2): 96-99. <http://dx.doi.org/10.13031/2013.26773>
- Huntington, J. L., Hegewisch, K. C., Daudert, B., Morton, C. G., Abatzoglou, J. T., McEvoy, D. J. y Erickson, T. (2017). Climate engine: Cloud computing and visualization of climate and remote sensing data for advanced natural resource monitoring and process understanding. *Bulletin of the American Meteorological Society*, 98(11), 2397-2410. <http://dx.doi.org/10.1175/BAMS-D-15-00324.1>
- Juarez, J., Trentin, G. y Heinen, M. (2018). Determinación de evapotranspiración de referencia a partir de modelos de inteligencia artificial. En *X Congreso de AgroInformática (CAI)-JAIIO 47* (pp. 160-168). Sociedad Argentina de Informática e Investigación Operativa.
- Liakos, K. G., Busato, P., Moshou, D., Pearson, S. y Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors*, 18(8), 2674. <https://doi.org/10.3390/s18082674>.
- Lv, M., Xu, Z. y Lv, M. (2020). Evaluating hydrological processes of the atmosphere-vegetation interaction model and ME-RRA-2 at global scale. *Atmosphere*, 12(1), 16.
- McCuen, R. H., Knight, Z. y Cutter, A. G. (2006). Evaluation of the Nash-Sutcliffe Efficiency index. *Journal of Hydrologic Engineering*, 11(6), 597-602. [https://doi.org/10.1061/\(ASCE\)1084-0699\(2006\)11:6\(597\)](https://doi.org/10.1061/(ASCE)1084-0699(2006)11:6(597))
- Monteith J. L. y Unsworth M. H. (1990). *Principles of Environmental Physics*. (2ª ed.). E. Arnold.
- Mu, Q. Z., Zhao y M.S., Running, S. W. (2013). MODIS Global Terrestrial Evapotranspiration (ET) product (NASA MOD16A2/A3). Algorithm Theoretical Basis Document. Collection 5. NASA Headquarters. *Numerical Terradynamic Simulation Group*, 268. [https://scholarworks.umd.edu/ntsg\\_pubs/268](https://scholarworks.umd.edu/ntsg_pubs/268)
- Orte, F., Lusi, A., Carmona, F., D'Elia, R., Faramiñán, A. y Wolfram, E. (2021). Comparison of NASA-POWER solar radiation data with ground-based measurements in the south of South America. En *2021 XIX Workshop on Information Processing and Control (RPIC)* (pp. 1-4). IEEE. <https://doi.org/10.1109/RPIC53795.2021.9648428>
- Pereyra, F. X. (2003). *Ecoregiones de la Argentina*. Servicio Geológico Minero Argentino. <http://repositorio.segemar.gov.ar/308849217/2953>
- Ritter, A. y Muñoz-Carpena, R. (2013). Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology*, 480, 33-45. <https://doi.org/10.1016/j.jhydrol.2012.12.004>
- Ruhoff, A., de Andrade, B. C., Laipelt, L., Fleischmann, A. S., Siqueira, V. A., Moreira, A. A., Barbedo, R., Cyganski, G.L., Fernandez, G.M.R., Brêda, J.P.L.F., Paiva, R.C.D.d., Meller, A., Teixeira, A.d.A., Araújo, A.A., Fuckner, M.A., Biggs, T. (2022). Global Evapotranspiration Datasets Assessment Using Water Balance in South America. *Remote Sensing*, 14(11), 2526. <https://doi.org/10.3390/rs14112526>
- Running, S. W., Mu, Q. Z. y Zhao, M. S. (2017). MOD16A2 MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V006. *NASA EOSDIS Land Processes DAAC*, 6. <https://doi.org/10.5067/MODIS/MOD16A2.006>
- Shrestha, N. K. y Shukla, S. (2015). Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment. *Agricultural and Forest Meteorology*, 200, 172-184. <https://doi.org/10.1016/j.agrformet.2014.09.025>
- Weinzettel P. y Usunoff E. (2001). Cálculo de la recarga mediante la aplicación de la ecuación de Darcy en la zona no saturada. En A. Medina, J. Carrera (Eds.), *Las caras del agua subterránea. Tomo II* (pp. 225-261). IGME.