

## NUEVA BASE DE DATOS DE PRECIPITACIONES MENSUALES DE LA REPÚBLICA ARGENTINA (PMRAv1), 2000-2022

Juan J. Gaitan<sup>1</sup>, Lucio Biancari<sup>2</sup>

<sup>1</sup>Departamento de Tecnología, Universidad Nacional de Luján. CONICET, Luján, Buenos Aires, Argentina

<sup>2</sup>Cátedra de Ecología e IFEVA, Facultad de Agronomía, Universidad de Buenos Aires. CONICET, Buenos Aires, Argentina

(Manuscrito recibido el 18 de septiembre de 2023, en su versión final el 5 de marzo de 2024)

### RESUMEN

La representación precisa de los patrones espacio-temporales de precipitación es un insumo esencial para numerosas aplicaciones ambientales. Sin embargo, la estimación de los patrones de precipitación derivados únicamente de pluviómetros está sujeta a grandes incertidumbres, especialmente en regiones con escasez de datos. Presentamos una nueva base de datos de Precipitaciones Mensuales de la República Argentina (PMRAv1) para el período 2000-2022, con 5 km de resolución espacial. PMRAv1 utiliza una metodología basada en regresión de bosques aleatorios (Regression Random Forest) para combinar datos mensuales de mediciones terrestres (entre 142 y 227 estaciones en cada mes), cuatro productos globales de precipitación estimada por satélite, modelos de circulación atmosféricas globales o interpolación de datos medidos en terreno (TERRACLIMATE, ERA5-LAND, GPMv6 y PERSIANN-CDR) con el objetivo de mejorar la estimación de las precipitaciones mensuales en Argentina. La metodología desarrollada pudo mejorar la representación espacio-temporal de la precipitación al permitir la fusión de múltiples fuentes de información satelital y mediciones en terreno. La validación realizada utilizando el 30% de los datos de pluviómetros mostró que PMRAv1 mejora significativamente los parámetros RMSE, MAE, EM y  $R^2$  en comparación con los cuatro productos globales de precipitación. Además, el producto resultó más estable en la predicción de los valores mensuales observados, al presentar un menor desvío estándar en los tres parámetros de ajuste. La base de datos PMRAv1 se pone a disposición de los usuarios en varios formatos. El término 'v1' (versión 1) hace referencia a que se considera que este producto tendrá sucesivas versiones en el futuro que permitan actualizarla y mejorar la precisión de las estimaciones. Asimismo, el método presentado también podría ser utilizado para mejorar la estimación de otras variables climatológicas cuando se disponga de datos medidos en terreno.

*Palabras clave: Regression Random Forest, estaciones meteorológicas, estimación precipitaciones.*

## NEW MONTHLY PRECIPITATION DATABASE OF ARGENTINA (PMRAv1), 2000-2022

### ABSTRACT

The accurate representation of spatio-temporal precipitation patterns is an essential input for numerous environmental applications. However, the estimation of precipitation patterns derived solely from rain gauges is subject to large uncertainties, especially in data-scarce regions. We present a new database of Monthly Precipitation of the Argentine Republic (PMRAv1) for the period 2000-2022, with 5 km of spatial resolution. PMRAv1 employs a Regression Random Forest-based methodology to combine monthly data from ground measurements (ranging from 142 to 227 rain gauges in each month) and four global precipitation products derived from satellite-estimated precipitation, global atmospheric circulation models or interpolation of ground-measured data (TERRACLIMATE, ERA5-LAND, GPMv6 y PERSIANN-CDR) with the aim of improving the estimation of monthly precipitation in Argentina. The developed methodology enhanced the spatio-temporal representation of precipitation by allowing the fusion of multiple sources of satellite information and ground measurements. Validation performed using 30% of the rain gauge data showed that PMRAv1 significantly improves the parameters RMSE, MAE, EM and  $R^2$  compared to the four global precipitation products. Furthermore, the model was more stable in predicting the observed monthly values, presenting a lower standard deviation in the three adjustment parameters. The PMRAv1 database is available to users in several formats. The term 'v1' (version 1) refers to the fact that this product will have successive future versions in the future that will update and refine the precision of the estimates. Likewise, the method presented could also be used to improve the estimation of other climatological variables when ground-based data are available.

*Keywords: Regression Random Forest, rain gauges, precipitation estimation.*

## 1. INTRODUCCIÓN

La precipitación es una variable clave en el estudio de los flujos de materia y energía que ocurren en el planeta Tierra, afectando diversos procesos hidrológicos, meteorológicos y ecológicos. La adquisición de datos de precipitación de alta resolución espacio-temporal es esencial para investigar los procesos de superficie y el cambio climático global, dado que constituye un parámetro de entrada fundamental de la meteorología de superficie en varios modelos de fisiología vegetal, ecología, hidrología y otros campos (Andersen et al., 2002; Stisen et al., 2012;

Qi et al., 2018; Zhang et al., 2020). Sin embargo, la precipitación es una de las variables más difíciles de representar a nivel regional debido a su naturaleza discreta y a la marcada heterogeneidad espacial y temporal que presenta (Almonacid et al., 2021). La precipitación se mide comúnmente con estaciones de pluviómetros, con una alta precisión en ubicaciones específicas (Villarini et al., 2008). La precisión de la representación de patrones espaciales de precipitación depende de la densidad y la configuración de la red de medición (García et al., 2008; Adhikary et al., 2015). Las redes pluviométricas de terreno están distribuidas escasa y desigualmente en

la mayoría de las regiones del mundo (Sharifi et al., 2019, Beck et al., 2020), por lo que el uso exclusivo de mediciones terrestres para estimar la distribución espacial de la precipitación a escala regional está sujeto a grandes incertidumbres (Woldemeskel et al., 2013).

Desde la década de 1980, la estimación de la precipitación basada en la tecnología de sensores remotos se ha desarrollado rápidamente, proporcionando información, espacialmente continua, sobre la distribución de la precipitación a alta resolución espacial y temporal. Se han desarrollado bases de datos grilladas globales o cuasi-globales de precipitación, interpoladas a distinta resolución espacial a partir de redes globales de estaciones meteorológicas, información satelital y modelado numérico (Almonacid et al., 2021). Los productos globales de precipitación (PGP) abordan las limitaciones antes mencionadas de las redes pluviométricas y están cada vez más disponibles para el público, mejorando significativamente nuestra comprensión de las características de la precipitación (Chen y Gao, 2018; Gupta et al., 2019; Markonis et al., 2019). Diferentes estudios han demostrado que estos PGP logran representar adecuadamente la distribución espacial de esta variable y las tendencias de largo plazo a escala continental, pero pueden presentarse diferencias significativas con la distribución real del fenómeno a escalas regionales y para estudios hidrológicos a escala de cuenca (Getirana et al., 2011; Pessacg et al., 2015; Bianchi et al., 2016). Según Zhang et al. (2021) los PGP están sujetos a grandes incertidumbres y sesgos que surgen de los algoritmos de recuperación, las mediciones indirectas y las deficiencias de los sensores. Por ejemplo, las falsas detecciones de precipitación por sensores infrarrojos si existen nubes frías que no precipitan, y la infravaloración de eventos localizados de tormenta por sensores de microondas. Por lo tanto, sigue existiendo la necesidad de mejorar la estimación de la distribución espacio-temporal de la precipitación mediante

la combinación de diferentes fuentes de datos, como PGP e información terrestre (Xie y Xiong, 2011). En los últimos años se han puesto en marcha numerosos estudios para mejorar la precisión y reducir los sesgos de los PGP en cuanto a la estimación de la intensidad de la precipitación (Báez-Villanueva et al., 2020; Zhang et al., 2021).

La fusión es un método para combinar información útil de varios productos de precipitación en un solo producto que tiene una mejor precisión (Chao et al., 2018). En los últimos años se han desarrollado varios enfoques basados en algoritmos de Machine Learning (ML) y Deep Learning (DL) que demostraron ser muy adecuados para fusionar múltiples PGP con datos de observación terrestre para estimar la variabilidad espacio-temporal de la precipitación (Nguyen et al., 2023; Mohammed et al., 2023). Random Forest (RF; Breiman, 2001) es un método perteneciente al grupo de algoritmos ML que ha mostrado ventajas para fusionar múltiples PGP con datos de observación terrestre: i) es una técnica precisa y capaz de manejar tamaños de muestra pequeños y se comporta bien incluso cuando algunas variables explicativas aportan baja información a la predicción y cuando se utilizan varias covariables, principalmente porque no produce estimaciones sesgadas ni conduce a un sobreajuste (Biau y Scornet, 2016); ii) tiene una buena capacidad para mapear relaciones no lineales entre las covariables y la variable de salida; iii) aunque RF es una técnica no espacial, puede considerar indirectamente covariables geográficas (p. ej., coordenadas geográficas) (Báez-Villanueva et al., 2020).

Los objetivos del presente trabajo son: i) generar y poner a disposición de los usuarios una base de datos de Precipitación Mensual de la República Argentina versión 1 (PMRAv1) para el período 2000-2022 y ii) mostrar una posible aplicación de este producto: elaborar un mapa de la tendencia de las precipitaciones anuales en dicho período para PMRAv1 y los cuatro PGP. Para ello se presenta un procedimiento de fusión

basado en Random Forest que aprovecha la combinación de información de diferentes PGP y series temporales de precipitación medidas en pluviómetros de terreno. Este método de fusión se basa en tres supuestos clave (Báez-Villanueva et al., 2020): i) las mediciones de precipitación en la red de pluviómetros son precisas en la escala de puntos; ii) los PGP generalmente están sesgados pero contienen información útil sobre los patrones espacio-temporales de precipitación; y iii) la combinación de diferentes PGP y datos pluviométricos puede brindar una mejor representación de la variabilidad espacio-temporal de precipitación que cualquier producto individual.

## 2. MATERIALES Y MÉTODOS

### 2.1. Área de estudio

El territorio americano de la República Argentina abarca una superficie de 2,80 millones de km<sup>2</sup> y es el octavo país del mundo en extensión. La geografía es muy diversa, encontrándose mayormente llanuras en el este, serranías en el centro y montañas en el oeste. El país se encuentra atravesado, en su extremo occidental, de norte a sur por la cordillera de los Andes. El pico más alto de Argentina es el Aconcagua, de 6960,8 m s. n. m. Existe una gran variedad de climas determinados principalmente por la extensión del país y la presencia de la Cordillera de los Andes hacia el oeste. Dos tercios del país están definidos por condiciones áridas o semiáridas a partir de una marcada diagonal árida desde el noroeste hacia el centro-este del país. Los climas subtropicales húmedos y húmedos predominan en la región Pampeana y noreste del país, mientras que los áridos y fríos se encuentran en el sur y oeste. Estas características se traducen en una alta heterogeneidad en la precipitación media anual para todo el territorio de la República Argentina, la cual varía desde menos de 200 mm en la mencionada diagonal árida, hasta alrededor de 2000 mm en el noreste y en la cordillera patagónica.

### 2.2. Adquisición de datos

#### 2.2.1. Datos pluviométricos

Se recopilieron datos de precipitación diaria, entre enero de 2000 y diciembre de 2022, de estaciones meteorológicas provenientes de redes oficiales de la República Argentina: Servicio Meteorológico Nacional (SMN, <https://www.smn.gov.ar/>, accedido el 10/09/2023), red meteorológica del Instituto Nacional de Tecnología Agropecuaria (INTA, <http://siga.inta.gov.ar/>, accedido el 10/09/2023), Base de Datos del Sistema Nacional de Información Hídrica (SNIH, <https://snih.hidricosargentina.gov.ar/Filtros.aspx>, accedido el 10/09/2023); Autoridad Interjurisdiccional de las Cuencas de los ríos Limay, Neuquén y Negro (AIC, <http://www.aic.gov.ar/sitio/home>, accedido el 10/09/2023) y redes meteorológicas de las provincias de La Pampa (<https://lapampa.re-desclimaticas.com/accounts/login/?next=/>, accedido el 10/09/2023), Córdoba (<https://clima.bccba.org.ar/accounts/login/?next=/>, accedido el 10/09/2023) y San Luis (<http://www.clima.sanluis.gov.ar/Index.aspx>, accedido el 10/09/2023).

Los registros diarios fueron agregados a escala mensual, considerándose sin datos aquellos meses con menos de 28 días medidos. Las series temporales de las estaciones que cumplieron este requisito fueron evaluadas a partir de la detección de valores extremos de precipitación siguiendo el protocolo de exploración gráfica de Zuur et al. (2010). El análisis de valores extremos en cada estación tenía el objetivo de detectar errores en el proceso de medición o en la entrada de los datos a las planillas de cálculo. Sin embargo, los valores más extremos detectados correspondían a eventos de precipitación elevada validados con otros registros (por ej. artículos periódicos). Por lo tanto, no fue necesario excluir ningún valor extremo de las series temporales. Como resultado se compiló una base de datos de 385 estaciones con series temporales incompletas (Gaitán y Biancari, 2024). Diciembre de 2022 fue el mes con menor

cantidad datos (163 estaciones) y diciembre de 2011 fue el mes con mayor cantidad de datos (243 estaciones) (Figura 1).

### 2.2.2. Covariables

Se seleccionó un conjunto de 6 covariables potencialmente predictoras de la precipitación mensual: cuatro PGP y dos variables geográficas (latitud y longitud). La resolución del producto final PMRAv1 y la original de cada covariable fue llevada a una resolución común de 5000 m, que es similar a la del PGP de mayor detalle espacial (TERRACLIMATE).

La selección de los PGP se realizó de acuerdo a los siguientes criterios: i) que se encuentren disponibles en la plataforma Google Earth Engine (GEE); ii) que abarquen la totalidad del territorio de la República Argentina y iii) que cuenten con datos de precipitaciones mensuales (o que se puedan agregar a dicha escala temporal) para el periodo 2000-2022. De acuerdo con estos criterios se seleccionaron los siguientes cuatro PGP:

ERA5-LAND (ee. ImageCollection: 'ECMWF/ERA5\_LAND/MONTHLY\_AGGR') es un conjunto de datos de reanálisis que combina datos del modelo de predicción numérica ECMWF (European Centre for Medium-Range Weather Forecasts) y observaciones de todo el mundo con una resolución espacial de 11132 metros (Muñoz Sabater et al., 2021).

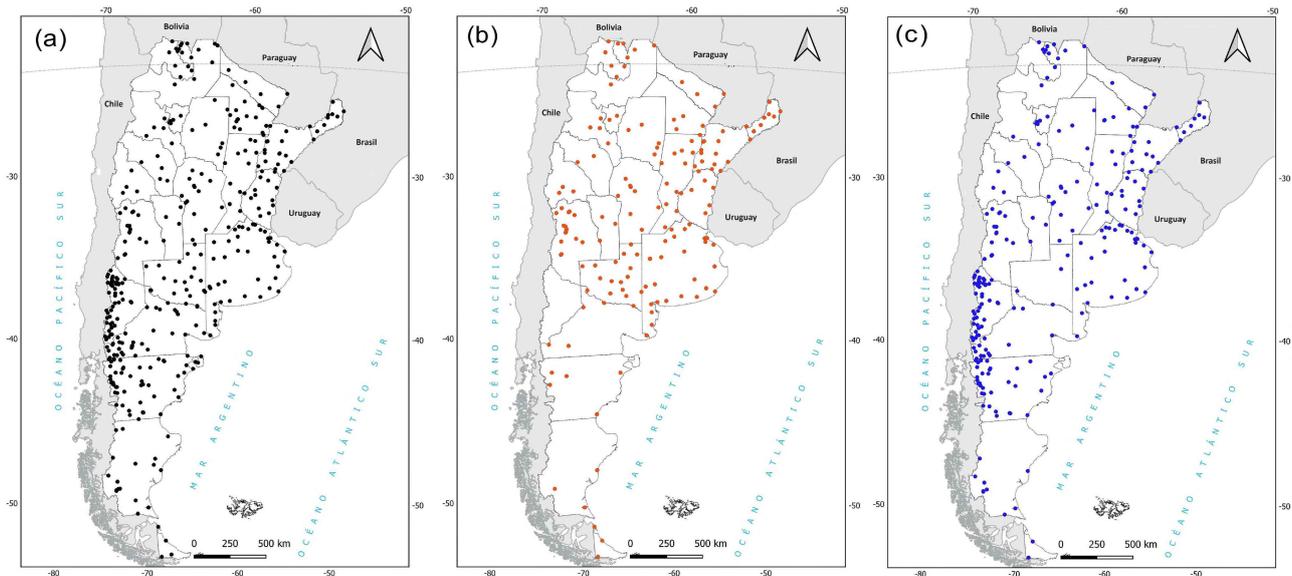
TERRACLIMATE (ee. ImageCollection: 'IDAHO\_EPSCOR/TERRACLIMATE') utiliza la interpolación de datos medidos en terreno, combinando normales climatológicas de alta resolución espacial del conjunto de datos WorldClim version 2 (Fick et al., 2017) con datos de resolución espacial más gruesa, pero que varían en el tiempo de CRU Ts4.0 (Harris et al., 2017) y el reanálisis japonés de 55 años (JRA55; Kobayashi, et al., 2015). La resolución espacial es de 4638 metros (Abatzoglou, et al., 2018).

GPMv6 (ee. ImageCollection: 'NASA/GPMc\_L3/IMERG\_V06') la Medición de Precipitación Global (GPM) es una misión satelital internacional que comprende un conjunto de satélites que llevan instrumentos de microondas pasivos y activos diseñados para observar la estructura y la intensidad de las precipitaciones y nevadas. Junto con las observaciones geoestacionarias infrarrojas, estos datos permiten calcular y publicar productos de precipitación global con una resolución espacial de 11132 metros (Huffman et al., 2019).

PERSIANN-CDR (ee. ImageCollection: 'NOAA/PERSIANN-CDR') estima la precipitación a partir de información de detección remota mediante un algoritmo basado en redes neuronales artificiales que utiliza datos infrarrojos (IR) del satélite geoestacionario GridSat-B1. Las estimaciones logradas se ajustan utilizando el producto de precipitación GPCP, generando grillas con una resolución espacial es de 27830 metros (Ashouri et al., 2015).

### 2.3. Modelización y validación

En este estudio utilizamos el algoritmo Regression Random Forest para estimar las precipitaciones mensuales del período 2000-2022 en la República Argentina mediante la fusión de información de diferentes PGP e información de terreno (datos de pluviómetros). Este algoritmo construye un conjunto de árboles de regresión (en este estudio utilizamos 200 árboles). Cada uno de estos árboles estima el resultado en cada píxel y el valor final resulta del promedio de los 200 árboles (Breiman, 2001). Cada árbol se construye a partir de una muestra seleccionada al azar del conjunto de datos de terreno. Esto permite una estimación robusta del error con el conjunto de muestras restantes, las llamadas muestras Out-Of-Bag (OOB). Los valores de las muestras OOB excluidas se estiman a partir de las muestras seleccionadas y, al contrastar los valores estimados con los observados, se pueden calcular diferentes métricas para evaluar el rendimiento de los productos.



*Figura 1:* Total de estaciones meteorológicas utilizadas para modelar PMRAv1 ( $n = 385$ ) (a), y meses con menor (diciembre 2022,  $n = 163$ ) (b) y mayor (diciembre 2011,  $n = 243$ ) (c) cantidad de estaciones disponibles.

Para elaborar la base de datos grillados PMRAv1 se ajustaron modelos entre enero de 2000 y diciembre de 2022 (276 modelos mensuales). La performance de cada modelo mensual fue evaluada mediante un proceso de validación cruzada con 10 corridas. Para ello, el conjunto original de datos pluviométricos fue dividido al azar en 10 conjuntos en los cuales el 70 % fue usado para calibrar el modelo y el 30 % para validar. Cuantificamos la performance de los modelos mensuales mediante la media del error cuadrático medio (RMSE), del error absoluto medio (MAE), del error medio (EM) y del coeficiente de determinación ( $R^2$ ), en las 10 corridas de la validación cruzada (Legates y McCabe, 1999).

#### 2.4. Tendencia de la precipitación anual (2000 – 2022)

Los datos de precipitación mensual fueron agregados a escala anual y se cartografió su tendencia temporal a lo largo de los 23 años para PMRAv1 y los cuatro PGP. Para ello se calculó la pendiente de la regresión lineal entre el tiempo (variable  $x$ ) y la precipitación anual (variable  $y$ ) en cada píxel y se expresó como

porcentaje en relación a la precipitación medial anual del período. Se identificaron las áreas con tendencias significativas para un nivel de probabilidad  $p < 0.05$ .

### 3. RESULTADOS

La Tabla I resume los estadísticos promedios de los 276 meses de la serie temporal que permiten evaluar el desempeño del producto PMRAv1 en comparación a cuatro PGP (TERRACLIMATE, ERA5-LAND, GPMv6 y PERSIANN-CDR). La base PMRAv1 mostró una mejora al disminuir entre un 15 % y 20 % el RMSE y el MAE, y aumentar un 12 % el  $R^2$  en relación a los promedios de todos los meses para los dos PGP de mejor ajuste (TERRACLIMATE y GPMv6). Además, PMRAv1 tuvo los valores de desvío estándar más bajos en comparación a los otros productos para RMSE, MAE, EM y  $R^2$  (Tabla I). El EM fue cercano a 0 para PMRAv1 (promedio de toda la serie = 0.66 mm), mientras que TERRACLIMATE en general subestimó la media de la precipitación mensual alrededor de 5 mm y ERA5-LAND, GPMv6 y PERSIANN-CDR la sobreestimaron entre 7 mm y 16 mm. La tendencia del ajuste de

los productos indica en general una disminución en RMSE y  $R^2$  a lo largo del tiempo para PMRAv1 (Figura 2). El EM para PMRAv1 fue estable y cercano a 0 a lo largo del tiempo (EM PMRAv1 =  $0.723-0.0004x$ ,  $p=0.76$ ), mientras que GPMv6 aumentó su sobreestimación (EM GPMv6 =  $1.38+0.039x$ ,  $p<0.0001$ ) y ERA5-LAND disminuyó su sobreestimación (EM ERA5-LAND =  $23.72-0.056x$ ,  $p<0.0001$ ) a lo largo del tiempo. Los  $R^2$  de PMRAv1, TERRACLIMATE, ERA5-LAND y GPMv6 disminuyen a lo largo de la serie temporal ( $p<0.001$ ), mientras PERSIANN-CDR tiene una pendiente positiva pero no significativa ( $p=0.13$ ). El análisis de tendencia de las precipitaciones en el período 2000-2022 muestra, para PMRAv1, una disminución en la región centro-norte del país y en la región cordillerana central, mientras que se observa una tendencia de aumento en la estepa patagónica. Las áreas con tendencia, estadísticamente significativa, negativa representaron el 25.2 % de la superficie, mientras que las áreas con tendencia positiva fueron el 1.3%. Para los cuatro PGP en promedio, la superficie con tendencia negativa fue del 29.5 % (variando entre 16.1 % en GPMv6 y 44.5 % en ERA5-LAND) y la superficie con tendencia positiva fue del 0.7 % (entre 0.0 % en ERA5-LAND y 1.6 % en GPMv6). La comparación entre PMRAv1 y los cuatro PGP en la distribución espacial de las áreas con tendencia significativa negativa, positiva y sin tendencia, indica una coincidencia global del 76.6 % entre PMRAv1 y TERRACLIMATE, del 73.1 % con PERSIANN-CDR, del 72.3 % con GPMv6 y del 68.0 % con ERA5-LAND (Figura 3).

La base de datos PMRA se pone a disposición de los usuarios en diferentes formatos:

1. APP: mediante el siguiente enlace se accede a una APP desarrollada en GEE que permite visualizar los 23 mapas anuales de precipitación. Asimismo, la aplicación permite delimitar un área de interés y generar un gráfico de la serie temporal de precipitaciones mensuales. Además, estos datos se pueden descargar en formato .csv: <https://ee-precipitacionesmensuales.projects.earthengine.app/view/pmrav1>
2. Drive: en el siguiente enlace de Drive se puede descargar la serie temporal de imágenes en formato ráster de las precipitaciones mensuales: <https://drive.google.com/drive/folders/1c4OUapwH5QwLwZ-oxQV1-Q6wrQkhV7Gy?usp=sharing>
3. GEE: el siguiente enlace permite el acceso a la serie temporal de imágenes en formato ráster de las precipitaciones mensuales: <https://code.earthengine.google.com/32467858bba614f981bbbefc94269776>

#### 4. DISCUSIÓN Y CONCLUSIONES

Las estimaciones de precipitación basadas en satélites, reanálisis e interpolaciones de mediciones en terreno brindan una oportunidad sin precedentes para numerosas aplicaciones hidrológicas, meteorológicas y ambientales. A pesar de las continuas mejoras de los PGP, todavía existen diferentes tipos de desajustes en la mayoría de ellos. Aquí presentamos un método capaz de obtener estimaciones de precipitación mejoradas al combinar información de PGP disponibles públicamente y datos de estaciones pluviométricas. El algoritmo de Regression Random Forest pudo mejorar la representación espacio-temporal de la precipitación al permitir la fusión de múltiples fuentes de información satelital y mediciones en terreno, lo cual es consistente con los resultados obtenidos en otros estudios (Verdin et al., 2016; Manz et al., 2016).

La validación, que se llevó a cabo utilizando el 30 % de los datos de pluviómetros, reveló una mejora significativa en los parámetros RMSE, MAE, EM y  $R^2$  de PMRAv1, en comparación con los cuatro PGP individuales. Además, el producto resultó más estable en la predicción de los valores mensuales observados, al presentar un menor desvío estándar en los cuatro parámetros de ajuste. La disminución en el parámetro RMSE a lo largo del tiempo, indica

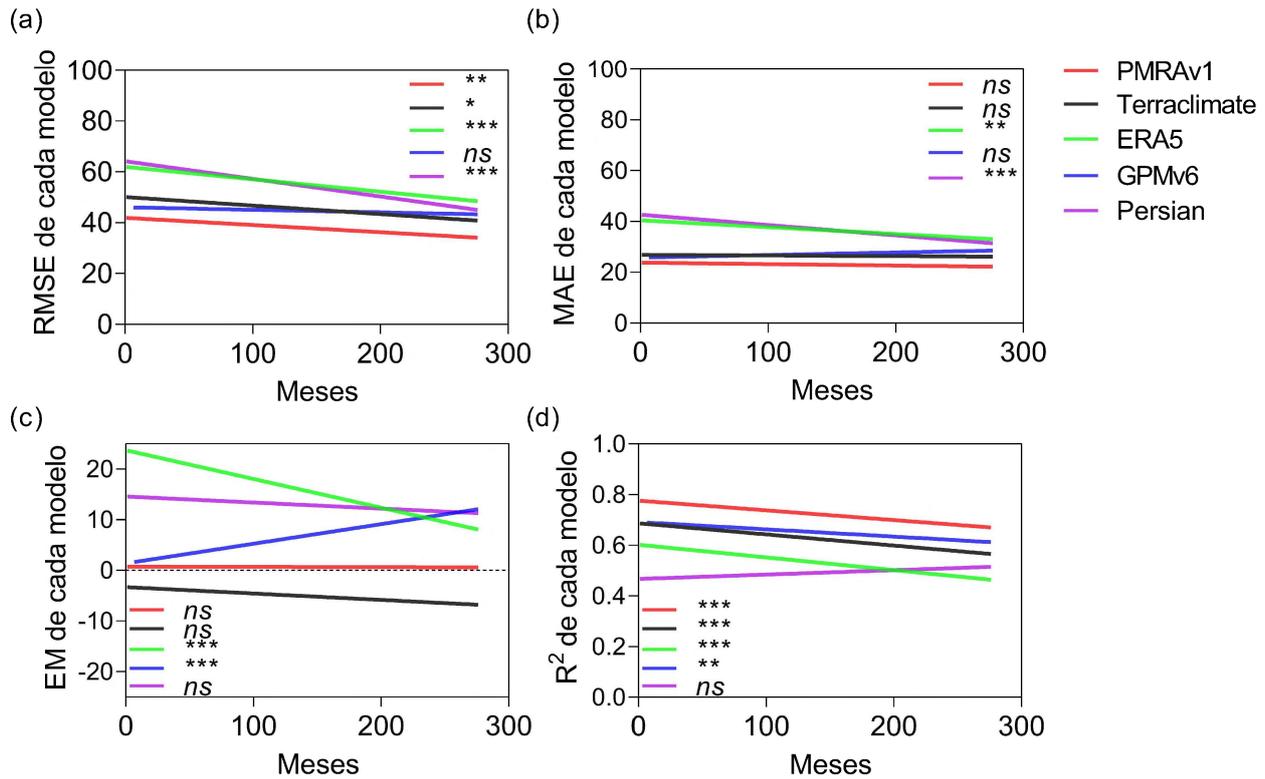


Figura 2: Ajuste de cada producto de precipitación a lo largo de los 276 meses estudiados. Cada panel representa los distintos parámetros de ajuste estudiados: (a) RMSE, (b) MAE, (c) EM y, (d)  $R^2$ . En cada panel se detalla el nivel de significancia de la prueba de hipótesis para pendientes distintas de 0: \* =  $p < 0.05$ , \*\* =  $p < 0.01$ , \*\*\* =  $p < 0.001$  y ns = no significativo.

que PMRAv1 es cada vez más certero en la predicción de la media de los valores observados (Legates y McCabe, 1999). Sin embargo, las reducciones observadas en el  $R^2$  de los PGP (a excepción de PERSIANN-CDR) indicarían que a lo largo del tiempo explican una porción menor de la variabilidad de los datos. Dado que PMRAv1 fusiona la información de los otros modelos, esto podría explicar que también disminuya su  $R^2$  en el tiempo.

En el análisis de tendencias hubo una coincidencia alta entre PMRAv1 y los otros productos evaluados (>70%), pero se observaron diferencias en algunas áreas específicas. Por ejemplo, mientras algunos productos subestimaron las tendencias negativas para el área centro-norte del país y en la región cordillerana central, otros

subestimaron las tendencias positivas en la estepa patagónica. Las áreas con tendencias negativas de precipitación observadas con PMRAv1 coinciden en gran medida con las áreas que poseen tendencias significativas de déficit hídrico y de disminución en la humedad del suelo (Peretti et al., 2023). Contar con un producto de precipitaciones validado para Argentina es de gran importancia para mejorar las proyecciones del cambio climático y sus impactos. Por ejemplo, en los Andes se han propuesto cambios en la distribución espacial de plantas y biomas relacionados al cambio climático (Tovar et al., 2022), y también se han asociado las tendencias históricas de pérdida en la masa de glaciares con eventos de sequía prolongados y disminución en las precipitaciones (Falaschi et al., 2023). A su vez, en la estepa patagónica donde la disponibilidad de agua es

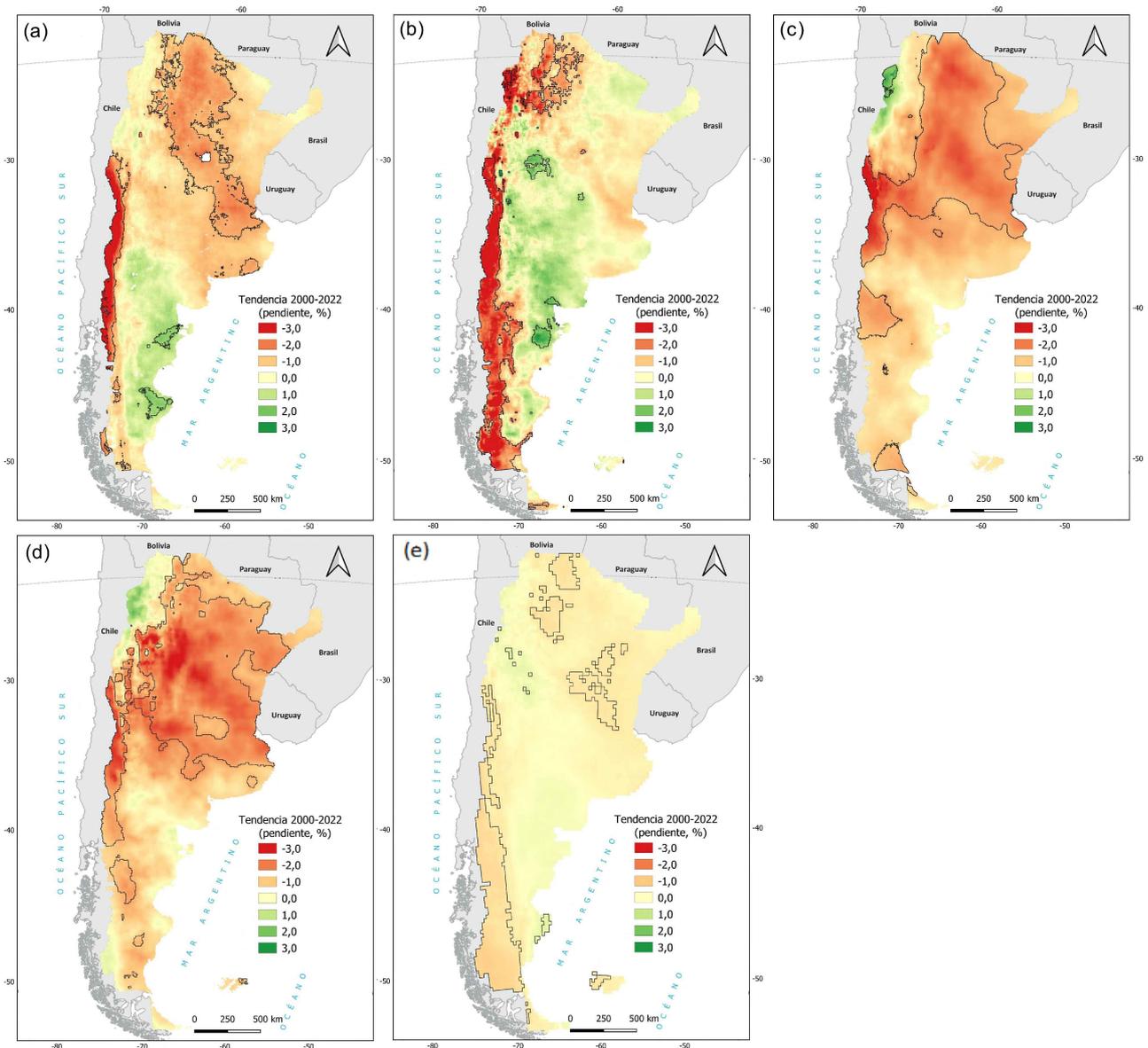


Figura 3: Pendiente de la tendencia de la precipitación durante el período 2000-2022, expresada como porcentaje de la media del período, para PMRAv1 (a) y cuatro productos globales de precipitación: GPMv6 (b), TERRACLIMATE (c), ERA5-LAND (d) y PERSIANN-CDR (e). Las líneas sobre el mapa indican las áreas con tendencia estadísticamente significativa ( $p < 0.05$ ).

la principal limitante para la supervivencia y crecimiento de las plantas, tendencias positivas en la precipitación podrían tener efectos sobre la cobertura vegetal y la coexistencia entre especies leñosas y herbáceas (Biancari et al., 2023).

En este estudio ponemos a disposición de los usuarios la base de datos PMRAv1. El

término ‘v1’ (versión 1) hace referencia a que consideramos que este producto tendrá sucesivas versiones en el futuro que permitan actualizarla y mejorar la precisión de las estimaciones. Para mejorar las estimaciones identificamos tres posibles vías: i) incrementar la densidad de datos de pluviómetro: estudios previos han mostrado que la precisión de los modelos fusionados mediante RF tiende a incrementarse

*Tabla I:* Resumen de los parámetros de ajuste RMSE (raíz del error cuadrático medio), MAE (error absoluto medio), EM (error medio) y  $R^2$  para PMRAv1 y los cuatro productos globales de precipitación (TERRACLIMATE, ERA5-LAND, GPMv6 y PERSIANN-CDR). Se muestran medias, desvíos estándar, valores mínimos y máximos para los 276 meses de la serie temporal 2000-2022.

	Modelo	Media	Desvío estándar	Mínimo	Máximo
RMSE	PMRAv1	38	13.9	12.7	81.8
	TERRACLIMATE	45.5	19.3	12.9	108.3
	ERA5-LAND	55.2	16.5	17.4	106.4
	GPMv6	44.7	17.5	13.7	122
	PERSIANN-CDR	54.6	20	20.6	152
MAE	PMRAv1	23	8.15	7.25	44.2
	TERRACLIMATE	26.5	10.3	7.58	58.26
	ERA5-LAND	36.7	11.1	10	68.9
	GPMv6	27.2	10.6	8.07	77.6
	PERSIANN-CDR	37	13.1	12.5	122.7
EM	PMRAv1	0.66	1.83	-4.49	6.41
	TERRACLIMATE	-5.05	8.84	-40	11.3
	ERA5-LAND	15.9	8.97	-7.05	46.5
	GPMv6	6.85	8.82	-16.8	57.3
	PERSIANN-CDR	12.9	8.05	-4.3	62
$R^2$	PMRAv1	0.72	0.09	0.31	0.87
	TERRACLIMATE	0.63	0.12	0.11	0.84
	ERA5-LAND	0.53	0.14	0.12	0.83
	GPMv6	0.65	0.12	0.15	0.86
	PERSIANN-CDR	0.49	0.15	0.03	0.83

al incluir un mayor número de datos de terreno (Báez-Villanueva et al., 2020; Zhang et al., 2021); ii) modificar los PGP utilizados: las versiones futuras se pueden mejorar incluyendo solamente aquellos productos globales que tengan mejor ajuste a los datos observados y iii) regionalizar el país: por ejemplo, ajustando modelos por región de acuerdo a patrones de precipitación homogéneos. Luego, comparar si el producto consolidado a nivel nacional mejora la estimación respecto a un único producto nacional.

La base de datos generada, PMRAv1, mejora la estimación de los PGP para la República Argentina, y ha sido puesta a disposición de manera abierta en diversos formatos.

Contar con productos climáticos validados y abiertos de alcance nacional puede traducirse en aplicaciones y modelos más certeros para diversas disciplinas que requieran datos de precipitación como insumo. En este trabajo se demostró una posible aplicación de este producto para estimar tendencias en las precipitaciones, lo cual tiene gran valor para caracterizar con mayor certeza el cambio climático. Finalmente, el método presentado en este estudio también se podría utilizar para mejorar la estimación espacio-temporal de otras variables climatológicas cuando se disponga de datos medidos en terreno.

#### REFERENCIAS

Abatzoglou JT, Dobrowski SZ, Parks SA,

- Hegewisch KC, 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>
- Adhikary SK, Yilmaz AG, Muttill N, 2015: Optimal design of rain gauge network in the Middle Yarra River catchment, Australia. *Hydrological Processes*, 29(11), 2582-2599. <http://doi.org/10.1002/hyp.10389>
- Almonacid L, Pessacg N, Diaz BG, Bonfili O, Peri PL, 2021: Nueva base de datos reticulada de precipitación para la provincia de Santa Cruz, Argentina. *Centro Argentino de Meteorólogos (CAM)*. <https://doi.org/10.24215/1850468Xe007>
- Andersen J, Dybkjaer G, Jensen KH, Refsgaard JC, Rasmussen K, 2002: Use of remotely sensed precipitation and leaf area index in a distributed hydrological model. *Journal of Hydrology*, 264(1-4), 34-50. [http://doi.org/10.1016/S0022-1694\(02\)00046-X](http://doi.org/10.1016/S0022-1694(02)00046-X)
- Ashouri H, Hsu KL, Sorooshian S, Braithwaite DK, Knapp KR, Cecil LD, Nelson BR, Prat OP, 2015: Persiann-CDR daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bull. Am. Meteorol. Soc.*, 96, 69–83. <http://doi.org/10.1175/BAMS-D-13-00068.1>
- Baez-Villanueva OM, Zambrano-Bigiarini M, Beck HE, McNamara I, Ribbe L, Nauditt A, Birkel C, Verbist K, Giraldo-Osorio JD, Thinh NX, 2020: RF-MEP: A novel Random Forest method for merging gridded precipitation products and ground-based measurements. *Remote Sensing of Environment*, 239, 111606. <https://doi.org/10.1016/j.rse.2019.111606>
- Biancari L, Oñatibia GR, Gaitán JJ, Aguiar MR, 2023: Coexistence of grasses and shrubs in Patagonian steppes. Norm or exception? *Journal of Vegetation Science*, 34, e13177. <https://doi.org/10.1111/jvs.13177>
- Chao L, Zhang K, Li Z, Zhu Y, Wang J, Yu Z, 2018: Geographically Weighted Regression Based Methods for Merging Satellite and Gauge Precipitation. *Journal of Hydrology*, 558, 275–289. <https://doi.org/10.1016/j.jhydrol.2018.01.042>
- Falaschi D, Berthier E, Belart JM, Bravo C, Castro M, Durand M, Villalba R, 2023: Increased mass loss of glaciers in Volcán Domuyo (Argentinian Andes) between 1962 and 2020, revealed by aerial photos and satellite stereo imagery. *Journal of Glaciology*, 69(273), 40–56. <http://doi.org/10.1017/jog.2022.43>
- Fick SE, Hijmans RJ, 2017: WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.*, 37, 4302–4315. <http://doi.org/10.1002/joc.5086>
- Gaitan, Juan; Biancari, Lucio (2024). Anexo A - Series temporales. *Precipitacion mensual por estacion.xlsx*. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.25356649.v1>
- Garcia M, Peters-Lidard CD, Goodrich DC, 2008: Spatial interpolation of precipitation in a dense gauge network for monsoon storm events in the southwestern United States. *Water Resources Research*, 44(5). <http://doi.org/10.1029/2006WR005788>
- Harris I, Jones PD, Osborn TJ, Lister DH, 2017: CRU TS4.00: Climatic Research Unit (CRU) Time-Series (TS) version 4.00 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2015). Centre for Environmental Data Analysis. <https://doi.org/10.5285/edf8febfaad48abb2cbaf7d7e846a86>
- Huffman GJ, Stocker EF, Bolvin DT, Nelkin EJ, Jackson T, 2019: GPM IMERG final precipitation L3 half hourly 0.1 degree × 0.1 degree V06, Greenbelt, MD, Goddard earth sciences data and information services center (GES DISC). <https://doi.org/10.5067/GPM/IMERG/3B-HH/06>
- Kobayashi S, Ota Y, Harada Y, Ebata A, Moriya M, Onoda H, Onogi K, Kamahori H, Kobayashi C, Endo H, Miyaoka K, Takahashi K, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan*. Ser. II, 93(1), 5-48. <https://doi.org/10.2151/jmsj.2015-001>
- Legates DR, McCabe GJ, 1999: Evaluating

- the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.*, 35(1), 233–241. <http://doi.org/10.1029/1998WR900018>
- Manz B, Buytaert W, Zulkafli Z, Lavado W, Willems B, Robles LA, Rodríguez-Sánchez JP, 2016: High-resolution satellite-gauge merged precipitation climatologies of the Tropical Andes. *J. Geophys. Res. Atmos.*, 121, 1190–1207. <https://doi.org/10.1002/2015JD023788>
- Mohammed J, Mengiste Y, Singh V, 2023: Improving spatio-temporal precipitation estimates in data scarce river basins: an application of machine learning-based multi-source data merging. *Stochastic Environmental Research and Risk Assessment*, 37(4), 1353–1369. <http://doi.org/10.1007/s00477-022-02346-4>
- Muñoz-Sabater J, Dutra E, Agustí-Panareda A, Albergel C, Arduini G, Balsamo G, Boussetta S, Choulga M, Harrigan S, Hersbach H, Martens B, Miralles DG, Piles M, Rodríguez-Fernández NJ, Zsoter E, Buontempo C, Thépaut JN, 2021: ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data*, 13, 4349–4383. <https://doi.org/10.5194/essd-13-4349-2021>
- Nguyen GV, Le XH, Van LN, May DTT, Jung S, Lee G, 2023: Machine learning approaches for reconstructing gridded precipitation based on multiple source products. *Journal of Hydrology: Regional Studies*, 48, 101475. <https://doi.org/10.1016/j.ejrh.2023.101475>
- Nguyen GV, Le XH, Van LN, Jung S, Yeon M, Lee G, 2021: Application of random forest algorithm for merging multiple satellite precipitation products across South Korea. *Remote Sensing*, 13(20), 4033. <https://doi.org/10.3390/rs13204033>
- Peretti M, Spennemann PC, Long MEF, 2023: Trends in soil moisture content and water deficits in Argentina and the role of climate contribution. *Theor Appl Climatol*, 152, 1189–1201. <http://doi.org/10.1007/s00704-023-04428-x>
- Qi J, Markewitz D, Radcliffe D, 2018: Modelling the effect of changing precipitation inputs on deep soil water utilization. *Hydrological Processes*, 32(5), 672–686. <http://doi.org/10.1002/hyp.11452>
- Stisen S, Højberg AL, Troldborg L, Refsgaard JC, Christensen BSB, Olsen M, Henriksen HJ, 2012: On the importance of appropriate precipitation gauge catch correction for hydrological modelling at mid to high latitudes. *Hydrology and Earth System Sciences*, 16(11), 4157–4176. <http://doi.org/10.5194/hessd-9-3607-2012>
- Tovar, C., Carril, A.F., Gutiérrez, A.G., Ahrends, A., Fita, L., Zaninelli, P., Flombuam, P., Abarzúa, A.M., Alarcón, D., Aschero, V., Báez, S., Barros, A., Carilla, J., Ferrero, M.E., Flantua, S.G.A., Gonzáles, P., Menéndez, C.G., Pérez-Escobar, O.A., Pauchard, A., Ruscica, R.C., Särkinen, T., Sörensson, A.A., Srur, A., Villalba, R., y Hollingsworth, P.M., 2022: Understanding climate change impacts on biome and plant distributions in the Andes: Challenges and opportunities. *Journal of Biogeography*, 49(8), 1420–1442. <http://doi.org/10.1111/jbi.14389>
- Verdin, A., Funk, C., Rajagopalan, B., y Kleiber, W., 2016: Kriging and Local Polynomial Methods for Blending Satellite-Derived and Gauge Precipitation Estimates to Support Hydrologic Early Warning Systems. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5), 2552–256. <http://doi.org/10.1109/TGRS.2015.2502956>
- Villarini, G., y Krajewski, W.F., 2008: Empirically-based modeling of spatial sampling uncertainties associated with rainfall measurements by rain gauges. *Advances in Water Resources*, 31(7), 1015–1023. <https://doi.org/10.1016/j.advwatres.2008.04.007>
- Villarini, G., y Krajewski, W.F., 2009: Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall. *Surv. Geophys.*, 31, 107–129. <http://doi.org/10.1007/s10712-009-9079-x>

- Wang, L., Jiao, W., MacBean, N., Rulli, M.C., Manzoni, S., Vico, G., y D'Odorico, P., 2022: Dryland productivity under a changing climate. *Nature Climate Change*, 12, 981–994. <http://doi.org/10.1038/s41558-022-01499-y>
- Woldemeskel, F.M., Sivakumar, B., y Sharma, A., 2013: Merging gauge and satellite rainfall with specification of associated uncertainty across Australia. *Journal of Hydrology*, 499, 167-176. <http://doi.org/10.1016/j.jhydrol.2013.06.039>
- Xie, P., y Xiong, A.Y., 2011: A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. *Journal of Geophysical Research: Atmospheres*, 116(D21). <http://doi.org/10.1029/2011JD016118>
- Zhang, L., Li, X., Zheng, D., Zhang, K., Ma, Q., Zhao, Y., y Ge, Y., 2021: Merging multiple satellite-based precipitation products and gauge observations using a novel double machine learning approach. *Journal of Hydrology*, 594, 125969. <http://doi.org/10.1016/j.jhydrol.2021.125969>
- Zhang, L., Ren, D., Nan, Z., Wang, W., Zhao, Y., Ma, Q., y Wu, X., 2020: Interpolated or satellite-based precipitation? Implications for hydrological modeling in a meso-scale mountainous watershed on the Qinghai-Tibet Plateau. *Journal of Hydrology*, 583, 124629. <https://doi.org/10.1016/j.jhydrol.2020.124629>
- Zuur, A.F., Ieno, E.N., y Elphick, C.S., 2010: A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1, 3–14. <http://doi.org/10.1111/j.2041-210X.2009.00001.x>