

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

3

4

Juan J. Gaitan¹, Lucio Biancari²

5

6 ¹Departamento de Tecnología, Universidad Nacional de Luján. CONICET, Luján,
7 Buenos Aires, Argentina.

8

9

²Cátedra de Ecología e IFEVA, Facultad de Agronomía, Universidad de Buenos
Aires. CONICET, Buenos Aires, Argentina.

10

11

Autor correspondiente: Juan J. Gaitan, jgaitan@unlu.edu.ar

12

13

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

14

15

RESUMEN

16

17

18

19

20

21

22

23

24

25

26

27

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

Artículo en edición

28 la representación espacio-temporal de la precipitación al permitir la fusión de múltiples
 29 fuentes de información satelital y mediciones en terreno. La validación realizada
 30 utilizando el 30% de los datos de pluviómetros mostró que PMRAv1 mejora
 31 significativamente los parámetros RMSE, MAE, EM y R^2 en comparación con los cuatro
 32 productos globales de precipitación. Además, el producto resultó más estable en la
 33 predicción de los valores mensuales observados, al presentar un menor desvío estándar
 34 en los tres parámetros de ajuste. La base de datos PMRAv1 se pone a disposición de los
 35 usuarios en varios formatos. El término ‘v1’ (versión 1) hace referencia a que se considera
 36 que este producto tendrá sucesivas versiones en el futuro que permitan actualizarla y
 37 mejorar la precisión de las estimaciones. Asimismo, el método presentado también podría
 38 ser utilizado para mejorar la estimación de otras variables climatológicas cuando se
 39 disponga de datos medidos en terreno.

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

42

43 **NEW MONTHLY PRECIPITATION DATABASE OF ARGENTINA (PMRAv1), 2000-**
 44 **2022**

45

ABSTRACT

46 The accurate representation of spatio-temporal precipitation patterns is an essential input
 47 for numerous environmental applications. However, the estimation of precipitation
 48 patterns derived solely from rain gauges is subject to large uncertainties, especially in
 49 data-scarce regions. We present a new database of Monthly Precipitation of the Argentine
 50 Republic (PMRAv1) for the period 2000-2022, with 5 km of spatial resolution. PMRAv1
 51 employs a Regression Random Forest-based methodology to combine monthly data from
 52 ground measurements (ranging from 142 to 227 rain gauges in each month) and four
 53 global precipitation products derived from satellite-estimated precipitation, global
 54 atmospheric circulation models or interpolation of ground-measured data
 55 (TERRACLIMATE, ERA5-LAND, GPMv6 y PERSIANN-CDR) with the aim of
 56 improving the estimation of monthly precipitation in Argentina. The developed
 57 methodology enhanced the spatio-temporal representation of precipitation by allowing

58 the fusion of multiple sources of satellite information and ground measurements.
 59 Validation performed using 30% of the rain gauge data showed that PMRAv1
 60 significantly improves the parameters RMSE, MAE, EM and R^2 compared to the four
 61 global precipitation products. Furthermore, the model was more stable in predicting the
 62 observed monthly values, presenting a lower standard deviation in the three adjustment
 63 parameters. The PMRAv1 database is available to users in several formats. The term ‘v1’
 64 (version 1) refers to the fact that this product will have successive future versions in the
 65 future that will update and refine the precision of the estimates. Likewise, the method
 66 presented could also be used to improve the estimation of other climatological variables
 67 when ground-based data are available.

68 **Key Words:** Regression Random Forest, rain gauges, precipitation estimation.

69 1) INTRODUCCIÓN

70 La precipitación es una variable clave en el estudio de los flujos de materia y energía que
 71 ocurren en el planeta Tierra, afectando diversos procesos hidrológicos, meteorológicos y
 72 ecológicos. La adquisición de datos de precipitación de alta resolución espacio-temporal
 73 es esencial para investigar los procesos de superficie y el cambio climático global, dado
 74 que constituye un parámetro de entrada fundamental de la meteorología de superficie en
 75 varios modelos de fisiología vegetal, ecología, hidrología y otros campos (Andersen et
 76 al., 2002; Stisen et al., 2012; Qi et al., 2018; Zhang et al., 2020). Sin embargo, la
 77 precipitación es una de las variables más difíciles de representar a nivel regional debido
 78 a su naturaleza discreta y a la marcada heterogeneidad espacial y temporal que presenta
 79 (Almonacid et al., 2021). La precipitación se mide comúnmente con estaciones de
 80 pluviómetros, con una alta precisión en ubicaciones específicas (Villarini et al., 2008). La
 81 precisión de la representación de patrones espaciales de precipitación depende de la
 82 densidad y la configuración de la red de medición (García et al., 2008; Adhikary et al.,
 83 2015). Las redes pluviométricas de terreno están distribuidas escasa y desigualmente en
 84 la mayoría de las regiones del mundo (Sharifi et al., 2019, Beck et al., 2020), por lo que
 85 el uso exclusivo de mediciones terrestres para estimar la distribución espacial de la
 86 precipitación a escala regional está sujeto a grandes incertidumbres (Woldemeskel et al.,
 87 2013).

Artículo en edición

88 Desde la década de 1980, la estimación de la precipitación basada en la tecnología de
 89 sensores remotos se ha desarrollado rápidamente, proporcionando información,
 90 espacialmente continua, sobre la distribución de la precipitación a alta resolución espacial
 91 y temporal. Se han desarrollado bases de datos grilladas globales o cuasi-globales de
 92 precipitación, interpoladas a distinta resolución espacial a partir de redes globales de
 93 estaciones meteorológicas, información satelital y modelado numérico (Almonacid et al.,
 94 2021). Los productos globales de precipitación (PGP) abordan las limitaciones antes
 95 mencionadas de las redes pluviométricas y están cada vez más disponibles para el público,
 96 mejorando significativamente nuestra comprensión de las características de la
 97 precipitación (Chen y Gao, 2018; Gupta et al., 2019; Markonis et al., 2019). Diferentes
 98 estudios han demostrado que estos PGP logran representar adecuadamente la distribución
 99 espacial de esta variable y las tendencias de largo plazo a escala continental, pero pueden
 100 presentarse diferencias significativas con la distribución real del fenómeno a escalas
 101 regionales y para estudios hidrológicos a escala de cuenca (Getirana et al., 2011; Pessacq
 102 et al., 2015; Bianchi et al., 2016). Según Zhang et al. (2021) los PGP están sujetos a
 103 grandes incertidumbres y sesgos que surgen de los algoritmos de recuperación, las
 104 mediciones indirectas y las deficiencias de los sensores. Por ejemplo, las falsas
 105 detecciones de precipitación por sensores infrarrojos si existen nubes frías que no
 106 precipitan, y la infravaloración de eventos localizados de tormenta por sensores de
 107 microondas. Por lo tanto, sigue existiendo la necesidad de mejorar la estimación de la
 108 distribución espacio-temporal de la precipitación mediante la combinación de diferentes
 109 fuentes de datos, como PGP e información terrestre (Xie y Xiong, 2011). En los últimos
 110 años se han puesto en marcha numerosos estudios para mejorar la precisión y reducir los
 111 sesgos de los PGP en cuanto a la estimación de la intensidad de la precipitación (Báez-
 112 Villanueva et al., 2020; Zhang et al., 2021).

113 La fusión es un método para combinar información útil de varios productos de
 114 precipitación en un solo producto que tiene una mejor precisión (Chao et al., 2018). En
 115 los últimos años se han desarrollado varios enfoques basados en algoritmos de Machine
 116 Learning (ML) y Deep Learning (DL) que demostraron ser muy adecuados para fusionar
 117 múltiples PGP con datos de observación terrestre para estimar la variabilidad espacio-
 118 temporal de la precipitación (Nguyen et al., 2023; Mohammed et al., 2023). Random

119 Forest (RF; Breiman, 2001) es un método perteneciente al grupo de algoritmos ML que
 120 ha mostrado ventajas para fusionar múltiples PGP con datos de observación terrestre: i)
 121 es una técnica precisa y capaz de manejar tamaños de muestra pequeños y se comporta
 122 bien incluso cuando algunas variables explicativas aportan baja información a la
 123 predicción y cuando se utilizan varias covariables, principalmente porque no produce
 124 estimaciones sesgadas ni conduce a un sobreajuste (Biau y Scornet, 2016); ii) tiene una
 125 buena capacidad para mapear relaciones no lineales entre las covariables y la variable de
 126 salida; iii) aunque RF es una técnica no espacial, puede considerar indirectamente
 127 covariables geográficas (p. ej., coordenadas geográficas) (Báez-Villanueva et al., 2020).

128 Los objetivos del presente trabajo son: i) generar y poner a disposición de los usuarios
 129 una base de datos de Precipitación Mensual de la República Argentina versión 1
 130 (PMRAv1) para el período 2000-2022 y ii) mostrar una posible aplicación de este
 131 producto: elaborar un mapa de la tendencia de las precipitaciones anuales en dicho
 132 período para PMRAv1 y los cuatro PGP. Para ello se presenta un procedimiento de fusión
 133 basado en Random Forest que aprovecha la combinación de información de diferentes
 134 PGP y series temporales de precipitación medidas en pluviómetros de terreno. Este
 135 método de fusión se basa en tres supuestos clave (Báez-Villanueva et al., 2020): i) las
 136 mediciones de precipitación en la red de pluviómetros son precisas en la escala de puntos;
 137 ii) los PGP generalmente están sesgados pero contienen información útil sobre los
 138 patrones espacio-temporales de precipitación; y iii) la combinación de diferentes PGP y
 139 datos pluviométricos puede brindar una mejor representación de la variabilidad espacio-
 140 temporal de precipitación que cualquier producto individual.

141 2) MATERIALES Y MÉTODOS

142 2.1. Área de estudio

143 El territorio americano de la República Argentina abarca una superficie de 2,80 millones
 144 de km² y es el octavo país del mundo en extensión. La geografía es muy diversa,
 145 encontrándose mayormente llanuras en el este, serranías en el centro y montañas en el
 146 oeste. El país se encuentra atravesado, en su extremo occidental, de norte a sur por la
 147 cordillera de los Andes. El pico más alto de Argentina es el Aconcagua, de 6960,8 m s.
 148 n. m. Existe una gran variedad de climas determinados principalmente por la extensión
 149 del país y la presencia de la Cordillera de los Andes hacia el oeste. Dos tercios del país

Artículo en edición

150 están definidos por condiciones áridas o semiáridas a partir de una marcada diagonal árida
 151 desde el noroeste hacia el centro-este del país. Los climas subtropicales húmedos y
 152 húmedos predominan en la región Pampeana y noreste del país, mientras que los áridos y
 153 fríos se encuentran en el sur y oeste. Estas características se traducen en una alta
 154 heterogeneidad en la precipitación media anual para todo el territorio de la República
 155 Argentina, la cual varía desde menos de 200 mm en la mencionada diagonal árida, hasta
 156 alrededor de 2000 mm en el noreste y en la cordillera patagónica.

157 **2.2. Adquisición de datos**

158 **2.2.1. Datos pluviométricos**

159 Se recopilaron datos de precipitación diaria, entre enero de 2000 y diciembre de 2022, de
 160 estaciones meteorológicas provenientes de redes oficiales de la República Argentina:
 161 Servicio Meteorológico Nacional (SMN, <https://www.smn.gov.ar/>,
 162 10/09/2023), red meteorológica del Instituto Nacional de Tecnología Agropecuaria
 163 (INTA, <http://siga.inta.gov.ar/>,
 164 accedido el 10/09/2023), Base de Datos del Sistema
 165 Nacional de Información Hídrica (SNIH,
 166 <https://snih.hidricosargentina.gov.ar/Filtros.aspx>,
 167 accedido el 10/09/2023); Autoridad
 168 Interjurisdiccional de las Cuencas de los ríos Limay, Neuquén y Negro (AIC,
 169 <http://www.aic.gov.ar/sitio/home>,
 170 accedido el 10/09/2023) y redes meteorológicas de las
 171 provincias de La Pampa (<https://lapampa.redesclimaticas.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).

172 Los registros diarios fueron agregados a escala mensual, considerándose sin datos
 173 aquellos meses con menos de 28 días medidos. Las series temporales de las estaciones
 174 que cumplieron este requisito fueron evaluadas a partir de la detección de valores
 175 extremos de precipitación siguiendo el protocolo de exploración gráfica de Zuur et al.
 176 (2010). El análisis de valores extremos en cada estación tenía el objetivo de detectar
 177 errores en el proceso de medición o en la entrada de los datos a las planillas de cálculo.
 178 Sin embargo, los valores más extremos detectados correspondían a eventos de
 179 precipitación elevada validadas con otros registros (por ej. artículos periodísticos). Por lo

180 tanto, no fue necesario excluir ningún valor extremo de las series temporales. Como
 181 resultado se compiló una base de datos de 385 estaciones con series temporales
 182 incompletas (Anexo A, <https://doi.org/10.6084/m9.figshare.25356649.v1>). Diciembre de
 183 2022 fue el mes con menor cantidad de datos (163 estaciones) y diciembre de 2011 fue el
 184 mes con mayor cantidad de datos (243 estaciones) (Figura 1).

185 **2.2.2. Covariables**

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

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

197 ERA5-LAND (ee. ImageCollection: ‘ECMWF/ERA5_LAND/MONTHLY_AGGR’) es
 198 un conjunto de datos de reanálisis que combina datos del modelo de predicción numérica
 199 ECMWF (European Centre for Medium-Range Weather Forecasts) y observaciones de
 200 todo el mundo con una resolución espacial de 11132 metros (Muñoz Sabater et al., 2021).

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

207 GPMv6 (ee. ImageCollection: ‘NASA/GPM_L3/IMERG_V06’) la Medición de
 208 Precipitación Global (GPM) es una misión satelital internacional que comprende un

209 conjunto de satélites que llevan instrumentos de microondas pasivos y activos diseñados
210 para observar la estructura y la intensidad de las precipitaciones y nevadas. Junto con las
211 observaciones geoestacionarias infrarrojas, estos datos permiten calcular y publicar
212 productos de precipitación global con una resolución espacial de 11132 metros (Huffman
213 et al., 2019).

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

220 **2.3. Modelización y validación**

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

233 Para elaborar la base de datos grillados PMRAv1 se ajustaron modelos entre enero de
234 2000 y diciembre de 2022 (276 modelos mensuales). La performance de cada modelo
235 mensual fue evaluada mediante un proceso de validación cruzada con 10 corridas. Para
236 ello, el conjunto original de datos pluviométricos fue dividido al azar en 10 conjuntos en
237 los cuales el 70 % fue usado para calibrar el modelo y el 30 % para validar. Cuantificamos
238 la performance de los modelos mensuales mediante la media del error cuadrático medio

Artículo en edición

239 (RMSE), del error absoluto medio (MAE), del error medio (EM) y del coeficiente de
240 determinación (R^2), en las 10 corridas de la validación cruzada (Legates y McCabe,
241 1999).

242 **2.4. Tendencia de la precipitación anual (2000 – 2022)**

243 Los datos de precipitación mensual fueron agregados a escala anual y se cartografió su
244 tendencia temporal a lo largo de los 23 años para PMRAv1 y los cuatro PGP. Para ello se
245 calculó la pendiente de la regresión lineal entre el tiempo (variable x) y la precipitación
246 anual (variable y) en cada píxel y se expresó como porcentaje en relación a la
247 precipitación medial anual del período. Se identificaron las áreas con tendencias
248 significativas para un nivel de probabilidad $p < 0.05$.

249

250 **3) RESULTADOS**

251 La Tabla I resume los estadísticos promedios de los 276 meses de la serie temporal que
252 permiten evaluar el desempeño del producto PMRAv1 en comparación a cuatro PGP
253 (TERRACLIMATE, ERA5-LAND, GPMv6 y PERSIANN-CDR). La base PMRAv1
254 mostró una mejora al disminuir entre un 15% y 20% el RMSE y el MAE, y aumentar un
255 12% el R^2 en relación a los promedios de todos los meses para los dos PGP de mejor
256 ajuste (TERRACLIMATE y GPMv6). Además, PMRAv1 tuvo los valores de desvío
257 estándar más bajos en comparación a los otros productos para RMSE, MAE, EM y R^2
258 (Tabla I). El EM fue cercano a 0 para PMRAv1 (promedio de toda la serie = 0.66 mm),
259 mientras que TERRACLIMATE en general subestimó la media de la precipitación
260 mensual alrededor de 5 mm y ERA5-LAND, GPMv6 y PERSIANN-CDR la
261 sobreestimaron entre 7 mm y 16 mm.

262 La tendencia del ajuste de los productos indica en general una disminución en RMSE y
263 R^2 a lo largo del tiempo para PMRAv1 (Figura 2). El EM para PMRAv1 fue estable y
264 cercano a 0 a lo largo del tiempo ($EM_{PMRAv1} = 0.723 - 0.0004x$, $p = 0.76$), mientras que
265 GPMv6 aumentó su sobreestimación ($EM_{GPMv6} = 1.38 + 0.039x$, $p < 0.0001$) y ERA5-
266 LAND disminuyó su sobreestimación ($EM_{ERA5-LAND} = 23.72 - 0.056x$, $p < 0.0001$) a
267 lo largo del tiempo. Los R^2 de PMRAv1, TERRACLIMATE, ERA5-LAND y GPMv6

Artículo en edición

268 disminuyen a lo largo de la serie temporal ($p < 0.001$), mientras PERSIANN-CDR tiene
269 una pendiente positiva pero no significativa ($p = 0.13$).

270 El análisis de tendencia de las precipitaciones en el período 2000-2022 muestra, para
271 PMRAv1, una disminución en la región centro-norte del país y en la región cordillerana
272 central, mientras que se observa una tendencia de aumento en la estepa patagónica. Las
273 áreas con tendencia, estadísticamente significativa, negativa representaron el 25.2% de la
274 superficie, mientras que las áreas con tendencia positiva fueron el 1.3%. Para los cuatro
275 PGP en promedio, la superficie con tendencia negativa fue del 29.5% (variando entre
276 16.1% en GPMv6 y 44.5% en ERA5-LAND) y la superficie con tendencia positiva fue
277 del 0.7% (entre 0.0% en ERA5-LAND y 1.6% en GPMv6). La comparación entre
278 PMRAv1 y los cuatro PGP en la distribución espacial de las áreas con tendencia
279 significativa negativa, positiva y sin tendencia, indica una coincidencia global del 76.6%
280 entre PMRAv1 y TERRACLIMATE, del 73.1% con PERSIANN-CDR, del 72.3% con
281 GPMv6 y del 68.0% con ERA5-LAND (Figura 3).

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

283 1. **APP:** mediante el siguiente enlace se accede a una APP desarrollada en GEE que
284 permite visualizar los 23 mapas anuales de precipitación. Asimismo, la aplicación
285 permite delimitar un área de interés y generar un gráfico de la serie temporal de
286 precipitaciones mensuales. Además, estos datos se pueden descargar en formato
287 .csv:

288 <https://ee-precipitacionesmensuales.projects.earthengine.app/view/pmrav1>

289 2. **Drive:** en el siguiente enlace de Drive se puede descargar la serie temporal de
290 imágenes en formato ráster de las precipitaciones mensuales:

291 [https://drive.google.com/drive/folders/1c4OUapwH5QwLwZ-oxQV1-
292 Q6wrQkhV7Gy?usp=sharing](https://drive.google.com/drive/folders/1c4OUapwH5QwLwZ-oxQV1-Q6wrQkhV7Gy?usp=sharing)

293 3. **GEE:** el siguiente enlace permite el acceso a la serie temporal de imágenes en
294 formato ráster de las precipitaciones mensuales:

295 <https://code.earthengine.google.com/32467858bba614f981bbbec94269776>

296

297 **4) DISCUSIÓN Y CONCLUSIONES**

Artículo en edición

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

308 La validación, que se llevó a cabo utilizando el 30% de los datos de pluviómetros, reveló
 309 una mejora significativa en los parámetros RMSE, MAE, EM y R^2 de PMRAv1, en
 310 comparación con los cuatro PGP individuales. Además, el producto resultó más estable
 311 en la predicción de los valores mensuales observados, al presentar un menor desvío
 312 estándar en los cuatro parámetros de ajuste. La disminución en el parámetro RMSE a lo
 313 largo del tiempo, indica que PMRAv1 es cada vez más certero en la predicción de la
 314 media de los valores observados (Legates y McCabe, 1999). Sin embargo, las reducciones
 315 observadas en el R^2 de los PGP (a excepción de PERSIANN-CDR) indicarían que a lo
 316 largo del tiempo explican una porción menor de la variabilidad de los datos. Dado que
 317 PMRAv1 fusiona la información de los otros modelos, esto podría explicar que también
 318 disminuya su R^2 en el tiempo.

319 En el análisis de tendencias hubo una coincidencia alta entre PMRAv1 y los otros
 320 productos evaluados (>70%), pero se observaron diferencias en algunas áreas específicas.
 321 Por ejemplo, mientras algunos productos subestimaron las tendencias negativas para el
 322 área centro-norte del país y en la región cordillerana central, otros subestimaron las
 323 tendencias positivas en la estepa patagónica. Las áreas con tendencias negativas de
 324 precipitación observadas con PMRAv1 coinciden en gran medida con las áreas que
 325 poseen tendencias significativas de déficit hídrico y de disminución en la humedad del
 326 suelo (Peretti et al., 2023). Contar con un producto de precipitaciones validado para
 327 Argentina es de gran importancia para mejorar las proyecciones del cambio climático y
 328 sus impactos. Por ejemplo, en los Andes se han propuesto cambios en la distribución

Artículo en edición

329 espacial de plantas y biomas relacionados al cambio climático (Tovar et al., 2022), y
 330 también se han asociado las tendencias históricas de pérdida en la masa de glaciares con
 331 eventos de sequía prolongados y disminución en las precipitaciones (Falaschi et al.,
 332 2023). A su vez, en la estepa patagónica donde la disponibilidad de agua es la principal
 333 limitante para la supervivencia y crecimiento de las plantas, tendencias positivas en la
 334 precipitación podrían tener efectos sobre la cobertura vegetal y la coexistencia entre
 335 especies leñosas y herbáceas (Biancari et al., 2023).

336 En este estudio ponemos a disposición de los usuarios la base de datos PMRAv1. El
 337 término ‘v1’ (versión 1) hace referencia a que consideramos que este producto tendrá
 338 sucesivas versiones en el futuro que permitan actualizarla y mejorar la precisión de las
 339 estimaciones. Para mejorar las estimaciones identificamos tres posibles vías: i)
 340 incrementar la densidad de datos de pluviómetro: estudios previos han mostrado que la
 341 precisión de los modelos fusionados mediante RF tiende a incrementarse al incluir un
 342 mayor número de datos de terreno (Báez-Villanueva et al., 2020; Zhang et al., 2021); ii)
 343 modificar los PGP utilizados: las versiones futuras se pueden mejorar incluyendo
 344 solamente aquellos productos globales que tengan mejor ajuste a los datos observados y
 345 iii) regionalizar el país: por ejemplo, ajustando modelos por región de acuerdo a patrones
 346 de precipitación homogéneos. Luego, comparar si el producto consolidado a nivel
 347 nacional mejora la estimación respecto a un único producto nacional.

348 La base de datos generada, PMRAv1, mejora la estimación de los PGP para la República
 349 Argentina, y ha sido puesta a disposición de manera abierta en diversos formatos. Contar
 350 con productos climáticos validados y abiertos de alcance nacional puede traducirse en
 351 aplicaciones y modelos más certeros para diversas disciplinas que requieran datos de
 352 precipitación como insumo. En este trabajo se demostró una posible aplicación de este
 353 producto para estimar tendencias en las precipitaciones, lo cual tiene gran valor para
 354 caracterizar con mayor certeza el cambio climático. Finalmente, el método presentado en
 355 este estudio también se podría utilizar para mejorar la estimación espacio-temporal de
 356 otras variables climatológicas cuando se disponga de datos medidos en terreno.

357

358 **5) REFERENCIAS**

- 359 Abatzoglou, J.T., Dobrowski, S.Z., Parks, S.A. y Hegewisch, K.C., 2018: Terraclimate,
360 a high-resolution global dataset of monthly climate and climatic water balance from
361 1958-2015, *Scientific Data* 5:170191.
- 362 Adhikary, S. K., Yilmaz, A. G., y Muttill, N., 2015: Optimal design of rain gauge network
363 in the Middle Yarra River catchment, Australia. *Hydrological Processes*, 29(11),
364 2582-2599.
- 365 Almonacid, L., Pessacg, N., Diaz, B. G., Bonfili, O., y Peri, P. L., 2021: Nueva base de
366 datos reticulada de precipitación para la provincia de Santa Cruz, Argentina. *Centro*
367 *Argentino de Meteorólogos (CAM)*.
- 368 Andersen, J., Dybkjaer, G., Jensen, K. H., Refsgaard, J. C., y Rasmussen, K., 2002: Use
369 of remotely sensed precipitation and leaf area index in a distributed hydrological
370 model. *Journal of Hydrology*, 264(1-4), 34-50.
- 371 Ashouri, H., Hsu, K.L., Sorooshian, S., Braithwaite, D.K., Knapp, K.R., Cecil, L.D.,
372 Nelson, B.R. y Prat, O.P., 2015: Persiann-CDR daily precipitation climate data
373 record from multisatellite observations for hydrological and climate studies. *Bull.*
374 *Am. Meteorol. Soc.*, 96, 69–83.
- 375 Baez-Villanueva, O.M., Zambrano-Bigiarini, M., Beck, H. E., McNamara, I., Ribbe, L.,
376 Nauditt, A., ... y Thinh, N.X., 2020: RF-MEP: A novel Random Forest method for
377 merging gridded precipitation products and ground-based measurements. *Remote*
378 *Sensing of Environment*, 239, 111606.
- 379 Biancari, L., Oñatibia, G.R., Gaitán, J.J. y Aguiar, M.R., 2023: Coexistence of grasses
380 and shrubs in Patagonian steppes. Norm or exception? *Journal of Vegetation*
381 *Science*, 34, e13177.
- 382 Chao, L., Zhang, K., Li, Z., Zhu, Y., Wang, J., Yu, Z., 2018: Geographically Weighted
383 Regression Based Methods for Merging Satellite and Gauge Precipitation. *Journal*
384 *of Hydrology*. 558, 275–289.
- 385 Falaschi, D., Berthier, E., Belart, J.M.C., Bravo, C., Castro, M., Durand, M., Villalba, R.,
386 2023: Increased mass loss of glaciers in Volcán Domuyo (Argentinian Andes)
387 between 1962 and 2020, revealed by aerial photos and satellite stereo imagery.
388 *Journal of Glaciology* 69(273), 40–56.
- 389 Fick, S.E. y Hijmans, R.J., 2017: WorldClim 2: new 1-km spatial resolution climate
390 surfaces for global land areas. *Int. J. Climatol.* 37, 4302–4315.

- 391 Garcia, M., Peters-Lidard, C. D., y Goodrich, D. C., 2008: Spatial interpolation of
392 precipitation in a dense gauge network for monsoon storm events in the
393 southwestern United States. *Water Resources Research*, 44(5).
- 394 Harris, I., Jones, P.D., Osborn, T.J., y Lister, D.H., 2017: CRU TS4.00: Climatic Research
395 Unit (CRU) Time-Series (TS) version 4.00 of high-resolution gridded data of
396 month-by-month variation in climate (Jan. 1901- Dec. 2015). Centre for
397 Environmental Data Analysis. doi:10.5285/edf8febdaad48abb2cbaf7d7e846a86.
- 398 Huffman, G.J., Stocker, E.F., Bolvin, D.T., Nelkin, E.J., y Jackson T., 2019: GPM
399 IMERG final precipitation L3 half hourly 0.1 degree \times 0.1 degree V06, Greenbelt,
400 MD, Goddard earth sciences data and information services center (GES DISC).
- 401 Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., ... y Takahashi, K.,
402 2015: The JRA-55 reanalysis: General specifications and basic characteristics.
403 *Journal of the Meteorological Society of Japan. Ser. II*, 93(1), 5-48.
- 404 Legates, D. R., y McCabe, G. J., 1999: Evaluating the use of “goodness-of-fit” measures
405 in hydrologic and hydroclimatic model validation. *Water Resour.*
406 *Res.*, 35(1), 233–241.
- 407 Manz, B., Buytaert, W., Zulkafli, Z., Lavado, W., Willems, B., Robles, L.A.
408 y Rodríguez-Sánchez, J.P., 2016: High-resolution satellite-gauge merged
409 precipitation climatologies of the Tropical Andes. *J. Geophys. Res.*
410 *Atmos.*, 121, 1190–1207.
- 411 Mohammed, J., Mengiste, Y., y P. Singh, V., 2023: Improving spatio-temporal
412 precipitation estimates in data scarce river basins: an application of machine
413 learning-based multi-source data merging. *Stochastic Environmental Research and*
414 *Risk Assessment*, 37(4), 1353-1369.
- 415 Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo,
416 G., Bousssetta, S., Choulga, M., Harrigan, S., Hersbach, H., Martens, B., Miralles,
417 D. G., Piles, M., Rodríguez-Fernández, N. J., Zsoter, E., Buontempo, C., and
418 Thépaut, J.N., 2021: ERA5-Land: a state-of-the-art global reanalysis dataset for land
419 applications. *Earth Syst. Sci. Data*, 13, 4349–4383.
- 420 Nguyen, G. V., Le, X. H., Van, L. N., Jung, S., y Lee, G., 2023: Machine learning
421 approaches for reconstructing gridded precipitation based on multiple source
422 products. *Journal of Hydrology: Regional Studies*, 48, 101475.

- 423 Nguyen, G. V., Le, X. H., Van, L. N., Jung, S., Yeon, M., y Lee, G., 2021: Application
 424 of random forest algorithm for merging multiple satellite precipitation products
 425 across South Korea. *Remote Sensing*, 13(20), 4033.
- 426 Peretti, M., Spennemann, P.C. y Long, M.E.F., 2023: Trends in soil moisture content and
 427 water deficits in Argentina and the role of climate contribution. *Theor Appl*
 428 *Climatol* 152, 1189–1201.
- 429 Qi, J., Markewitz, D., y Radcliffe, D., 2018: Modelling the effect of changing
 430 precipitation inputs on deep soil water utilization. *Hydrological Processes*, 32(5),
 431 672-686.
- 432 Stisen, S., Højberg, A. L., Troldborg, L., Refsgaard, J. C., Christensen, B. S. B., Olsen,
 433 M., y Henriksen, H. J., 2012: On the importance of appropriate precipitation gauge
 434 catch correction for hydrological modelling at mid to high latitudes. *Hydrology and*
 435 *Earth System Sciences*, 16(11), 4157-4176.
- 436 Tovar, C., Carril, A.F., Gutiérrez, A.G., Ahrends, A., Fita, L., Zaninelli, P., ... y
 437 Hollingsworth, P.M., 2022: Understanding climate change impacts on biome and
 438 plant distributions in the Andes: Challenges and opportunities. *Journal of*
 439 *Biogeography*, 49(8), 1420-1442.
- 440 Verdin, A., Funk, C., Rajagopalan, B. y Kleiber, W., 2016: Kriging and Local Polynomial
 441 Methods for Blending Satellite-Derived and Gauge Precipitation Estimates to
 442 Support Hydrologic Early Warning Systems. *IEEE Transactions on Geoscience and*
 443 *Remote Sensing*, 54(5), 2552-256.
- 444 Villarini, G., y Krajewski, W. F., 2008: Empirically-based modeling of spatial sampling
 445 uncertainties associated with rainfall measurements by rain gauges. *Advances in*
 446 *Water Resources*, 31(7), 1015-1023.
- 447 Villarini, G. y Krajewski, W.F. Review of the Different Sources of Uncertainty in Single
 448 Polarization Radar-Based Estimates of Rainfall., 2009: *Surv. Geophys.* 31, 107–
 449 129.
- 450 Wang L, Jiao W, MacBean N, Rulli MC, Manzoni S, Vico G, y D’Odorico P., 2022:
 451 Dryland productivity under a changing climate. *Nature Climate Change* 12: 981–
 452 994.

- 453 Woldemeskel, F. M., Sivakumar, B., y Sharma, A., 2013: Merging gauge and satellite
454 rainfall with specification of associated uncertainty across Australia. *Journal of*
455 *Hydrology*, 499, 167-176.
- 456 Xie, P., y Xiong, A. Y., 2011: A conceptual model for constructing high-resolution gauge-
457 satellite merged precipitation analyses. *Journal of Geophysical Research:*
458 *Atmospheres*, 116(D21).
- 459 Zhang, L., Li, X., Zheng, D., Zhang, K., Ma, Q., Zhao, Y., y Ge, Y., 2021: Merging
460 multiple satellite-based precipitation products and gauge observations using a novel
461 double machine learning approach. *Journal of Hydrology*, 594, 125969.
- 462 Zhang, L., Ren, D., Nan, Z., Wang, W., Zhao, Y., Zhao, Y., Ma, Q., y Wu, X., 2020:
463 Interpolated or satellite-based precipitation? Implications for hydrological modeling
464 in a meso-scale mountainous watershed on the Qinghai-Tibet Plateau. *Journal of*
465 *Hydrology* 583, 124629.
- 466 Zuur, A.F., Ieno, E.N., Elphick, C.S., 2010: A protocol for data exploration to avoid
467 common statistical problems. *Methods in Ecology and Evolution* 1: 3–14.
468

469 Figuras y tablas

470

471

472

473

474

475

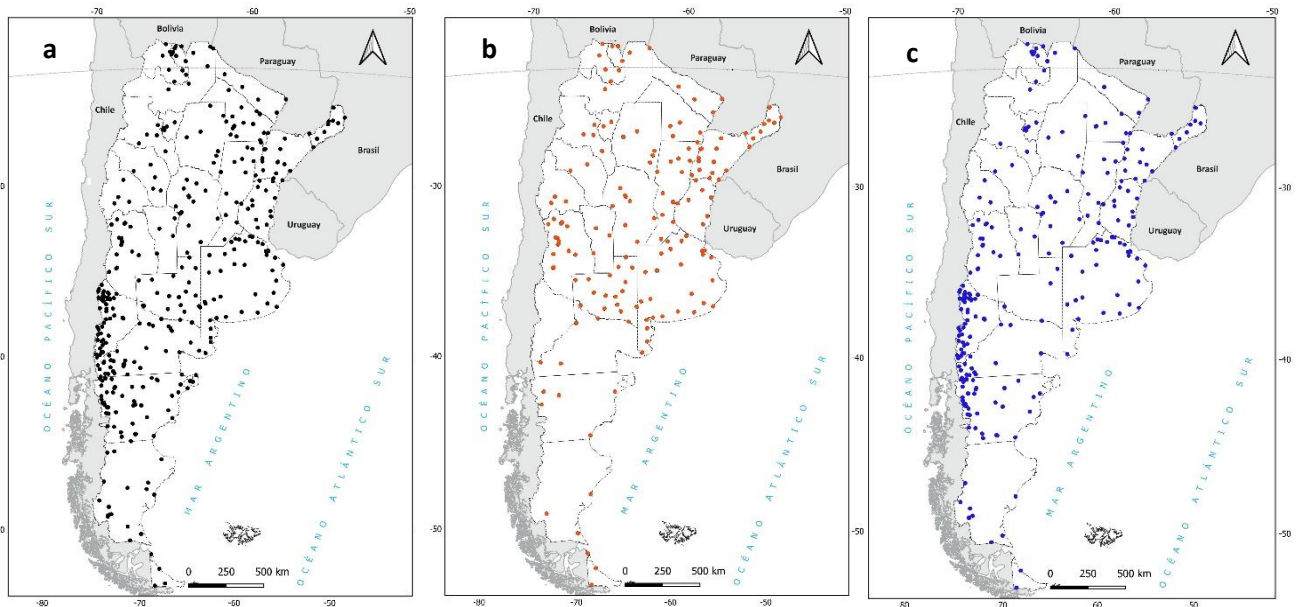
476

477

478

479

480



481

482 **Figura 1.** Total de estaciones meteorológicas utilizadas para modelar PMRAv1 (n = 385)

483 (a), y meses con menor (diciembre 2022, n = 163) (b) y mayor (diciembre 2011, n = 243)

484 (c) cantidad de estaciones disponibles.

485

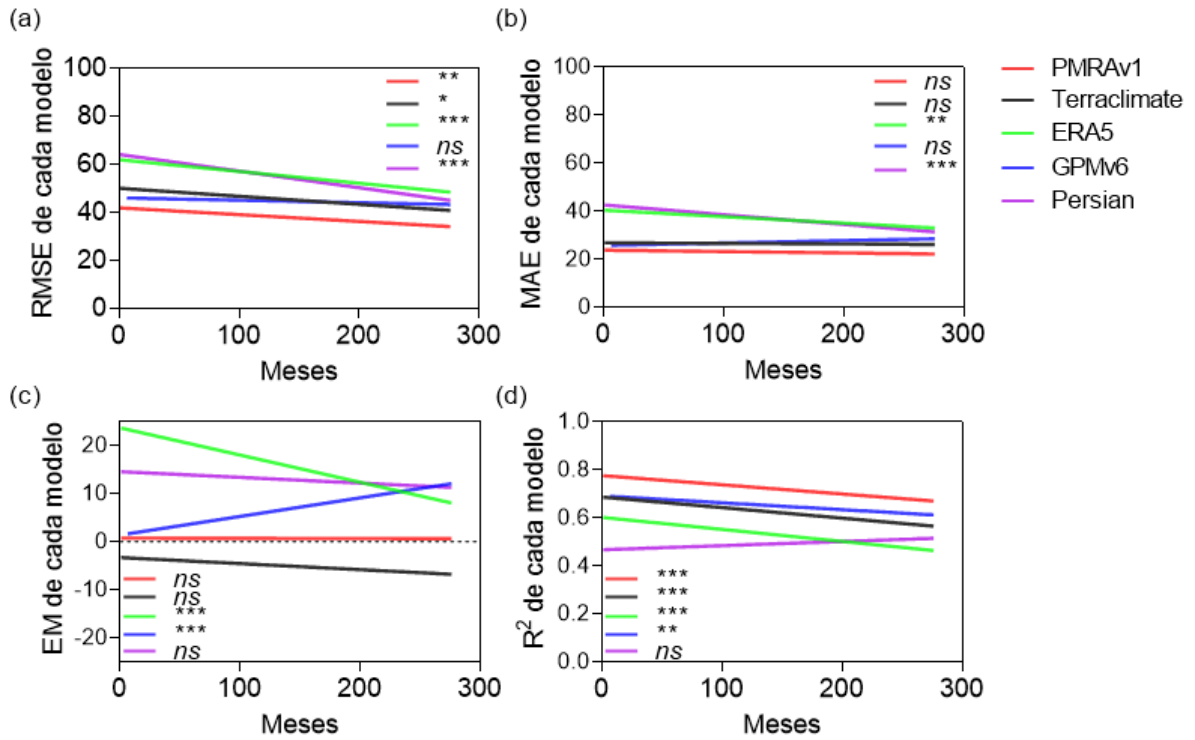
486 **Tabla I.** Resumen de los parámetros de ajuste RMSE (raíz del error cuadrático medio),
 487 MAE (error absoluto medio), EM (error medio) y R^2 para PMRAv1 y los cuatro productos
 488 globales de precipitación (TERRACLIMATE, ERA5-LAND, GPMv6 y PERSIANN-
 489 CDR). Se muestran medias, desvíos estándar,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

491

	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
R²	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

492

493



494

495 **Figura 2.** Ajuste de cada producto de precipitación a lo largo de los 276 meses estudiados.

496 Cada panel representa los distintos parámetros de ajuste estudiados: (a) RMSE, (b) MAE,

497 (c) EM y, (d) R². En cada panel se detalla el nivel de significancia de la prueba de hipótesis

498 para pendientes distintas de 0: * = p<0.05, ** = p<0.01, *** = p<0.001 y ns = no

499 significativo.

500

501

502

503

504

505

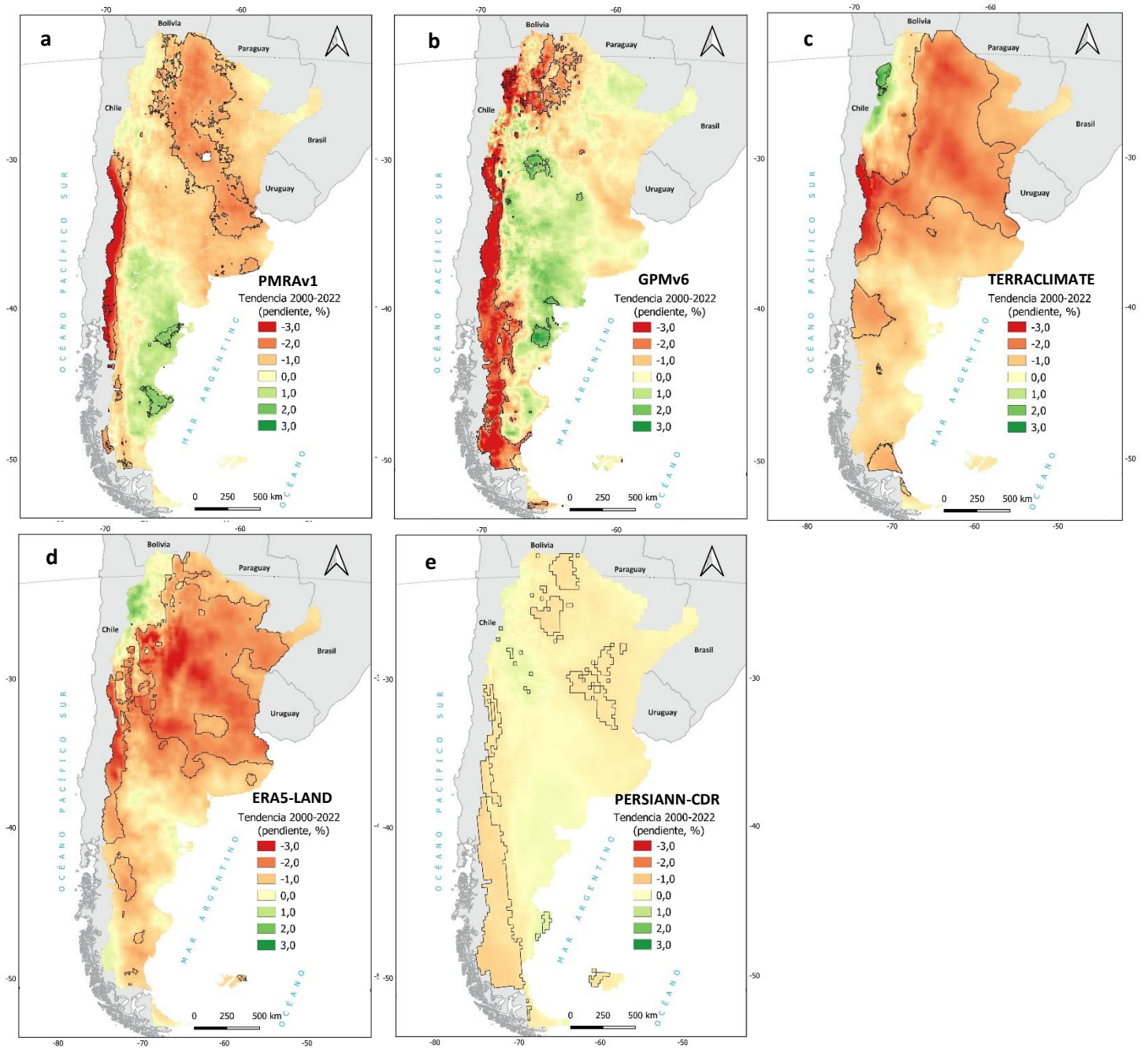
506

507

508

509

510



529

530 **Figura 3.** Pendiente de la tendencia de la precipitación durante el período 2000-2022,
531 expresada como porcentaje de la media del período, para PMRAv1 (a) y cuatro productos
532 globales de precipitación: GPMv6 (b), TERRACLIMATE (c), ERA5-LAND (d) y

- 533 PERSIANN-CDR (e). Las líneas sobre el mapa indican las áreas con tendencia
534 estadísticamente significativa ($p < 0.05$).