



PERÚ

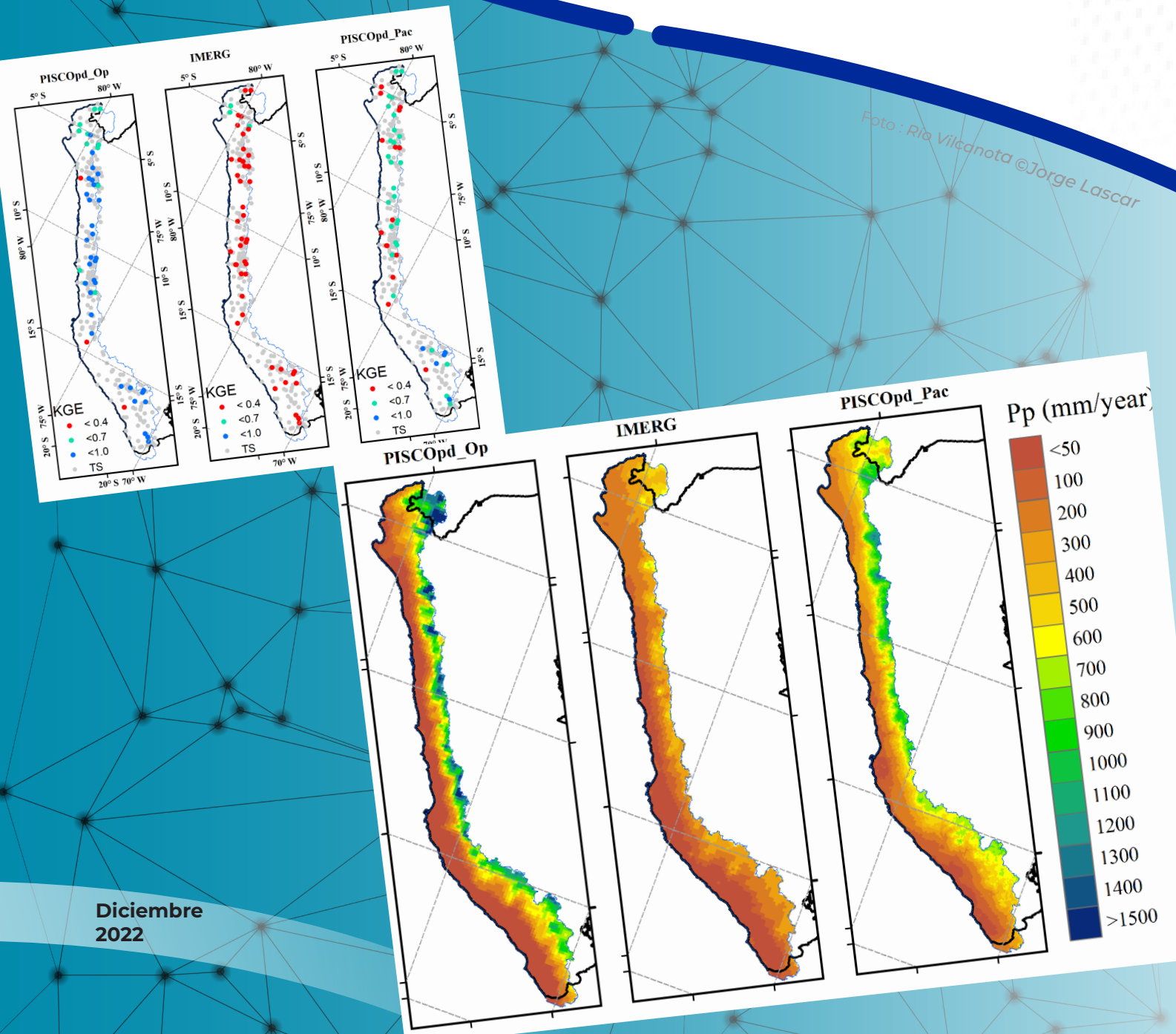
Ministerio del Ambiente



# Conjunto de datos de precipitación grillada diaria de alta resolución para la vertiente del Pacífico en Perú mediante una interpolación espacial Random Forest

Dirección de Hidrología - DHI  
Subdirección de Estudios e Investigaciones Hidrológicas

Foto : Río Vilcanota ©Jorge Lascar



Diciembre 2022

**Citar:**

**Millán, C. (2023).** Conjunto de datos de precipitación grillada diaria de alta resolución para la vertiente del Pacífico en Perú mediante una Interpolación Espacial Random Forest. Estudio Final. Servicio Nacional de Meteorología e Hidrología del Perú –SENAMHI, (2022).

# Conjunto de datos de precipitación grillada diaria de alta resolución para la vertiente del Pacífico en Perú mediante una interpolación espacial Random Forest

ESTUDIO FINAL





## **DIRECTORIO**

PhD. Guillermo Antonio Baigorria Paz  
PRESIDENTE EJECUTIVO DEL SENAMHI

Ing. Oscar Gustavo Felipe Obando  
DIRECTOR DE HIDROLOGÍA DEL SENAMHI

PhD. Waldo Sven Lavado Casimiro  
SUBDIRECTOR DE ESTUDIOS E INVESTIGACIONES HIDROLÓGICAS DEL SENAMHI

### **ReESPONSABLE DEL ESTUDIO**

Ing. Carlos Millán Arancibia  
Especialista en Hidrología  
Subdirección de Estudios e Investigaciones Hidrológicas (SEH)

### **REVISOR DEL ESTUDIO**

PhD. Waldo Lavado Casimiro  
Subdirector de Estudios e Investigaciones Hidrológicas del SENAMHI

*Diciembre 2022*

*Lima - Perú*

## Contenido

1.	Antecedentes & resumen.....	7
2.	Métodos y conjuntos de datos.....	9
2.1	Área de estudio .....	9
2.2	Fuente de datos observados en tierra.....	9
2.2.1	Estaciones pluviométricas convencionales. ....	9
2.2.2	Estaciones pluviométricas automáticas. ....	9
2.3	Fuentes de datos grillados. ....	11
2.4	Metodología.....	12
3.	Validación técnica .....	12
3.1	Validación del modelo.....	13
3.2	Validación independiente con estaciones pluviométricas automáticas	14
4.	Comparación a nivel de series temporales.....	16
5.	Notas de uso .....	19
6.	Referencias.....	20
7.	Apéndice:.....	23

## Contenido de figuras:

<b>Figura 01.</b> El área de estudio, la vertiente del Pacífico peruano, representado según la lluvia media anual (izq.) y su variación altitudinal (der.).....	8
<b>Figura 02.</b> Resumen del esquema metodológico.....	12
<b>Figura 03.</b> Correlación de Pearson entre los rendimientos KGE de los productos de lluvia. ....	13
<b>Figura 04.</b> Rendimiento KGE ´ de los pluviómetros convencionales separadas para la validación inicial del modelo. TS = Terrain Stations.....	14
<b>Figura 05.</b> a) Rendimiento KGE ´ (superior) y b) Sesgo (inferior) para los pluviómetros automáticos.....	15
<b>Figura 06.</b> Distribución espacial del promedio interanual de Precipitación (Pp) para el periodo de referencia 2014-2020 para PISCOpd_Op (left) IMERG (mid) y PISCOpd_Pac (right). 16	16
<b>Figura 07.</b> Serie histórica mensual de Precipitación (Pp) para el periodo de referencia 2014-2020 para PISCOpd_Op (rojo) IMERG (verde) y PISCOpd_Pac (azul).....	17
<b>Figura 08.</b> Climatología mensual de Precipitación (Pp) para el periodo de 2014-2020 para PISCOpd_Op (rojo) IMERG (verde) y PISCOpd_Pac (azul).....	18
<b>Figura 09.</b> Serie histórica anual de Precipitación (Pp) para el periodo de referencia 2014-2020 para PISCOpd_Op (rojo) IMERG (verde) y PISCOpd_Pac (azul).....	19

## Contenido de tablas:

<b>Tabla 01.</b> Conjuntos de datos meteorológicos de precipitación grillada diarios existentes para Perú (RS significa teledetección).....	7
<b>Tabla 02.</b> Covariables espacio-temporales consideradas.....	11
<b>Tabla 03.</b> Recopilación y sistematización de investigaciones para la especialización de la precipitación mediante técnicas de aprendizaje automático.....	10

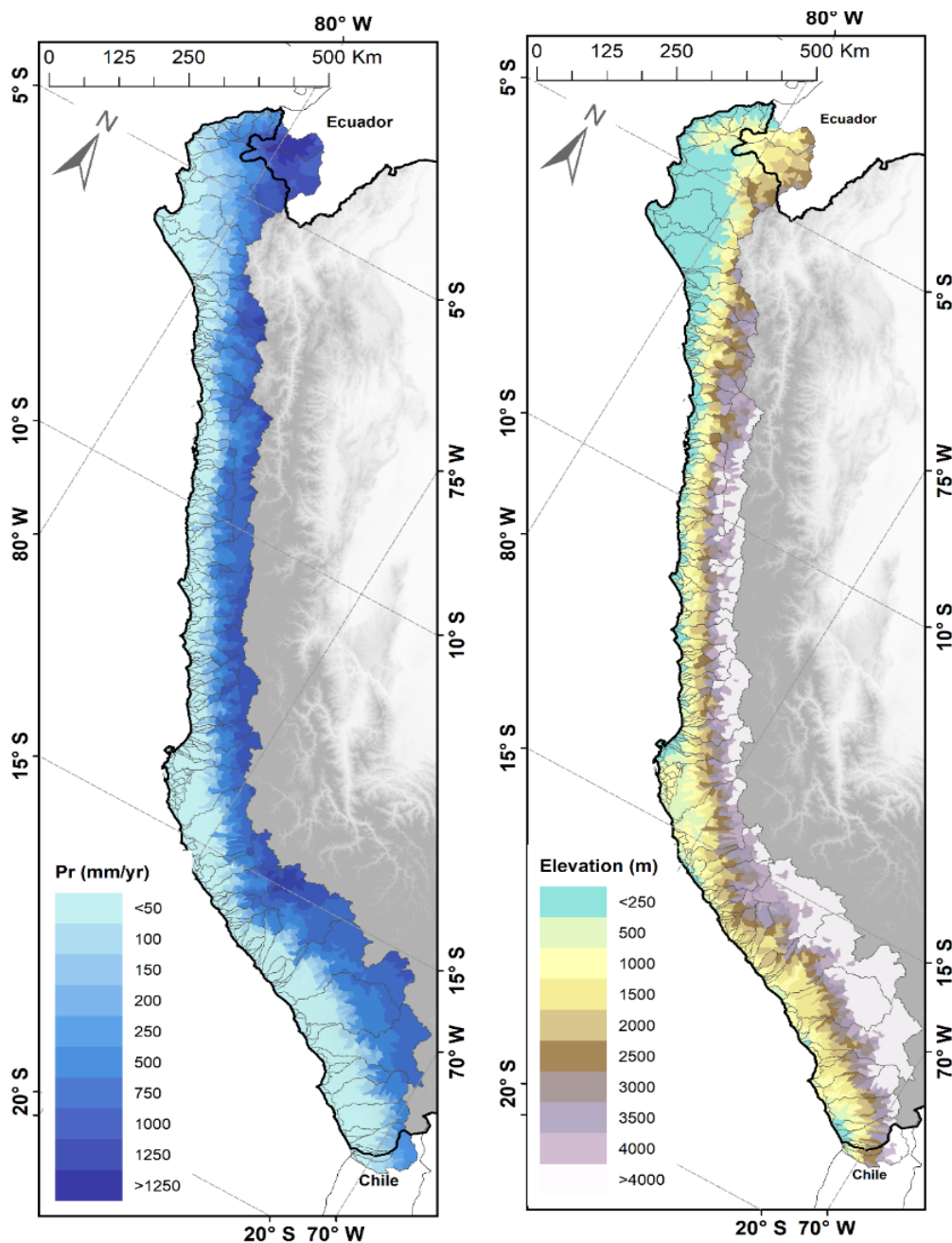
## Conjunto de datos de precipitación grillada diaria de alta resolución para la vertiente del Pacífico en Perú mediante una interpolación espacial Random Forest

### 1. Antecedentes & resumen

Existen diferentes conjuntos de datos de precipitación grillada para la región de la vertiente del Pacífico en la parte occidental del Perú, los cuales tienen diferentes escalas espaciales y temporales, en la Tabla 1 se muestra a detalle algunos conjuntos de datos disponibles para la zona de interés. A pesar de ello, la principal limitación de estos productos es la resolución espacial gruesa, la cual genera un nivel de incertidumbre y representatividad para representar los patrones meteorológicos localizado a nivel de quebradas y pequeñas cuencas a lo largo de la vertiente del Pacífico cuyas áreas son menores al tamaño de grilla de estos productos. Los datos de precipitación son insumos de entrada esenciales para numerosos modelos y análisis a escala local en diferentes campos de investigación, por ejemplo, en análisis de la oferta de agua para la agricultura, determinar patrones meteorológicos de las lluvias (Haslinger et al. 2014; Huerta and Lavado-Casimiro 2021), hidrología (Lee, Im, and Bae 2019; Llauca et al. 2021), clima y cambio climático (Sippel et al. 2020), riesgos hidrometeorológicos (Petritsch and Hasenauer 2014) y sistemas de alerta temprana (Millan et al. 2021).

**Tabla 01.** Conjuntos de datos meteorológicos de precipitación grillada diarios existentes para Perú (RS significa teledetección).

Nombre del conjunto de datos	Abreviatura	Ref.	Tipo	Resolución espacial
Misión de medición de precipitaciones tropicales/recuperaciones integradas de múltiples satélites para la medición de precipitaciones globales	TRMM/IMERG	(Huffman, G. J., Bolvin, D. T. & Nelkin 2014)	Basado en RS	0.1° (~10 kilómetros)
Estimación de precipitación a partir de información de teledetección utilizando redes neuronales artificiales	PERSIANN	(Nguyen et al. 2019)	Basado en RS	0.04° (~4 kilómetros)
Centro de Predicción del Clima global de temperatura y precipitación	CPC	(PSL n.d.)	Basado en estaciones	0.5° (~50 kilómetros)
Conjuntos de datos observacional grillado diario ensamblados	E-OBS	(Comes et al. 2018)	Basado en estaciones	0.1° (~10 kilómetros)
Peruvian Interpolated data of SENAMHI's Climatological and Hydrological Observations	PISCO	(Aybar et al. 2020)	Basado en estaciones y RS	0.1° (~10 kilómetros)
Administración Nacional Oceánica y Atmosférica (NOAA) - CIRES 20th Century Reanalysis	NOAA-CIRES	(Compo et al. 2011)	Reanálisis	2.5° (~250 kilómetros)
ERA-Interino	ERA-Interino	(Dee et al. 2011)	Reanálisis	80 kilómetros
ERA5 (cada hora, pero se puede agregar a una resolución diaria)	ERA5	(Muñoz Sabater 2019)	Reanálisis	0.25° (~25 kilómetros)
A novel high-resolution gridded precipitation dataset for Peruvian and Ecuadorian watersheds – development and hydrological evaluation	RAIN4PE	(Fernández et al., 2021)	Basado en estaciones y RS	0.1° (~10 kilómetros)



**Figura 01.** El área de estudio, la vertiente del Pacífico peruano, representado según la lluvia media anual (izq.) y su variación altitudinal (der.).

Perú cuenta con el conjunto de datos PISCOp (Aybar et al. 2020), cuya base de datos actualmente abarca el periodo entre 1981 y 2019, a una resolución espacial de 0.1°, y cuenta con un producto operativo PISCO\_Op. A pesar de que su nivel de certidumbre es mayor que la de otros productos globales debido a que fue calibrada y generada a partir del conjunto de datos de estaciones terrestres de lluvia del SENAMHI, aun la resolución sigue siendo gruesa para las quebradas y pequeñas cuencas en todo el Perú.



Durante los últimos años, Perú ha tomado un especial interés en la gestión de riesgos de desastres asociados a los peligros hidrometeorológicos, razón por la que es necesario mejorar la representatividad de los datos operativos de precipitación continuamente. Además, la lluvia es el principal desencadenante de peligros de inundaciones repentinas, crecidas de detritos, flujos de detritos o escombros, flujos de lodos y lahares, generalmente conocidos como huaycos en el Perú.

Con esto en mente, se desarrolló el conjunto de datos **PISCOpd\_PacRF**, un conjunto de datos de precipitación grillada a una resolución espacial de 5 km en toda la costa del Perú, para la vertiente del Pacífico, para el periodo de 2014-2020. Para generar este conjunto de datos se utilizó un Procedimiento de Combinación basado en bosques aleatorios o *Random Forest Merging Procedure* (RF-MEP) (Baez-Villanueva et al. 2020). Se seleccionó RF-MEP pues combina covariables medioambientales y observaciones de estaciones cercanas, en orden para predecir valores en lugares sin observaciones, esto es de especial interés en el monitoreo de peligros hidrometeorológicos en zonas con escasos o insuficientes datos.

## 2. Métodos y conjuntos de datos

### 2.1 Área de estudio

El área de estudio comprende la vertiente del Pacífico en la región costera del Perú (Figura 01), la cual se encuentra en el franco occidental del país. Cubre un área de ~297594 km<sup>2</sup> y se extiende de este (-69.35° W) a oeste (-81.33° W) con un alto gradiente altitudinal, esta variabilidad orográfica está definida por los Andes, que comienza por encima de los 5000 msnm hasta desembocar en el Océano Pacífico. Incluye las cuencas transfronterizas con Ecuador en el norte (-3.40° N) y Chile en el sur (-18.51° S).

El sistema hidroclimático peruano está influenciado por la Cordillera de los Andes, contrastando las condiciones de contorno oceánico y la distribución de la masa terrestre (Garreaud et al. 2009) que describen gran parte de su variabilidad estacional e interanual de las precipitaciones. Esta región presenta mayores variaciones de precipitación que las otras dos principales regiones hidrológicas del Perú: el Amazonas y los drenajes endorreicos del Titicaca (Lavado Casimiro et al. 2012).

### 2.2 Fuente de datos observados en tierra

**2.2.1 Estaciones pluviométricas convencionales.** El conjunto de datos de precipitación diaria observada proviene de las estaciones meteorológicas convencionales administradas por el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI) las cuales se utilizaron para generar la interpolación espacial de la lluvia. Se utilizó esta base de datos, pues es la base de datos de precipitación operativa diaria que se actualiza en los servidores de SENAMHI y se recopila entre las 7:00 am hasta las 10:00 am. Esta misma base de datos se proyecta como insumo para actualizar PISCOpd\_PacRF.

**2.2.2 Estaciones pluviométricas automáticas.** Adicionalmente se incluyeron estaciones pluviométricas automáticas, administradas por SENAMHI, las cuales se usarán exclusivamente para el proceso de validación. En total se incluyeron 87 estaciones automáticas exclusivamente para la validación del producto.

Tabla 02. Recopilación y sistematización de investigaciones para la especialización de la precipitación mediante técnicas de aprendizaje automático

Estudio	Método(s) de fusión	Covariables	Región	Producto(s)	Número de estaciones técnicas de validación)	Resultados principales
(Sekulić et al. 2021)	Random Forest Spatial Interpolation methodology (RFSI)	Estaciones de OGIMET, Estaciones automáticas, DEM y W1, IMERG V06B, ensamble E-OBS	Serbia		61 (5-fold leave-location-out cross-validation)	El primer conjunto de datos diarios meteorológicos grilados de alta resolución (0.01°) para Serbia
(Turini et al. 2021)	Random forest (RF)	IMERG-V06 y GOES-16	Ecuador	Recuperación de llluvias por RF	22 (datos independientes para la validación)	Un nuevo algoritmo basado en satélite para la recuperación de llluvias con alta resolución espacio-temporal para Ecuador.
(Baez-Villanueva et al. 2020) the estimation of precipitation patterns derived solely from rain gauges is subject to large uncertainties. We present the Random Forest based Merging Procedure (RF-MEP	Random Forest based Merging Procedure (RF-MEP)	Estaciones, PERSIANN-CDR, ERA-Interim, CHIRPSv2, CMORPHv1 y TRMM 3V42v7. Slope, aspect, Köppen-Geiger climate classification, land cover type	Chile	RFMEP <sub>sp</sub> y RF-MEP <sub>sp</sub>	369 (258/111)	El desarrollo de un método de fusión basado en Random Forest, el cual generó dos productos con valores de KGE superiores al producto de referencia (MSWEPv2.2) demostrando que la combinación de productos de precipitación y mediciones en tierra generan una mejor representación de la variabilidad espacio temporal de la precipitación.
(Sekulić et al. 2020)	Random Forest based Merging Procedure (RF-MEP)	Datos sintéticos, estaciones diarias de GHCN, IMERG-V06A, Tmax, Tmin, DEM	Cataluña (España)	Mapas interpolados	87 (5-fold leave-location-out cross-validation)	La introducción de un método novedoso de interpolación espacial que puede producir resultados interpolados espaciales precisos.
(Kumar et al. 2019)	Support vector machine-based regression (SVR) model	Datos grilados diarias de IMD, TMPA NRT SRE 3B42RT v7, humedad de suelo de ASCAT (H101 y H16)	Ashti catchment (India)	Producto corregido de precipitación grilada	Datos grilados de lluvia IMD (optimización de parámetros del modelo)	El desarrollo de un método de integración de datos de satélite de llluvias y humedad de suelo mediante técnicas de aprendizaje automático mediante el modelo de support vector machine, logrando reducir el sesgo y RMSE.
(Hengl et al. 2018) but spatial location of points (geography)	Random forest for spatial predictions framework (RF-sp)	Distancias de amortiguación euclidianas, CHELSA, DEM	Suiza	Producto de lluvia mejorado espacialmente	467 (5-fold leave-location-out cross-validation)	Un método de predicción espacial basado en random forest usando como principal covariable las distancias más cercanas como predictoras de la información por efecto de la proximidad espacial

## 2.3 Fuentes de datos grillados.

**DEM.** Un DEM con resolución espacial de 30 arco-segundos (~90m) del proyecto HydroSHED (Lehner and Grill 2013) basado primariamente en los datos de elevación del SRTM. A partir de este producto, se generó el aspecto y la pendiente, como variables descriptoras de la fisiografía y relieve accidentado de la vertiente del Pacífico. Se incluyó la latitud para representar la variación de las lluvias a lo largo de la vertiente del Pacífico de norte a sur asociadas a los patrones gobernantes de la variabilidad climática en Perú y Sudamérica; y la longitud para representar la prominente orografía por la cordillera de los Andes.

**IMERG.** IMERG (Huffman, G. J., Bolvin, D. T. & Nelkin 2014) es un algoritmo que combina información de múltiples fuentes, como estimaciones de precipitación por microondas por satélite, estimaciones de satélites infrarrojos calibrados por microondas, medidores de precipitación y otros estimadores de precipitación para estimar la precipitación sobre la mayor parte de la superficie de la Tierra. Uno de los productos de IMERG son los mapas (grillados) de estimaciones de precipitación diaria. Las estimaciones de precipitación V06B de la versión de ejecución final de IMERG se utilizaron para desarrollar PISCO\_Op\_Pac. Las estimaciones de IMERG son una covariable espacio-temporal con una resolución espacial de 10km y una resolución temporal de un día. La versión usada del conjunto de datos IMERG, basadas en GPM, cubre el período de 2014-2020. El conjunto de datos IMERG se utilizó como una covariable de escala más gruesa para la precipitación. Este dataset fue remuestreado a una resolución espacial de 5km.

**Precipitación media mensual.** El SENAMHI generó un conjunto de datos de mapas climatológicos para Perú, los cuales cuentan con precipitaciones medias mensuales y se tomó en cuenta estas covariables espaciales mensuales para representar la variabilidad climática de la precipitación en la vertiente del Pacífico peruano.

La tabla 03 resume las variables usadas en la generación de PISCOpd\_PacRF.

**Tabla 03.** Covariables espacio-temporales consideradas.

Nombre de la covariable	Variable	Tipo de covariable	Resolución espacial	Resolución temporal
DEM	Altitud (m)	Ambiental	90 m	-
Pendiente	Inclinación (%)	Ambiental	90 m	-
Aspecto	Orientación (°)	Ambiental	90 m	-
Longitud y Latitud	Ubicación (°)	Espacial	90 m	-
IMERG	Pr (mm)	Ambiental multitemporal	0.1° (~10 kilómetros)	Diaria
Precipitación media mensual	Pr (mm/month)	Ambiental multitemporal	0.1° (~10 kilómetros)	Mensual

## 2.4 Metodología

El método usado se basa en RF-MEP (Baez-Villanueva et al. 2020) por sus siglas en inglés *Random Forest based MERging Procedure*. RF-MEP se basa en tres supuestos clave: i) las mediciones de Precipitación de las estaciones de pluviómetros son precisas en la escala de puntos; ii) los productos de Precipitación son generalmente sesgados pero contienen información útil sobre los patrones espacio-temporales de Precipitación; y iii) la combinación de diferentes productos Precipitación y datos de pluviómetros puede proporcionar una mejor representación de la variabilidad espacio-temporal de Precipitación que cualquier producto individual. RF-MEP utiliza RF para predecir la distribución espacial de la Precipitación mediante la fusión de información de diferentes productos cuadrículados (conocidos como covariables) e información terrestre de calidad controlada a una escala temporal seleccionada (por ejemplo, diaria, mensual o anual). Las predicciones individuales se generan a partir de un número definido por el usuario de árboles de decisión basados en muestras de bootstrap utilizando las covariables como predictores. La predicción final se calcula como el promedio de las predicciones individuales. La Figura 02 resume los pasos del método utilizado.

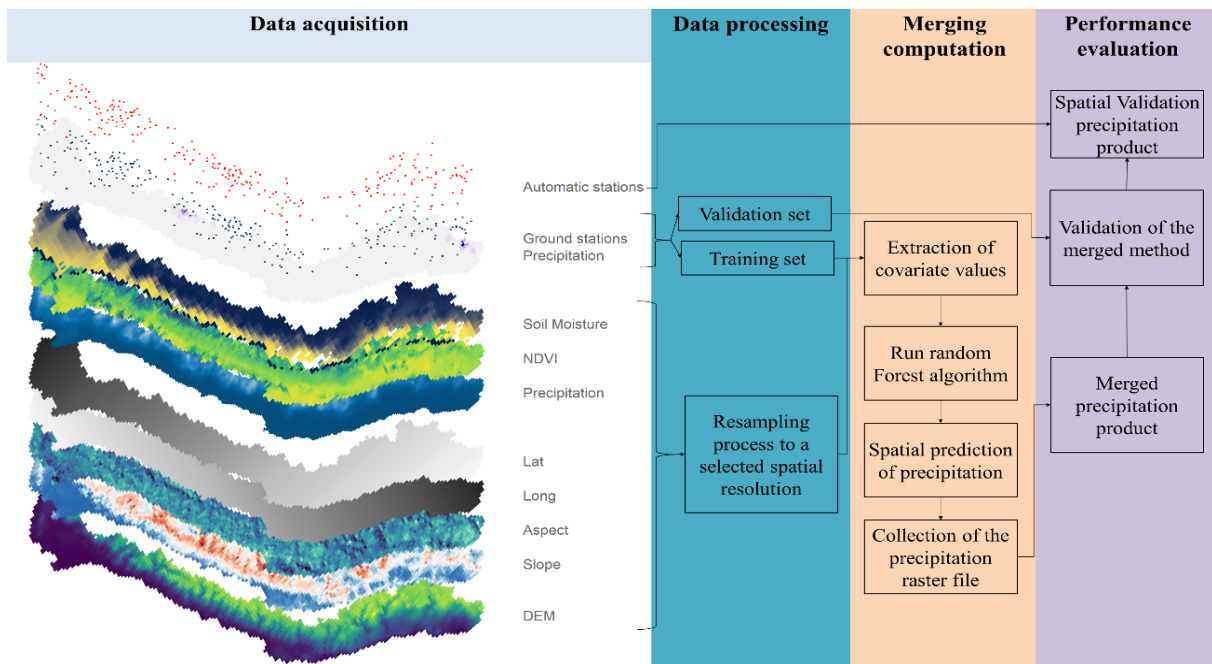


Figura 02. Resumen del esquema metodológico.

## 3. Validación técnica

La validación de PISCOpd\_PacRF fue evaluada en tres fases: i) validación del modelo utilizando la eficiencia Kling-Gupta modificada (KGE') incluida en el paquete hidroGOF (Zambrano-Bigiarini 2017). El valor óptimo para el KGE es uno; ii) validación independiente con estaciones automáticas; y iii) comparación con productos existentes para el área de análisis a diferentes escalas temporales. Los productos usados para la comparación fueron la versión operativo de PISCO (Aybar et al. 2020) y el producto IMERG-Early (Huffman, G. J., Bolvin, D. T. & Nelkin 2014). Se utilizó IMERG-Early pues es el producto con la menor latencia de los productos de GPM y es adecuado para sistemas de alerta temprana.

### 3.1 Validación del modelo

La validación interna del modelo se realizó de forma independiente al separar el 25 % de los pluviómetros convencionales utilizados para la generación del modelo en la etapa de entrenamiento. El total de pluviómetros convencionales usadas en la generación del modelo fueron 245 estaciones para la región del Pacífico peruano, de las cuales 196 (75 %) fueron usadas para el entrenamiento del modelo RF-MEP y 49 (25 %) fueron usadas para validar el modelo. Se realizó la validación mediante un análisis point-to-pixel, en el que las mediciones de las estaciones de pluviómetros se comparan con los valores de celdas de cuadrícula correspondientes de los productos de precipitación grillados bajo el supuesto de que las mediciones de pluviómetros son valores representativos en sus celdas de cuadrícula respectivas, este análisis es ampliamente utilizado para evaluar el rendimiento de los productos de precipitación (Baez-Villanueva et al. 2018, 2020; Dinku et al. 2007; Gao and Liu 2013; Li, Yang, and Hong 2013; Thiemig et al. 2012; Zambrano-Bigiarini et al. 2017).

La Figura 04 muestra el KGE' para las estaciones separadas para esta validación. Se observa que el producto PISCOpd\_Op tiene buen desempeño en la mayoría de sus estaciones, sobre todo en la zona de los andes, sin embargo, las estaciones que tuvieron menos desempeño se encuentran a niveles altitudinales bajos, esto debido a que estas estaciones registran tasas bajas de precipitación por lo que son más sensibles a la incertidumbre de cada producto. El producto IMERG nativo no tiene ninguna estación que registre buenos desempeños ( $KGE' > 0.7$ ), por lo que no se recomienda su uso de forma directa sin algún tipo de corrección en su aplicación. PISCOpd\_PacRF registra mayor cantidad de estaciones con rendimientos aceptables ( $0.4 < KGE' < 0.7$ ), sin embargo, se observa que tiene desempeños menores que los de PISCOpd\_Op, esto es congruente y esperado pues las estaciones usadas para generar PISCOpd\_PacRF fueron usadas en la generación del producto PISCOpd\_Op. Esta última condición fue la que nos llevó a realizar una validación con datos independientes y con diferente ubicación espacial que las estaciones, los cuales son descritos en la fase dos de la validación.

La Figura 03 muestra la correlación entre el rendimiento KGE de las estaciones versus su ubicación (*lon* y *lat*) y altitud (*z*). PISCOpd\_Op tiene una correlación alta significativa ( $p=0.05$ ) de 0.62 con la altitud de las estaciones, esto quiere decir que mientras a menor altura se encuentre la estación, menor será su rendimiento. Mientras que PISCOpd\_Pac tiene una correlación moderada de 0.42.

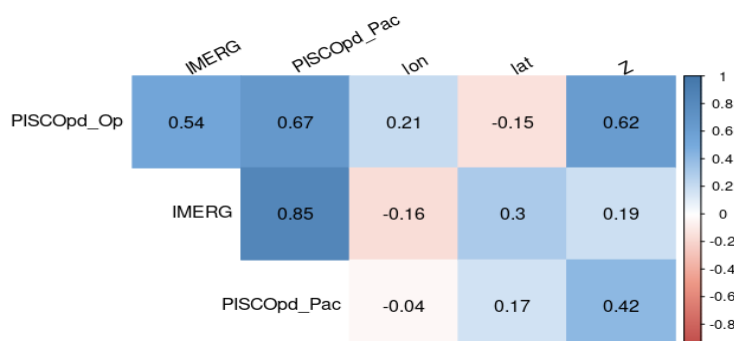
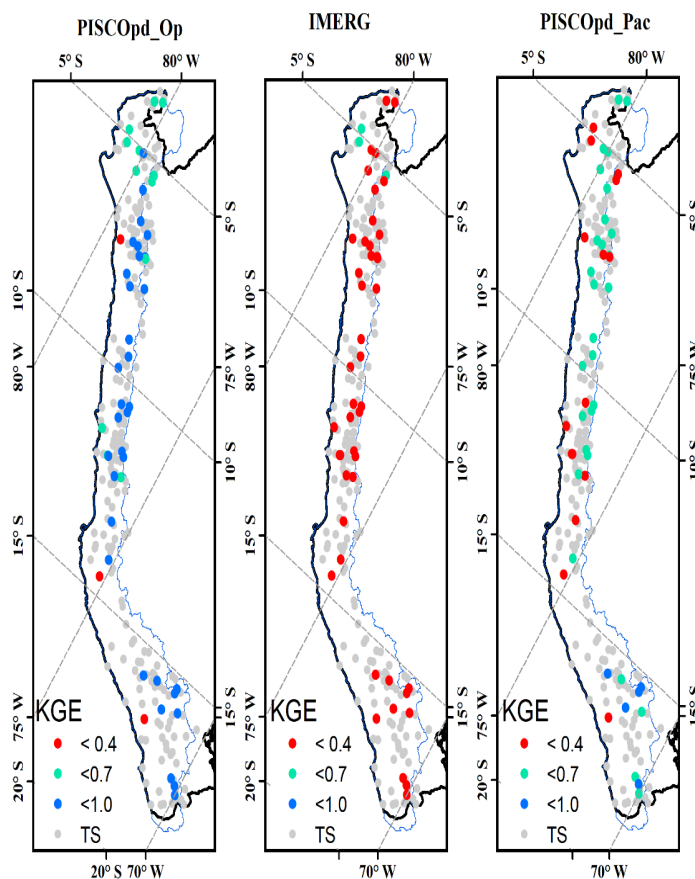


Figura 03. Correlación de Pearson entre los rendimientos KGE de los productos de lluvia.





**Figura 04.** Rendimiento KGE' de los pluviómetros convencionales separadas para la validación inicial del modelo. TS = Terrain Stations.

### 3.2 Validación independiente con estaciones pluviométricas automáticas

La validación independiente se realizó con estaciones automáticas del servicio hidrometeorológico de Perú, SENAMHI, de las cuales se filtraron y se consideraron 87 estaciones pluviométricas automáticas. Para la evaluación del rendimiento se usaron 2 métricas para caracterizar el error y/o desempeño i) índice de eficiencia Kling-Gupta modificada KGE y ii) el error simple (Bias).

Los rendimientos de eficiencia de KGE' evidencian un mejor rendimiento de la precipitación para PISCOpd\_PacRF sobre PISCOpd\_Op e IMERG-Early nativo (Figura 05 a). La Fig. 05b muestra el sesgo en las estaciones de los tres productos comparados, PISCOpd\_Pac disminuye de forma general el sesgo (más estaciones con coloración verde claro) mejorando la estimación de las lluvias, sin embargo, se observa una variabilidad del sesgo a lo largo de las costas del Pacífico (norte, central y sur). En el Pacífico norte, se observa que hay una mejora de la subestimación de la lluvia (con estaciones de coloración rojiza y naranja en PISCOpd\_Op e IMERG, que pasan a color verde en PISCOpd\_PacRF), sin embargo, se observa que algunos pluviómetros sobrestiman las lluvias (coloración verde oscura y azul). La zona del Pacífico central si tiene una clara disminución del sesgo en las estaciones para PISCOpd\_PacRF. Mientras que en el Pacífico sur se observa mayor variabilidad del sesgo con estaciones que subestiman y otras que tienen una sobrestimación, sin embargo, aún hay predominancia de pluviómetros con coloración verde claro en comparación de sus pares.

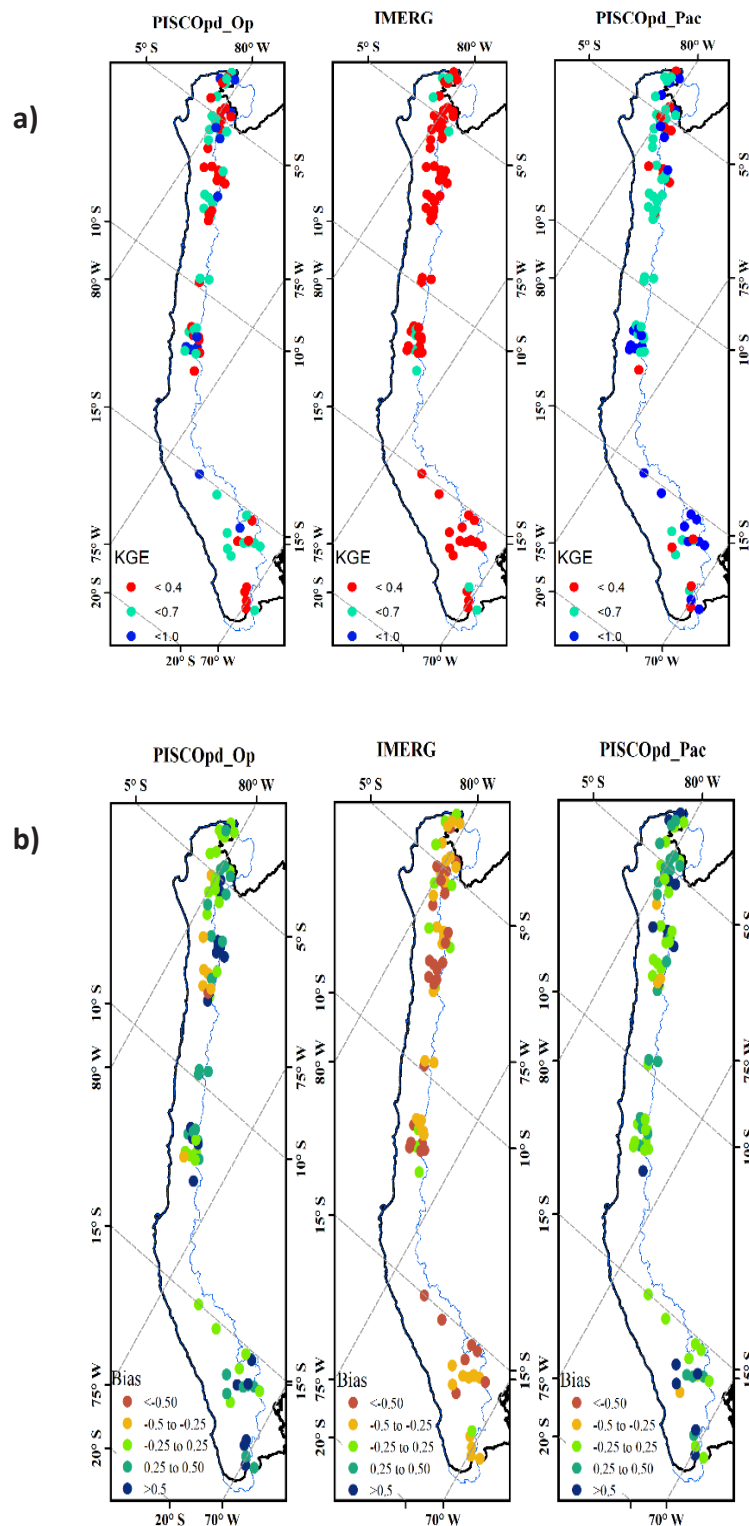
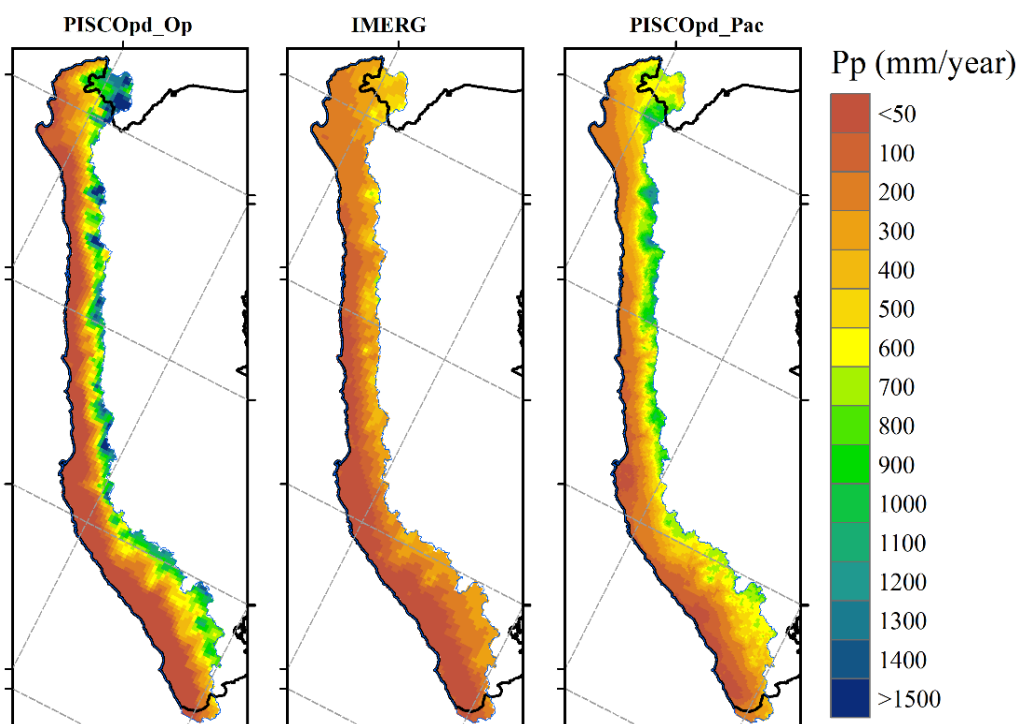


Figura 05. a) Rendimiento KGE' (superior) y b) Sesgo (inferior) para los pluviómetros automáticos.

#### 4. Comparación a nivel de series temporales

La comparación de los productos analizados se realizó a cuatro niveles: i) comparación de la variabilidad espacial del promedio interanual de la lluvia, ii) variabilidad mensual regional, iii) variabilidad del promedio mensual regional, y iv) variabilidad interanual regional. Acerca de las regiones, se usó las regiones hidrológicas de la Autoridad Nacional del Agua (ANA), que discretiza el Pacífico en 5 regiones administrativas de los recursos hídricos.

En primer lugar, se comparó la variabilidad espacial interanual para los 3 productos. La Figura 06 muestra la distribución espacial de la Pp media anual (2014-2020) para PISCOpd\_Op, IMERG y PISCOpd\_PacRF, donde se observa que PISCOpd\_Op tiene tasas superiores a PISCOpd\_PacRF, en contraste IMERG tiene una clara subestimación de las precipitaciones para el Pacífico. Existe similitud en la distribución espacial entre PISCOpd\_Op y PISCOpd\_PacRF, debido a que ambos fueron generados con similares bases de datos pluviométricos, sin embargo, también se observa la variabilidad en cuanto a magnitudes debido a que fueron generados con métodos de interpolación diferentes.

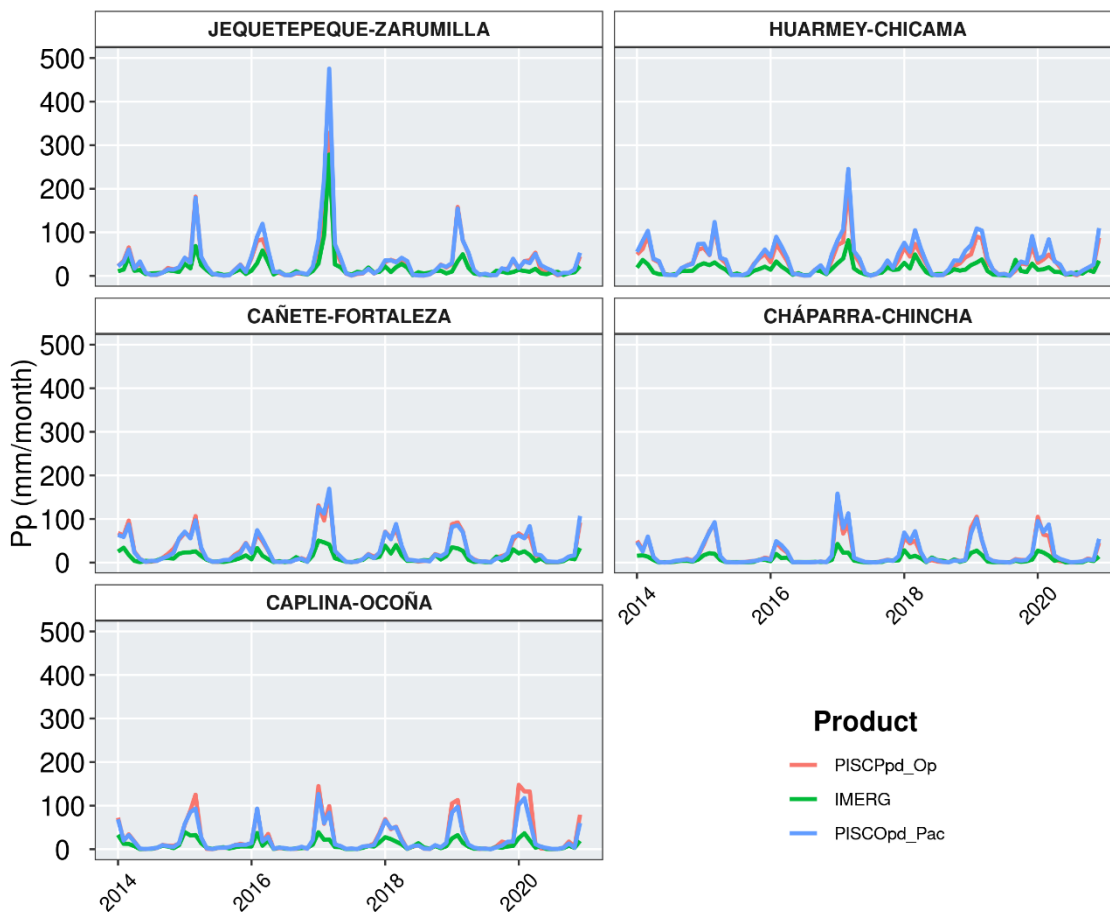


**Figura 06.** Distribución espacial del promedio interanual de Precipitación (Pp) para el periodo de referencia 2014-2020 para PISCOpd\_Op (left) IMERG (mid) y PISCOpd\_Pac (right).

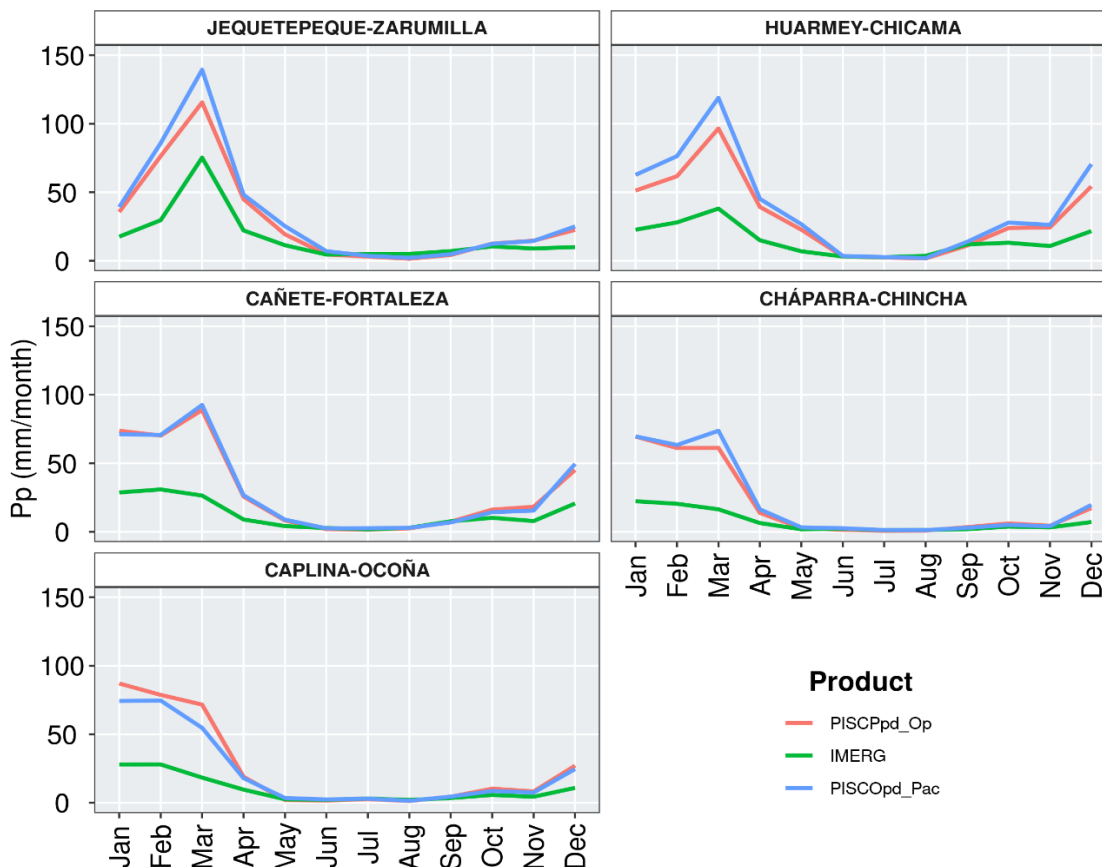
La comparación de la serie histórica mensual se presenta en la Figura 07, donde se resalta El Niño Costero del 2017 (Peng et al. 2019; Takahashi et al. 2018) con altos incrementos de las tasas de lluvia, sobre todo en las costas del Pacífico norte (región Jequetepeque – Zarumilla y Huarvey-Chicama) y con una menor modulación en el Pacífico central (Cañete- Fortaleza y Cháparra-Chincha).



Se generó una climatología mensual de referencia para el periodo de análisis (2014-2020), la cual se muestra en la Figura 08. Se observa que los tres productos representan la señal de las precipitaciones para la climatología en Perú con el periodo de lluvias comprendido entre los meses de diciembre a abril. Sin embargo, se observa que IMERG es la que subestima las lluvias conforme a los resultados de esta investigación. Esta comparativa evidencia más fácilmente que cuando se obtiene los promedios areales mensuales a nivel regional, PISCOpd\_PacRF tiene las mayores tasas de lluvias, con excepción de la región sur Caplina-Ocoña, en comparación con PISCOpd\_Op, a pesar de que contiene zonas con mayores tasas de lluvia (Figura 04), esto debido a la mejor distribución de las lluvias del producto PISCOpd\_PacRF generado con el método de RF-MEP (Baez-Villanueva et al. 2020) evidenciado con las métricas de rendimiento (Figura 05).

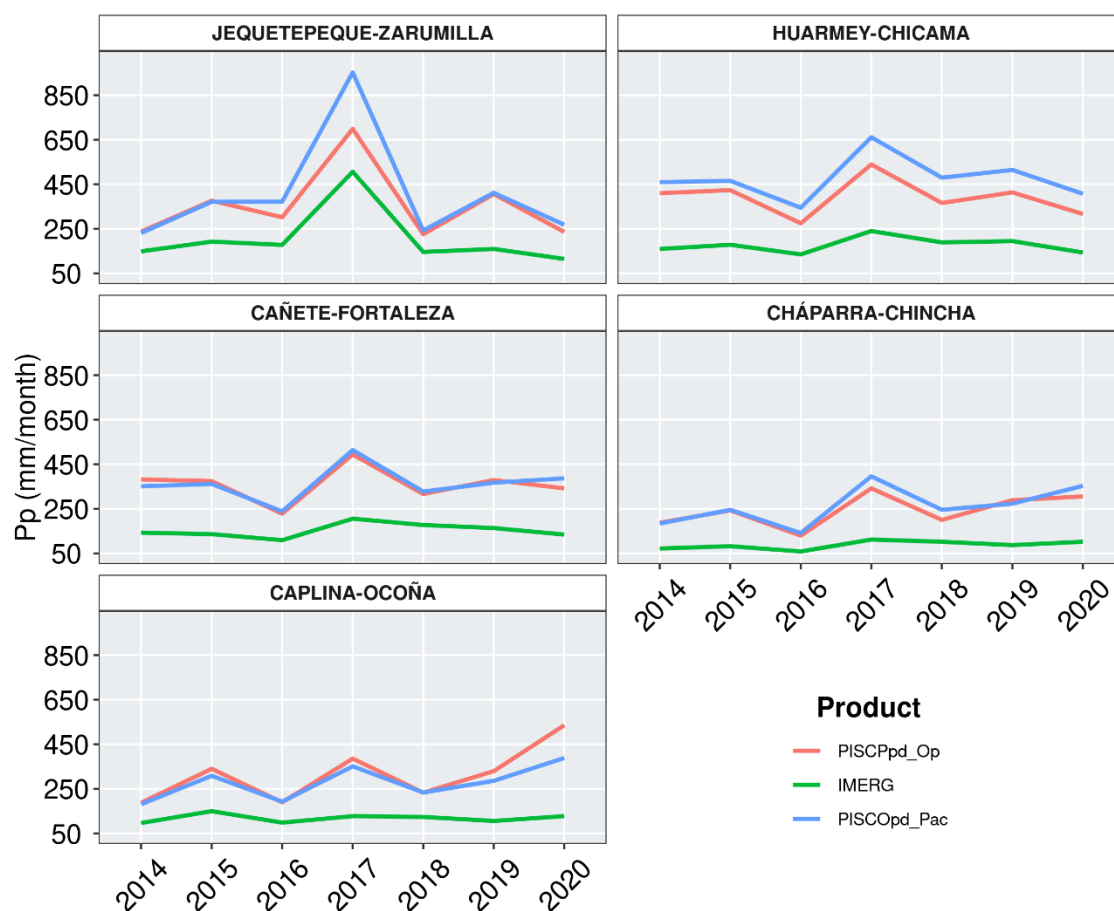


**Figura 07.** Serie histórica mensual de Precipitación (Pp) para el periodo de referencia 2014-2020 para PISCOpd\_Op (rojo) IMERG (verde) y PISCOpd\_Pac (azul).



**Figura 08.** Climatología mensual de Precipitación (Pp) para el periodo de 2014-2020 para PISCOpd\_Op (rojo) IMERG (verde) y PISCOpd\_Pac (azul).

La comparación de la serie histórica anual para el periodo de análisis (2014-2020), se muestra en la Figura 09. Igual que las series mensuales y climáticas, se observa que los tres productos presentan una similar señal de las precipitaciones, sin embargo, varían en sus intensidades. IMERG es la que cuenta con las menores intensidades de precipitación. Se vuelva a resaltar el hecho de que las tasas de lluvia regionales de PISCOpd\_PacRF son mayores en el Pacífico norte Jequetepeque-Zarumilla y Huarney-Chicama, y hay una leve similitud para el Pacífico central (Cañete-Fortaleza) y bajo sur (Cháparra-Chincha), mientras que en el Pacífico Sur (Caplina-Ocoña) se observa que PISCOpd\_Op supera levemente las intensidades de PISCOpd\_PacRF.



**Figura 09.** Serie histórica anual de Precipitación (Pp) para el periodo de referencia 2014-2020 para PISCOpd\_Op (rojo) IMERG (verde) y PISCOpd\_Pac (azul).

## 5. Notas de uso

PISCOpd\_PacRF es el segundo conjunto de datos meteorológicos grillados diarios de alta resolución para Perú. El primero fue PISCO para todo Perú con una resolución espacial de 10 km y el segundo PISCOpd\_PacRF con una resolución espacial de 5 km. Si bien es cierto que este conjunto de datos fue generado para cerrar la brecha de datos de lluvias de alta resolución para el Pacífico y productos hidrometeorológicos operativos, también se puede utilizar en una amplia gama de áreas como agricultura, seguros, silvicultura, climatología, meteorología, hidrología, ecología, mapeo de suelos, planificación urbana o cualquier otro campo de investigación que necesite datos grillados con una alta resolución espacial. PISCOpd\_PacRF está en formato NetCDF que lo hace interoperable con cualquier software GIS, como SAGA GIS (<http://www.saga-gis.org/>), QGIS (<http://www.qgis.org/>), ArcGIS (<https://www.arcgis.com/>), además que está enfocado en su lectura mediante código abierto como R, Python, etc. Cabe señalar que los valores de PISCOpd\_PacRF se encuentran en mm/día. Los datos están disponibles gratuitamente <https://doi.org/10.6084/m9.figshare.17304272> bajo la licencia Creative Commons: CC BY 4.0.

## 6. Referencias

- Aybar, Cesar, Carlos Fernández, Adrian Huerta, Waldo Lavado, Fiorella Vega, and Oscar Felipe-Obando. 2020. "Construction of a High-Resolution Gridded Rainfall Dataset for Peru from 1981 to the Present Day." *Hydrological Sciences Journal* 65(5):770–85. doi: <https://doi.org/10.1080/02626667.2019.1649411>.
- Baez-Villanueva, Oscar M., Mauricio Zambrano-Bigiarini, Hylke E. Beck, Ian McNamara, Lars Ribbe, Alexandra Nauditt, Christian Birkel, Koen Verbist, Juan Diego Giraldo-Osorio, and Nguyen Xuan Thinh. 2020. "RF-MEP: A Novel Random Forest Method for Merging Gridded Precipitation Products and Ground-Based Measurements." *Remote Sensing of Environment* 239(January). doi: 10.1016/j.rse.2019.111606.
- Baez-Villanueva, Oscar Manuel, Mauricio Zambrano-Bigiarini, Lars Ribbe, Alexandra Nauditt, Juan Diego Giraldo-Osorio, and Nguyen Xuan Thinh. 2018. "Temporal and Spatial Evaluation of Satellite Rainfall Estimates over Different Regions in Latin-America." *Atmospheric Research* 213(April):34–50. doi: 10.1016/j.atmosres.2018.05.011.
- Compo, G. P., J. S. Whitaker, P. D. Sardeshmukh, N. Matsui, R. J. Allan, X. Yin, B. E. Gleason, R. S. Vose, G. Rutledge, P. Bessemoulin, S. BroNnimann, M. Brunet, R. I. Crouthamel, A. N. Grant, P. Y. Groisman, P. D. Jones, M. C. Kruk, A. C. Kruger, G. J. Marshall, M. Maugeri, H. Y. Mok, O. Nordli, T. F. Ross, R. M. Trigo, X. L. Wang, S. D. Woodruff, and S. J. Worley. 2011. "The Twentieth Century Reanalysis Project." *Quarterly Journal of the Royal Meteorological Society* 137(654):1–28. doi: 10.1002/qj.776.
- Cornes, Richard C., Gerard van der Schrier, Else J. M. van den Besselaar, and Philip D. Jones. 2018. "An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets." *Journal of Geophysical Research: Atmospheres* 123(17):9391–9409. doi: 10.1029/2017JD028200.
- Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J. J. Morcrette, B. K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J. N. Thépaut, and F. Vitart. 2011. "The ERA-Interim Reanalysis: Configuration and Performance of the Data Assimilation System." *Quarterly Journal of the Royal Meteorological Society* 137(656):553–97. doi: 10.1002/qj.828.
- Dinku, T., P. Ceccato, E. Grover-Kopec, M. Lemma, S. J. Connor, and C. F. Ropelewski. 2007. "Validation of Satellite Rainfall Products over East Africa's Complex Topography." *International Journal of Remote Sensing* 28(7):1503–26. doi: 10.1080/01431160600954688.
- Gao, Y. C., and M. F. Liu. 2013. "Evaluation of High-Resolution Satellite Precipitation Products Using Rain Gauge Observations over the Tibetan Plateau." *Hydrology and Earth System Sciences* 17(2):837–49. doi: 10.5194/hess-17-837-2013.
- Garreaud, René D., Mathias Vuille, Rosa Compagnucci, and José Marengo. 2009. "Present-Day South American Climate." *Palaeogeography, Palaeoclimatology, Palaeoecology* 281(3–4):180–95. doi: 10.1016/j.palaeo.2007.10.032.

- Haslinger, Klaus, Daniel Koffler, Wolfgang Schoner, and Gregor Laaha. 2014. "Exploring the Link between Meteorological Drought and Streamflow: Effects of Climate-Catchment Interaction." 5375–77. doi: dx.doi.org/10.1002/2013WR015051.
- Hengl, Tomislav, Madlene Nussbaum, Marvin N. Wright, Gerard B. M. Heuvelink, and Benedikt Gräler. 2018. "Random Forest as a Generic Framework for Predictive Modeling of Spatial and Spatio-Temporal Variables." *PeerJ* 2018(8). doi: 10.7717/peerj.5518.
- Huerta, Adrian, and Waldo Lavado-Casimiro. 2021. "Trends and Variability of Precipitation Extremes in the Peruvian Altiplano (1971–2013)." *International Journal of Climatology* 41(1):513–28. doi: 10.1002/joc.6635.
- Huffman, G. J., Bolvin, D. T. & Nelkin, E. J. 2014. "Integrated Multi-Satellite Retrievals for GPM (IMERG), Final Run, Version V06." Retrieved (ftp://arthurhou.pps.eosdis.nasa.gov/gpmdata/).
- Kumar, Ashish, Raaj Ramsankaran, Luca Brocca, and Francisco Munoz-Arriola. 2019. "A Machine Learning Approach for Improving Near-Real-Time Satellite-Based Rainfall Estimates by Integrating Soil Moisture." *Remote Sensing* 11(19):1–20. doi: 10.3390/rs11192221.
- Lavado Casimiro, Waldo Sven, Josyane Ronchail, David Labat, Jhan Carlo Espinoza, and Jean Loup Guyot. 2012. "Analyse de La Pluie et de l'écoulement Au Pérou (1969–2004) : Bassins Versants Du Pacifique, Du Lac Titicaca et de l'Amazone." *Hydrological Sciences Journal* 57(4):625–42. doi: 10.1080/02626667.2012.672985.
- Lee, Moon-Hwan, Eun-Soon Im, and Deok-Hyo Bae. 2019. "Impact of the Spatial Variability of Daily Precipitation on Hydrological Projections: A Comparison of GCM and RCM driven Cases in the Han River Basin, Korea." *Hydrological Processes* (April):2240–57. doi: 10.1002/hyp.13469.
- Lehner, Bernhard, and Günther Grill. 2013. "Global River Hydrography and Network Routing: Baseline Data and New Approaches to Study the World's Large River Systems." *Hydrological Processes* 27(15):2171–86. doi: 10.1002/hyp.9740.
- Li, Zhe, Dawen Yang, and Yang Hong. 2013. "Multi-Scale Evaluation of High-Resolution Multi-Sensor Blended Global Precipitation Products over the Yangtze River." *Journal of Hydrology* 500:157–69. doi: 10.1016/j.jhydrol.2013.07.023.
- Llauca, Harold, Waldo Lavado-Casimiro, Cristian Montesinos, William Santini, and Pedro Rau. 2021. "PISCO\_HyM\_GR2M: A Model of Monthly Water Balance in Peru (1981–2020)." *Water (Switzerland)* 13(8):1–19. doi: 10.3390/w13081048.
- Millan, Carlos, Waldo Lavado, Fiorella Vega, Oscar Felipe, Julia Acuña, and Ken Takahashi. 2021. "SILVIA : An Operational System to Monitoring Landslides Forced by Heavy Precipitations at National Scale in Peru." P. 10970 in *EGU General Assembly 20220*. Vol. 1.
- Muñoz Sabater, J. 2019. "ERA5-Land Hourly Data from 1981 to Present." *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*.

- Nguyen, Phu, Eric J. Shearer, Hoang Tran, Mohammed Ombadi, Negin Hayatbini, Thanh Palacios, Phat Huynh, Dan Braithwaite, Garr Updegraff, Kuolin Hsu, Bob Kuligowski, Will S. Logan, and Soroosh Sorooshian. 2019. "The CHRS Data Portal, an Easily Accessible Public Repository for PERSIANN Global Satellite Precipitation Data." *Scientific Data* 6:1–10. doi: 10.1038/sdata.2018.296.
- Peng, Qihua, Shang Ping Xie, Dongxiao Wang, Xiao Tong Zheng, and Hong Zhang. 2019. "Coupled Ocean-Atmosphere Dynamics of the 2017 Extreme Coastal El Niño." *Nature Communications* 10(1):1–10. doi: 10.1038/s41467-018-08258-8.
- Petritsch, Richard, and Hubert Hasenauer. 2014. "Climate Input Parameters for Real-Time Online Risk Assessment." *Natural Hazards* 70(3):1749–62. doi: 10.1007/s11069-011-9880-y.
- PSL, Physical Sciences Laboratory. n.d. "CPC Global Unified Gauge-Based Analysis of Daily Precipitation."
- Sekulić, Aleksandar, Milan Kilibarda, Gerard B. M. Heuvelink, Mladen Nikolić, and Branislav Bajat. 2020. "Random Forest Spatial Interpolation." *Remote Sensing* 12(10):1–29. doi: 10.3390/rs12101687.
- Sekulić, Aleksandar, Milan Kilibarda, Dragutin Protić, and Branislav Bajat. 2021. "A High-Resolution Daily Gridded Meteorological Dataset for Serbia Made by Random Forest Spatial Interpolation." *Scientific Data* 8(1):1–12. doi: 10.1038/s41597-021-00901-2.
- Sippel, Sebastian, Nicolai Meinshausen, Erich M. Fischer, Enikő Székely, and Reto Knutti. 2020. "Climate Change Now Detectable from Any Single Day of Weather at Global Scale." *Nature Climate Change* 10(1):35–41. doi: 10.1038/s41558-019-0666-7.
- Takahashi, K., Aliaga-Nestares, Avalos V., Bouchon G., A. M., Castro, L. Cruzado, B. Dewitte, D. Gutiérrez, W. Lavado-Casimiro, J. Marengo, A. G. Martínez, K. Mosquera-Vásquez, and N. Quispe. 2018. "The 2017 Coastal El Niño [in State of the Climate in 2017]." *Bulletin of the American Meteorological Society* 99:S210–S211. doi: <https://doi.org/10.1175/2018BAMSStateoftheClimate.1>.
- Thiemig, Vera, Rodrigo Rojas, Mauricio Zambrano-Bigiarini, Vincenzo Levizzani, and Ad De Roo. 2012. "Validation of Satellite-Based Precipitation Products over Sparsely Gauged African River Basins." *Journal of Hydrometeorology* 13(6):1760–83. doi: 10.1175/JHM-D-12-032.1.
- Turini, Nazli, Boris Thies, Natalia Horna, and Jörg Bendix. 2021. "Random Forest-Based Rainfall Retrieval for Ecuador Using GOES-16 and IMERG-V06 Data." *European Journal of Remote Sensing* 54(1):117–39. doi: 10.1080/22797254.2021.1884002.
- Zambrano-Bigiarini, Mauricio. 2017. "HydroGOF: Goodness-of-Fit Functions for Comparison of Simulated and Observed Hydrological Time Series."
- Zambrano-Bigiarini, Mauricio, Alexandra Nauditt, Christian Birkel, Koen Verbist, and Lars Ribbe. 2017. "Temporal and Spatial Evaluation of Satellite-Based Rainfall Estimates across the Complex Topographical and Climatic Gradients of Chile." *Hydrology and Earth System Sciences* 21(2):1295–1320. doi: 10.5194/hess-21-1295-2017.



## 7. Apéndice:

### Apéndice A

El nuevo producto operativo se está ejecutando, y corre todos los días en los servidores de SENAMHI, en el horario de las 12:00 pm ya que hay una latencia de 5 horas del producto IMERG early. A continuación, se presenta el código:

```
# title: "RFmerge in Peru"
# author: "Carlos Millan"
pacman::p_load(zoo, sf, raster)
setwd("/home/dhi/Scripts_DHI/PISCOpd_PacRF/RFmerge/")
rm(list = ls())
a <- Sys.time()
utm <- CRS("+init=epsg:32718") # WGS 84 / UTM zone 18S
wgs <- CRS("+proj=longlat +datum=WGS84 +no_defs") # WGS 84 / UTM zone 18S
vozdata=function(){
  ur=#####
  tabs = httr::GET(ur)
  tabla=XML::readHTMLTable(rawToChar(tabs$content), stringsAsFactors = F)[[2]]
  data=tabla[,3:6]
  row.names(data)=NULL
  names(data)=c("raingauge", 'y', 'x', 'obs')
  data$raingauge = paste("X", data$raingauge, sep = "") #as.numeric(as.vector(data$x))
  data$x=as.numeric(as.vector(data$x))
  data$y=as.numeric(as.vector(data$y))
  data$obs=as.numeric(as.vector(data$obs))
  data=na.omit(data)
  data=data[data$obs>=0,]
  data[c("raingauge", 'x', 'y', 'obs')]
}

datos = vozdata() #info del dia
StationsPp <- zoo::zoo(t(datos$obs), Sys.Date()-1)
colnames(StationsPp) <- datos$raingauge
```

```

# 2. Loading input data -----
# Daily covariates
IMERG10km <- raster(paste("IMERG_Peru /V06C /early_daily /IMERG_V06C_Early_
daily_", Sys.Date()-1, ".tif", sep = ""))
# PISCO10km <- raster(paste("/mnt/dhi-pisco/PISCOO/PISCO_Operativo", Sys.Date()-
1, ".tif", sep = ""))
# Monthly covariates
ZV_climat10km <- brick("Capas/ PP_ZonasVidaClimatologia/ ZonaVidas_climat.nc")
[[lubridate::month(Sys.Date()-1)]]
CHIRPS_clim5km <- brick("Capas/Climatologias/clima_chirps.tif")[[lubri-
date::month(Sys.Date()-1)]]
# Spatial covariates
PacificDEM1km <- raster("Capas/dem_1km/demPE_hs_wgs_1km.tif")
grilla.ref <- raster(nrows=2139, ncols=1459, res=0.01, xmn=-82.50, xmx=-67.91, ymn=-
19.85, ymx=1.54, crs="+proj=longlat +datum=WGS84 +no_defs")
PacificDEM.mask <- raster::resample(PacificDEM1km,grilla.ref)
# Study area
PacificSf.mask <- read_sf("Shp/Mask_Pac.shp")
PacificDEM.mask <- crop(PacificDEM1km,PacificSf.mask) %>% mask(PacificSf.mask)
# PacificDEM.mask <- aggregate(PacificDEM.mask,5)
# 3.3 Stations at the project -----
# Filter station by domain area
stations <- datos
stations <- st_as_sf(stations, coords = c('x', 'y'), crs = st_crs("+proj=longlat +da-
tum=WGS84 +no_defs"))
stations <- stations[PacificSf.mask,]

StationsPp <- StationsPp[,which(names(StationsPp) %in% stations$raingauge)]
stations <- stations[which(stations$raingauge %in% names(StationsPp)),]
# 4. Preparing input data -----
list.clim <- list(IMERG10km, ZV_climat10km, CHIRPS_clim5km) #PISCO10km,
climatos <- lapply(list.clim,FUN = function(x){

```



```

y <- resample(x,PacificDEM.mask) #Resampleamos
y <- crop(y,PacificDEM.mask) #Cortamos
y
})
slope.1km <- terrain(PacificDEM.mask,opt="slope")
aspect.1km <- terrain(PacificDEM.mask,opt="aspect")
lon <- init(PacificDEM.mask,"x")
lat <- init(PacificDEM.mask,"y")
covariates <- list(zonavida=climatos[[2]],
                  CHIRPS_clim=climatos[[3]],
                  dem = PacificDEM.mask, slope = slope.1km,
                  aspect =aspect.1km, long=lon, lat =lat
                  )
names(covariates) <- c("zonavida","CHIRPS_clim","dem","slope","aspect","long","lat")
#"imerg",
covariates.utm <- list(imerg=climatos[[1]],
                      zonavida=covariates[[1]],
                      CHIRPS_clim=covariates[[2]],
                      dem = covariates[[3]], slope = covariates[[4]],
                      # aspect =covariates[[5]],
                      long=covariates[[6]], lat =covariates[[7]]
                      )
compareRaster(covariates.utm)
st.coords <- st_coordinates(stations)
lon <- st.coords[, "X"]
lat <- st.coords[, "Y"]
StationsGIS <- data.frame(ID=stations[["raingauge"]], lon=lon, lat=lat)
## 7 Running RFmerge
m <- extent(PacificDEM.mask)
library(RFmerge)
rfmep <- RFmerge(x=StationsPp, metadata= StationsGIS, cov=covariates.utm,
                 id="ID", lat="lat", lon="lon", mask=m.def,

```

```
training=1, write2disk=F, #drty.out=drty.out,  
parallel="none", par.nnodes=par.nnodes, ED=F, ntree = 3000)  
rfmep[Which(rfmep<0)] <- 0  
writeRaster(rfmep, paste("/mnt/dhi-pisco/SEH/PISCOpd_PacRF/PISCOpd_  
PacRF_", Sys.Date()-1, ".tif", sep = ""),  
format="GTiff", overwrite=T, varname="Precipitation", varunit="mm",  
longname="Precipitation RFMEP")
```

## Apéndice B

Series históricas de las estaciones automáticas usadas para la validación.

