

UNIVERSIDAD NACIONAL AGRARIA

LA MOLINA

FACULTAD DE CIENCIAS



**“CORRECCIÓN DE ESTIMACIONES DE PRECIPITACIÓN POR
SATÉLITE GPM-IMERG USANDO TÉCNICA DE MEZCLA SOBRE
LAS CUENCAS CHILLÓN, RÍMAC Y LURÍN”**

Presentada por:

SANDRA THALÍA RIVADENEIRA MALLQUI

TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE

INGENIERO METEORÓLOGO

Lima-Perú

2019

UNIVERSIDAD NACIONAL AGRARIA

LA MOLINA

FACULTAD DE CIENCIAS

“CORRECCIÓN DE ESTIMACIONES DE PRECIPITACIÓN POR SATÉLITE GPM-
IMERG USANDO TÉCNICA DE MEZCLA SOBRE LAS CUENCAS CHILLÓN,
RÍMAC Y LURÍN”

Presentada por:

Sandra Thalía Rivadeneira Mallqui

Tesis para Optar el Título Profesional de:

INGENIERO METEORÓLOGO

Sustentada y aprobada por el siguiente Jurado:

Mg.Sc.Eusebio Idelmo Cisneros Tarmeño
PRESIDENTE

Ing. Franklin Delio Unsihuay Tovar
MIEMBRO

Ing. Hector Ladislao Huisacaina Soto
MIEMBRO

Mg.Sc. Victoria Doris Calle Montes
ASESORA

Ing. Adrian Marko Huerta Julca
CO-ASESOR

DEDICATORIA

A mis padres Walter y Bertha por el apoyo y amor incondicional que me han brindado a lo largo de mi formación universitaria y los valores que siempre me han inculcado para ser una mejor persona cada día.

A mi hermano Richard por ser un ejemplo de superación que me enseñó a vencer mis propios límites y que con sabiduría supo ser un guía en mi formación personal y profesional.

A mi hermana Carla por estar presente en los buenos y malos momentos, así como brindarme sus palabras de aliento en situaciones de adversidad.

AGRADECIMIENTOS

Agradezco a la Universidad Nacional Agraria la Molina por darme la oportunidad de formarme profesionalmente en su casa de estudios.

Un agradecimiento especial al Ing. Adrian Huerta por la oportunidad que me otorgó para realizar esta tesis, por su paciencia en este proceso y por apoyarme con sus conocimientos y sugerencias en la investigación. Así como, a la Mg.Sc Victoria Calle por su paciencia y apoyo en la resolución de mis dudas y consultas.

Agradecerles también a mis profesores de la carrera de meteorología por incentivar me a ser mejor profesional y transmitirme sus conocimientos.

Finalmente, un agradecimiento a Luis Víctor por estar a mi lado en este proceso, brindarme su apoyo y motivarme en la culminación de esta tesis.

ÍNDICE GENERAL

I. INTRODUCCIÓN.....	1
II. REVISIÓN DE LA LITERATURA.....	3
2.1 Clasificación climatológica en la vertiente occidental del Perú	3
2.2 Variabilidad de precipitación en la vertiente occidental del Perú	7
2.2.1 “El Niño Costero 2017” y sus efectos en la precipitación	10
2.3 Percepción remota de la precipitación	12
2.4 Datos satelitales GPM y sus productos IMERG.....	13
2.5 Espacialización de la precipitación.....	18
2.6 Métricas estadísticas	19
III. MATERIALES Y MÉTODOS	22
3.1 Materiales	22
3.1.1 Zona de estudio	22
3.1.2 Información observada.....	25
3.1.3 Información satelital.....	27
3.1.4 Equipos.....	28
3.2 Metodología.....	29
3.2.1 Efectuar el análisis exploratorio de datos pluviométricos (convencionales y automáticos) y completación de la información.	30
3.2.2 Evaluar y validar las estimaciones de precipitación de tres productos IMERG en base a estaciones pluviométricas.	32
3.2.3 Desarrollar y validar la técnica de mezcla para corregir las estimaciones de precipitación por satélite sobre el área de estudio.	35
IV. RESULTADOS Y DISCUSIÓN.....	40
4.1 Efectuar el análisis exploratorio de datos pluviométricos (convencionales y automáticas) y completación de información.	40
4.1.1 Análisis exploratorio de la información	40
4.1.2 Completación de datos	45
4.2 Evaluar y validar las estimaciones de precipitación de tres productos IMERG en base a estaciones pluviométricas.	48

4.2.1 Validación de datos diarios de precipitación en estaciones automáticas en base a convencionales.	48
4.2.2 Validación espacio-temporal de la precipitación estimada por tres productos IMERG en base a estaciones convencionales.	51
4.2.3 Validación espacio-temporal de la precipitación estimada por tres productos IMERG en base a estaciones automáticas.	56
4.3 Desarrollar y validar la técnica de mezcla para corregir las estimaciones de precipitación por satélite sobre el área de estudio.	63
4.3.1 Cálculo de residuales y estimación del mejor parámetro Idp	63
4.3.2 Obtención de producto de mezcla mediante técnica Regression IDW	65
4.3.3 Validación del método de mezcla mediante validación cruzada "LOOCV".....	69
V. CONCLUSIONES.....	73
VI. RECOMENDACIONES	75
VII. REFERENCIAS BIBLIOGRÁFICAS	76
VIII. ANEXOS	83

ÍNDICE DE TABLAS

Tabla 1: Nomenclatura del sistema de clasificación para la precipitación efectiva	4
Tabla 2: Nomenclatura del sistema de clasificación para la temperatura eficiente.....	4
Tabla 3: Nomenclatura del sistema de clasificación para las características de la distribución de la precipitación anual.	4
Tabla 4: Nomenclatura del sistema de clasificación para la humedad relativa media	5
Tabla 5: Cuadro resumen de los estadísticos empleados para el análisis de datos de precipitación.....	19
Tabla 6: Productos de estimación de precipitación con el algoritmo IMERG	27
Tabla 7: Estadísticos descriptivos de datos de precipitación diaria en estaciones convencionales.	43
Tabla 8 : Estadísticos descriptivos de datos de precipitación diaria en automáticas.....	44
Tabla 9: Resultado de estadísticos entre pares de estaciones convencionales y automáticas	50
Tabla 10: Altura y clasificación climática entre los pares de estaciones.....	50
Tabla 11: Estadísticos entre los datos de precipitación diaria del producto satelital y el registrado por el pluviómetro convencional, para dos periodos	52
Tabla 12: Estadísticos entre los datos de precipitación diaria del producto satelital y el registrado por el pluviómetro automático, para dos periodos y cada tiempo de acumulación.	57
Tabla 13: Estadísticos para cuatro estaciones entre los datos observados y el producto satelital IMERG-late sin corregir (SC) y la validación cruzada de IMERG- late corregido (C).	70

ÍNDICE DE FIGURAS

Figura 1: Precipitación para el periodo 1964-2011 en las 9 regiones climáticamente homogéneas	8
Figura 2: Las nueve regiones de lluvias homogéneas después de la regionalización. Proceso de agrupamiento y RVM(método del vector regional). Superficie interpolada de precipitación anual.	9
Figura 3: Evolución de la anomalía mensual de la TSM en el Pacífico ecuatorial entre diciembre 2016 y marzo 2017, utilizando datos proporcionados por NOAA/OAR/ESRL PSD	11
Figura 4: Constelación de satélites que contribuyen con la misión de GPM	15
Figura 5: Componentes del satélite GPM-Core o núcleo	16
Figura 6: Concepto de observación de precipitación por el observatorio GPM núcleo... ..	16
Figura 7: Ejemplo de un diagrama de Taylor	21
Figura 8: Mapa de la zona de estudio CHIRILU	24
Figura 9: Ubicación de las estaciones pluviométricas convencionales y automáticas en la región CHIRILU	26
Figura 10: Procedimientos de la metodología empleada.....	29
Figura 11: Validación cruzada dejando uno fuera LOOCV.....	39
Figura 12: Disponibilidad de estaciones pluviométricas convencional con información para el periodo 2000-2018, los puntos negros indican que al año existieron al menos 95 por ciento de los datos.	40
Figura 13: Disponibilidad de estaciones automáticas con información para el periodo 2007-2018, los puntos negros indican que al año existieron al menos 95 por ciento de los datos.	41
Figura 14: Gráficos boxplot de cuatro estaciones convencionales para acumulados mensuales de precipitación en el periodo 2000-2018.....	42
Figura 15: Mapa de estaciones por categorías de estación en la región CHIRILU	45
Figura 16: Estaciones pluviométricas convencionales que pasaron el proceso de completación de datos.....	46
Figura 17: Estaciones pluviométricas automáticas que pasaron el proceso de completación de datos.....	47

Figura 18: Pares de estaciones pluviométricas convencionales y automáticas en la región de estudio CHIRILU.....	49
Figura 19: Mapa de pares de estaciones pluviométricas automáticas y convencionales sobre la clasificación climática.....	51
Figura 20: Series de tiempo de las estaciones con mejores estadísticos para la época EFM-lluvioso y JJA-seco en estaciones convencionales.....	52
Figura 21: Boxplot de los estadísticos aplicados a todas las estaciones pluviométricas convencionales por periodo lluvioso y seco	53
Figura 22: Mapa de análisis espacial de los resultados estadísticos de correlación de pearson “R” (izquierda) y bias relativo “RB (mm)” (derecha) para el periodo lluvioso y seco, entre los datos observados y los tres productos del algoritmo IMERG. La clasificación climática está representada por “CC”.	54
Figura 23: Función de Distribución Acumulada de Probabilidad (CDF) entre datos de precipitación observada convencional y los tres productos del algoritmo IMERG para dos periodos.	56
Figura 24: Series de tiempo de las estaciones con mejores estadísticos para la época EFM-lluviosa y JJA-seco en estaciones automáticas.	58
Figura 25: Boxplot de los estadísticos aplicados a todas las estaciones pluviométricas automáticas por periodo lluvioso y seco.....	59
Figura 26: Resultados estadísticos de la validación, coeficiente de correlación de pearson (R) y el sesgo relativo o bias (RB) sobre cuatro tipos de climas para cinco tiempos de acumulación de precipitación	60
Figura 27: Función de Distribución Acumulada de Probabilidad (CDF) entre datos de precipitación observada convencional y los tres productos del algoritmo IMERG para dos periodos.	61
Figura 28: Histogramas de los parámetros Idp empleados en la interpolación RIDW para cada periodo lluvioso.	64
Figura 29: Mapa de valores observados calculados por la función lineal (izquierda) y mapa de valores residuales de la función lineal (derecha).	65
Figura 30: Gráficos de dispersión entre los datos observados de estaciones pluviométricas, producto satelital IMERG-Late y IMERG-Late corregido.....	66
Figura 31: Mapa del producto satelital IMERG-late (izquierda) el producto satelital corregido por técnica de mezcla o merging (derecha).....	67

Figura 32: Precipitaciones (mm/día) para el 10 de Marzo del 2017 sobre la cuenca Chillón	68
Figura 33: Precipitaciones (mm/día) para el 14 de Marzo del 2017 sobre la cuenca Rímac	68
Figura 34: Precipitación (mm/día) para el 14 de Marzo del 2017 sobre la cuenca Lurín ..	68
Figura 35: Mapa de análisis espacial de correlación de pearson R entre los datos observados vs producto satelital IMERG-Late sin corregir “R (SC)” y la validación cruzada de IMERG-Late corregido “R(C)”.	71
Figura 36: Mapa de análisis espacial de sesgo o bias relativo RB (mm) entre los datos observados vs producto satelital IMERG-Late sin corregir “RB (SC)” y la validación cruzada de IMERG-Late corregido “RB(C)”	71
Figura 37: Diagramas de Taylor para evaluar la eficiencia de los producto IMERG-Late sin corregir (SC) y IMERG-Late (C) validado por LOOCV.....	72

ÍNDICE DE ANEXOS

ANEXO 1: Información de estaciones pluviométricas disponibles	83
ANEXO 2: Gráficos de boxplot de estaciones convencionales y automáticas en la región de estudio CHIRILU de precipitación promedio (mm/mes) para el registro histórico.....	85
ANEXO 3: Información de estaciones pluviométricas seleccionadas	88
ANEXO 4: Gráficos de series de tiempo de estaciones convencionales y los tres productos satelitales IMERG-Early, IMERG-Late y IMERG-Final.....	89
ANEXO 5: Gráficos de series de tiempo de estaciones automáticas y los tres productos satelitales IMERG-Early, IMERG-Late y IMERG-Final.....	91
ANEXO 6: Código en R para el cálculo del método Regression IDW	93
ANEXO 7: RMSE entre los datos observados y el producto satelital IMERG-late sin corregir (RMSE_sc) y la validación cruzada de IMERG-late corregido (RMSE_c).....	95
ANEXO 8: Distribución altitudinal de los estadísticos de correlación R y sesgo RB entre los datos observados y el producto satelital IMERG-Late sin corregir (R_sc y RB_sc) y la validación cruzada LOOCV de IMERG-late corregido (R_c y RB_c).	98
ANEXO 9: Esquema resumen para la obtención del producto mezclado mediante la metodología planteada en la tesis	99

RESUMEN

Esta investigación tiene por objetivo corregir las estimaciones de precipitación por satélite GPM-IMERG sobre un área que abarca tres cuencas que pertenecen al departamento de Lima, por medio de una técnica de mezcla entre datos observados en tierra y del algoritmo que proporciona productos de precipitación Multi-satélite denominado IMERG V05 (Integrated Multi-Satellite Retrievals for GPM) para la reciente misión GPM (Global Precipitation Measurement), disponible en tiempo real a una baja resolución espacial. Para ello, se realizó un previo análisis de los datos medidos en tierra así como una validación entre los datos que provienen de dos tipos de registros pluviométricos (convencional y automático) y las estimaciones de tres productos GPM-IMERG; IMERG-Early e IMERG-Late (data en tiempo real) y IMERG-Final (después del tiempo real). Después de determinar cuál de estos tres productos se asemeja mejor a los datos reales, se aplicó la técnica de mezcla RIDW (Regression Inverse Distance Weighting) entre los datos observados y el producto satelital, posteriormente se evaluó en cuánto mejoró la detección cuantitativa y cualitativa de la precipitación al aplicar la técnica. Los resultados muestran que IMERG-Final presenta mejor correlación con respecto a los datos observados, sin embargo sobreestima la precipitación en periodos lluviosos y secos, mientras que IMERG-Late presenta valores de correlación cercano a IMERG-Final y el sesgo con respecto a los datos observados es menor por lo que se desempeña mejor para estimaciones diarias a comparación de IMERG-Early. La técnica de mezcla aplicada a los datos observados y satelitales IMERG-Late fue validada, evidenciando una mejora de las estimaciones con alta detección de eventos lluviosos, la correlación mejoró en aproximadamente 20 por ciento (0.4 a 0.6) y el RMSE disminuyó entre 1 a 3 mm en promedio para datos diarios. Por lo que el producto de mezcla obtenido, pretende ser una herramienta para múltiples aplicaciones, entre ellas; en la predicción de eventos hidrometeorológicos que requiere de información confiable en tiempo real.

PALABRAS CLAVES: precipitación, GPM, IMERG-Early, IMERG-Late, IMERG-Final, Técnica de mezcla.

ABSTRACT

This research aims to correct the estimates of precipitation by satellite GPM-IMERG on an area that covers three basins that belong to the department of Lima, by means of a merging technique between observed data on land and the algorithm that provides products of precipitation Multi-satellite called IMERG V05 (Integrated Multi-Satellite Retrievals for GPM) for the recent mission GPM (Global Precipitation Measurement), available in real time at a low spatial resolution. To this end, a preliminary analysis was made of the data measured on the ground as well as a validation between the data coming from two types of rainfall records (conventional and automatic) and the estimations of three GPM-IMERG products; IMERG-Early and IMERG-Late (real-time data) and IMERG-Final (after real time). After determining which of these three products is more similar to the actual data, the merging technique RIDW (Regression inverse Distance Weighting) was applied between the observed data and the satellite product, then it was evaluated in how much improved the detection quantitative and qualitative precipitation when applying the technique. The results show that IMERG-Final presents the best correlation with respect to the observed data, however overestimate the precipitation in periods rainy and dry, while IMERG-Late presents correlation values close to IMERG-Final and the bias with respect to the observed data is lower, so it performs better for daily estimates compared to IMERG-Early. The merging technique applied to the observed data and satellite IMERG-Late was validated, showing an improvement of estimates with high detection of rainy events, the correlation improved by approximately 20 percent (0.4 to 0.6) and the RMSE decreased between 1 to 3 mm in average for daily data. As a result of the product merging, it pretends to be a tool for multiple applications, among them; in predicting hydrometeorological events that requires reliable information.

KEY WORDS: Precipitation, GPM, IMERG-Early, IMERG-Late, IMERG-Final, merging technique.

I. INTRODUCCIÓN

Contar con estimaciones cuantitativas de precipitación cada vez más exactas es fundamental en la predicción de fenómenos meteorológicos extremos, en el modelamiento hidrológico, así como en la planificación de actividades agrícolas. Por ello es necesario aumentar la precisión en su estimación y para lograrlo se requiere de registros pluviométricos más confiable que cuenten con una distribución espacial adecuada. Sin embargo, en nuestro país la densidad de pluviómetros no es homogénea, esta es una gran desventaja cuando se quiere cuantificar la cantidad de precipitación sobre una región. Según Hu *et al.* (2019), la información precisa sobre la distribución de la precipitación es la base para comprender científicamente los cambios globales o regionales y mejorar nuestra capacidad para enfrentar los desastres naturales así como para optimizar la gestión de los recursos hídricos.

Al resultar complicada la medición de la precipitación en lugares de topografía compleja, el sensoramiento remoto se ha convertido en la actualidad en una herramienta ventajosa, que permite contar con información en aquellas zonas no accesibles y en donde medir la lluvia con pluviómetros sería casi imposible. En ese sentido, la información proveniente de productos de estimación de precipitación por satélites meteorológicos es una herramienta que nos ha ayudado a superar las limitaciones que presentan las redes de observación de precipitación por estaciones. Adicional a esto, recientemente se está haciendo uso de la unión de ambos tipos de información (estaciones y satelitales) para generar un producto de mezcla que estime con mayor exactitud, con el propósito de obtener un producto de calidad manteniendo la información de distribución espacial de la precipitación proporcionadas por satélites y la observación cuantitativa en estaciones (Martens *et al.* 2013).

Las regiones con características áridas como la costa peruana, se enfrentan a numerosos desafíos en el manejo de sus recursos hídricos y en situaciones con precipitaciones extremas se producen inundaciones devastadoras, con consecuencias más graves que en las regiones húmedas (Nashwan *et al.* 2019). Un reciente evento catastrófico, ocurrido durante el “El Niño Costero 2017”, generó fuertes lluvias que causaron desbordes, aluviones e inundaciones que afectaron a varias localidades (INDECI, 2017), así mismo desencadenó en la muerte de 114 personas, 38 728 viviendas colapsadas y 1.08 millón de afectados en el Perú. En Lima, las cuencas Chillón, Rímac y Lurín, también se vieron afectadas por el desborde de sus ríos causando daños materiales y pérdidas humanas, se estiman 16000 damnificados y 17 fallecidos (Bedoya, 2017). Durante el periodo se desarrollaron algunos eventos de precipitación de moderadas a fuerte intensidad entre el 14 de enero al 5 de abril del 2017, como el ocurrido el 03 de abril, que a consecuencias de las precipitaciones pluviales se originó el desborde del río Rímac, afectando viviendas en el distrito de Santa Eulalia (COEN, 2017). Estos episodios han demostrado la necesidad de contar con información de precipitación a una alta resolución temporal y espacial sobre esta región, como la proporcionada por la mezcla de datos observados y satelitales, la cual sería una herramienta de gran utilidad para pronosticadores, hidrólogos, operadores de embalse, así como en la gestión de riesgos y desastres.

El objetivo general de esta investigación es: corregir las estimaciones de precipitación por satélite GPM-IMERG a través de una técnica de mezcla entre datos observados por estaciones pluviométricas con datos del algoritmo IMERG (Integrated Multi-satellite Retrievals) sobre el área que comprende a las cuencas Chillón, Rímac y Lurín. Los objetivos específicos son: 1) Efectuar el análisis exploratorio de datos pluviométricos (convencionales y automáticos) y completación de información, 2) Evaluar y validar las estimaciones de precipitación de tres productos IMERG en base a estaciones pluviométricas, 3) Desarrollar y validar la técnica de mezcla para corregir las estimaciones de precipitación por satélite sobre el área de estudio.

II. REVISIÓN DE LA LITERATURA

2.1 Clasificación climatológica en la vertiente occidental del Perú

Debido a que el Perú se sitúa en una zona tropical deberíamos disfrutar de un clima cálido, con lluvias abundantes. Sin embargo, en el clima intervienen en una serie de factores que lo modifican decisivamente. Tres son los factores que determinan básicamente el clima del Perú: la situación del país en la zona intertropical, las modificaciones altitudinales que introduce la cordillera de los Andes y la Corriente Peruana o de Humboldt, cuyas aguas recorren las costas del país.

La clasificación climática del Perú se ha realizado con la necesidad de señalar las zonas del territorio que presentan características climáticas peculiares. La información base de esta clasificación está conformada por datos meteorológicos de 20 años que comprenden el periodo de 1965 a 1984 en el cual se cuenta con “índices climáticos” de acuerdo con el sistema de clasificación de “Werren Thornthwaite”. Para la construcción de esta clasificación se consideró las variables climáticas de precipitación, temperatura del aire y humedad, además de factores como (SENAMHI, 2002):

- Latitud y altitud.
- La cordillera de los andes.
- La corriente fría Marítima Peruana.
- El anticiclón del pacífico sur.
- La continentalidad, etc.

A continuación se muestran los cuadros con las características y nomenclatura del sistema de clasificación.

Tabla 1: Nomenclatura del sistema de clasificación para la precipitación efectiva

Valores del Índice Anual I	Carácter del clima	Símbolo	Vegetación Característica
128 o mayores	Muy lluvioso	A	Selva
64 a 127	Lluvioso	B	Bosque
32 a 63	Semiseco	C	Pastizal
16 a 31	Semiárido	D	Estepa
Menor de 16	Árido	E	Desierto

FUENTE: SENAMHI, 2002

Tabla 2: Nomenclatura del sistema de clasificación para la temperatura eficiente

Valores del Índice Anual I	Carácter del clima	Símbolo	Zonas
128 o mayores	Cálido	A'	Tropical
101 a 127	Semicálido	B'_1	
80 a 100	Templado	B'_2	Mesotermal
64 a 79	Semifrío	B'_3	
32 a 63	Frío	C'	Microtermal
16 a 31	Semifrígido	D'	
1 a 15	Frígido	E'	
0	Polar	F'	Hielo Perpetuo

FUENTE: SENAMHI, 2002

Tabla 3: Nomenclatura del sistema de clasificación para las características de la distribución de la precipitación anual.

Símbolo	Significado
r	Precipitación abundante en todas las estaciones del año
i	Con invierno seco
p	Con primavera seca
v	Con verano seco
o	Con otoño seco
d	Deficiencia de precipitación en todas las estaciones del año

FUENTE: SENAMHI, 2002

Tabla 4: Nomenclatura del sistema de clasificación para la humedad relativa media

Valores Medios Anuales H	Características del Clima	Símbolo
25% a 44%	Muy seco	H_1
45% a 64%	Seco	H_2
65% a 84%	Húmedo	H_3
85% a 100%	Muy húmedo	H_4

FUENTE: SENAMHI, 2002

Los valores de los índices mostrados se calculan en función al grado de eficacia de la precipitación y la temperatura ambiental por el método de Thornthwaite.

- Valor del índice anual I de la precipitación efectiva

$$I = \sum_{i=1}^{12} i, \text{ donde } i = 1.64 \left(\frac{P}{T+12.2} \right)^{10/9}$$

- Valor del índice anual I de la temperatura eficiente

$$I' = \sum_{i'=1}^{12} i', \text{ donde } i' = \frac{9T}{20}$$

Considerando para ambos índices:

i = índice mensual de jerarquía de precipitación

i' = índice mensual de jerarquía de temperatura

P = Precipitación promedio mensual (mm)

T = Temperatura promedio mensual(°C)

- Valores medios anuales del índice de Humedad “H”

$$I_h = I_E - 0.6I_D, \text{ donde } I_E = \frac{E}{ETP} * 100, I_D = \frac{D}{ETP} * 100$$

E = excesos mensuales de agua

D = déficits mensuales de agua

EPT = evapotranspiraciones potenciales medias mensuales (mm)

I_E = índice de exceso

I_D = índice de déficit

En base a la simbología mostrada, las clasificaciones climáticas que se encuentran en las cuencas: Chillón y Rímac y Lurín, son las siguientes (SENAMHI, 2002):

- B(i) D'H3 ó (CC0): Zona de clima semifrío, lluvioso, con deficiencia de lluvia en invierno con humedad relativa calificada como húmeda. Corresponde a este tipo climático los siguientes lugares: Surasaca, Marcapomacocha, Milpo, Cerro de Pasco, Pachachaca, Junín, Morococha, Pomacocha y Yauricocha.
- C(o,i,p)B'3H3 ó (CC2): Zona semiseca, semifrío, con deficiencia de lluvias en otoño, invierno y primavera, con humedad relativa calificada como húmeda. Corresponde a este tipo climático los siguientes lugares: Santiago de Chuco, Conchucos, Corongo, Bambamarca, Huaraz.
- C(o,i,p)B'2H3 ó (CC3): Zona semiseca, templada, con deficiencia de lluvias en otoño, invierno y primavera, con humedad relativa calificada como húmeda. Corresponde a este tipo climático los siguientes lugares: Huancabamba, Huambos, Santa Cruz, Cajamarca, Contumaza, Otuzco, Caraz, Canta, Matucana.
- E(d)B'1H3 ó (CC4): Zona desértica semicálida, con deficiencia de lluvias en todas las estaciones del año, y con humedad relativa calificada como húmeda. Corresponde a este tipo climático los siguientes lugares: Talara, Bayóvar, Jayanca, Ferreñafe, Tinajones, Lambayeque, Chiclayo, Pimentel, Cayaltí, Casa Grande, Cartavio, Jequetepeque, Laredo Trujillo, Paramonga, Huacho, Lomas de Lachay, Huaral, Lima Metropolitana, Cañete, Pampa de Majes, Vitor, La Joya, Moquegua, Pisco, Ocucaje, Palpa y Tacna.

2.2 Variabilidad de precipitación en la vertiente occidental del Perú

La precipitación a lo largo de la vertiente del pacífico se caracteriza por su variabilidad espacial y estacional, esta variación está relacionada con su ubicación meridional y por su influencia con la topografía de la Cordillera de los Andes (Waylen y Poveda 2002; Garreaud *et a.* 2009) que sobretodo en latitudes menores permite que los vientos alisios, provenientes del Atlántico Norte que fluyen hacia el oeste sean bloqueados por los Andes orientales del Perú, generando un máximo climatológico de precipitación en el mes de enero (Garreaud, 2001).

En ese sentido, la costa del pacífico del Perú, es un su mayoría, una “Zona seca” que solo episódicamente experimenta eventos de lluvia (Rau *et al.* 2016) muy fuertes como por ejemplo durante eventos “El Niño”. La usencia de las lluvias en la costa peruana esta atribuida principalmente a la corriente peruana (reduce la evaporación proveniente del océano pacífico) y a la influencia de la Cordillera de los Andes (actúa como barrera frente a los vientos alisios cargados de humedad proveniente del atlántico, que al elevarse pasa por niveles atmosféricos de menor presión), los vientos alisios al dilatase se enfrían por lo que tienen menor capacidad para retener vapor de agua, por lo que no cae lluvia en niveles de menor presión. Posteriormente, el aire se satura y disminuye considerablemente su temperatura, por lo que el vapor de agua se condensa y pasa del estado gaseoso al líquido, generándose la lluvia al lado oriental de la cordillera, mientras que algunas nubes que logran pasar la cordillera al lado occidental se internan mar adentro (Pareja, 1936).

A pesar de su predominante sequedad, las temporadas de lluvias sobre esta región también se pueden desarrollar debido a un ligero debilitamiento del anticiclón del pacífico sureste y el desplazamiento hacia el sur la zona de convergencia intertropical o ZCIT (Lavado *et al.* 2012). Además, en las zonas de mayor altitud, durante el “Invierno Altiplánico”, la precipitación es básicamente originada por tormentas convectivas que se desarrollan durante la tarde y primeras horas de la noche. Los días con actividad convectiva tienden normalmente a agruparse en secuencias de alrededor de una semana (episodios lluviosos), separados por secuencias de días sin convección y de similar duración (episodios secos) (Garreaud, 2001).

Diversos estudios sobre la precipitación a nivel interanual han demostrado que esta variable también puede ser regionalizada, un primer trabajo de Rau *et al.* (2016) para el periodo 1964-2011, que consistió en un procedimiento iterativo a dos etapas, empleando dos métodos clásicos: la claustrización particional k-means (Hartigan y Wong, 1979) con su evaluación por perfiles; y el vector regional (Brunet-Moret, 1979) con su evaluación por índices estadísticos RVM. Esta combinación hizo posible asimilar el comportamiento no-estacionario y no-Gaussiano de las series de precipitación en las 145 estaciones analizadas. La Figura 1, muestra las series interanuales de las nueve regiones marcadas por un gradiente latitudinal y altitudinal. En la Figura 2, estas regiones también representan a nueve zonas climáticamente homogéneas, donde la región 4 o R4 es la más seca del país con periodo húmedo de enero a marzo que representa cerca del 75 por ciento de la lluvia anual, por otro lado la región 6 o R6 muestra un patrón de lluvia heterogéneo con una temporada de lluvia de diciembre a marzo que representa el 85 por ciento del total anual (Rau *et al.* 2016). En este estudio se encontró que la variabilidad de la lluvia es más alta en las latitudes del norte y disminuye hacia el sur.

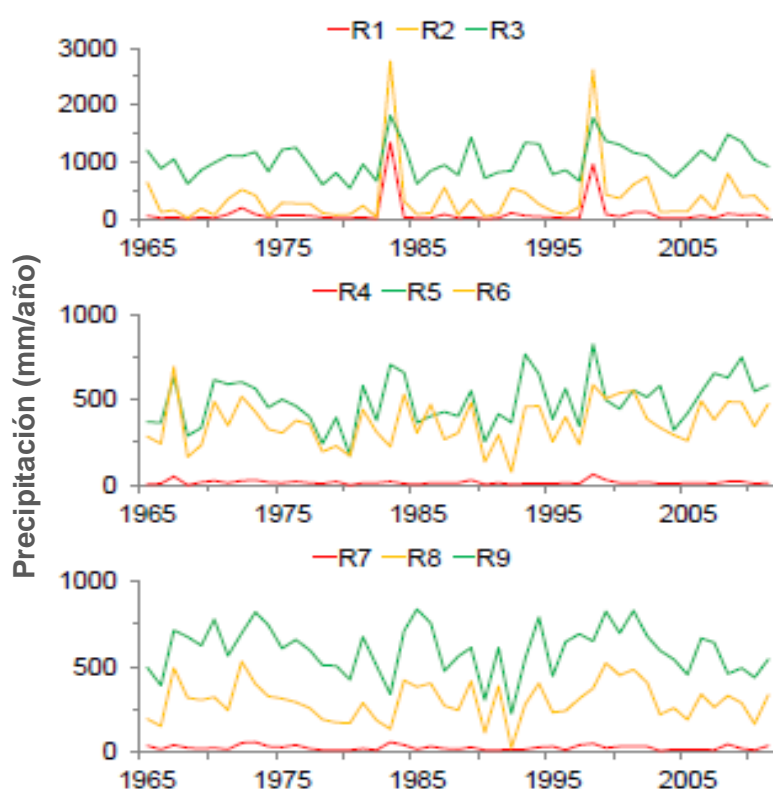


Figura 1: Precipitación para el periodo 1964-2011 en las 9 regiones climáticamente homogéneas

FUENTE: Rau *et al.* 2016

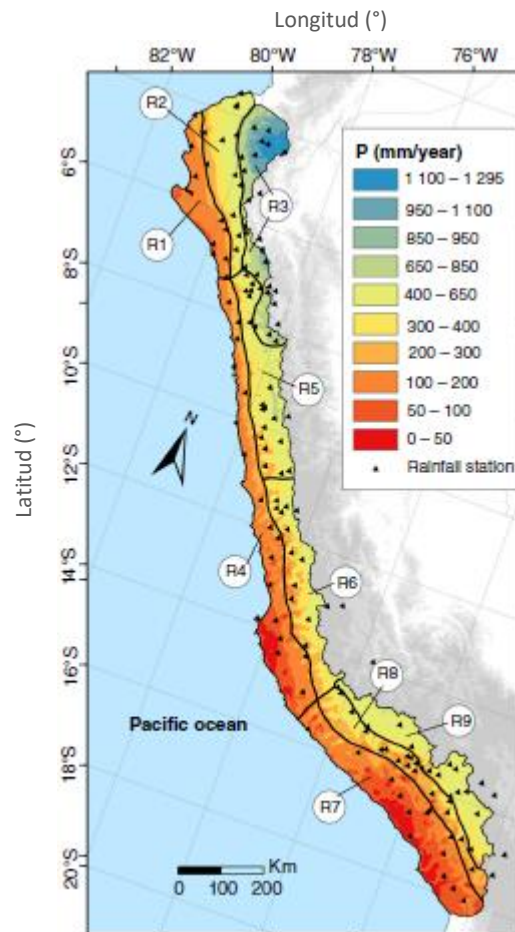


Figura 2: Las nueve regiones de lluvias homogéneas después de la regionalización. Proceso de agrupamiento y RVM(método del vector regional). Superficie interpolada de precipitación anual.

FUENTE: Rau *et al.* 2016

Así mismo, a una escala espacial más pequeña, la precipitación generada sobre la parte baja de las cuencas Chillón, Rímac y Lurín, se atribuye a la influencia del Anticiclón del Pacífico Sur (APSO) que genera un clima semicálido y desértico con escasas precipitaciones durante el año, con incrementos en su intensidad hacia el este. En algunas ocasiones existen alteraciones térmicas sobre la región sur de Lima, que incrementa la humedad y por consecuencia las lloviznas y neblina. En la parte media y alta de las cuencas predomina un clima húmedo y con precipitaciones en verano y ausencia de ésta en invierno. Además, los patrones de humedad y precipitación, están determinados por el ingreso de vientos cálidos y húmedos provenientes de la Amazonía hacia los Andes, así como la influencia de la topografía que influye en la descarga de humedad por medio de la precipitación provocando el desecamiento parcial de la atmósfera, este es causante de que disminuya gradualmente la precipitación en dirección hacia el oeste (SENAMHI, 2016).

La precipitación promedio anual en la cuenca del río Rímac es de alrededor 647.6 mm/año con oscilaciones de 525.3 a 874.6 mm/año (ANA, 2010). En el caso de la cuenca del río Chillón, la precipitación media anual es 450 mm/año (ANA, 2003). Por otro lado, las precipitaciones en la cuenca Lurín varían en función a su elevación, en la parte baja la precipitación puede ser cero durante todo el año y en la parte alta hasta 650 mm por año (ANA, 2004).

2.2.1 “El Niño Costero 2017” y sus efectos en la precipitación

Las lluvias que se producen durante el ENSO (El Niño–Southern Oscillation, por sus siglas en inglés) sobre la vertiente del pacífico generalmente a estado asociado con fuertes eventos que modifican las condiciones áridas y semiáridas sobre esta región (Woodman y Takahashi 2014). Esto ha ocasionado el interés de la comunidad científica por estudiar los eventos ENSO cada vez más cambiantes que se han producido a lo largo de los años. El evento desarrollado durante el periodo enero a mayo del 2017 a lo largo de las costas de Perú (Costa norte y central) y Ecuador (Costa Sur) (Takahashi, 2017), presentó ciertas particularidades. En principio, la denominación “El Niño Costero 2017” estuvo en la mira, ya que en Ecuador se rechazó el término. Por otro lado, investigadores internacionales y peruanos no detectaron su desarrollo en el pacífico, a pesar del monitoreo, ya que se desarrolló rápidamente. De acuerdo con el índice Niño Oceánico (ONI, por sus siglas en inglés) que se monitorea en la región 3.4 (5°N-5°S, 120°-170°W) mediante anomalías de la temperatura superficial del mar o TSM, cuando se desarrolla un fenómeno “El Niño” se presentan anomalías de + 0.5°C de la TSM durante cinco periodos consecutivos a lo largo de tres meses, lo que implica calentamiento en el pacífico central, sin embargo según el monitoreo de ONI, las condiciones durante “El Niño Costero 2017” estaban en transición a la fase neutra. Por otro lado, el monitoreo de la TSM en la región 1+2(90°W-80°W, 10°S-0°) del índice ICEN (Índice Costero El Niño) detectó el evento pero en intensidad débil (Ramírez y Briones, 2017).

El comportamiento del sistema océano-atmósfera durante el evento de “El Niño costero 2017” fue diferente a la dinámica común de ENSO. De acuerdo al análisis realizado por Rodríguez –Morata *et al.* (2018) propone que la causa de las anomalías en la precipitación durante el evento del 2017, estuvo vinculado a la presencia de un anticiclón profundo

ubicado junto a la costa chilena, sumado al desarrollo inusual de aguas cálidas frente a las costa del Perú en la región Niño 1+2. Este último, relacionado con un debilitamiento anómalo del flujo de vientos del oeste subtropical en nivel medio-superior de la tropósfera, lo que a su vez condujo a un debilitamiento del transporte de vientos del sureste en la costa, que pudo haber generado el calentamiento de la TSM en el pacífico oriental que impidió el afloramiento natural en la costa peruana (Garreaud, 2018). En la Figura 3, se observan las anomalías de la TSM, en (a) diciembre del 2016 existen anomalías negativas en el centro-este del Pacífico, en (b) enero 2017 anomalías negativas en la región 1+2, en (c y d) las anomalías presentaban sus valores máximos (Rodríguez-Morata *et al.* 2018).

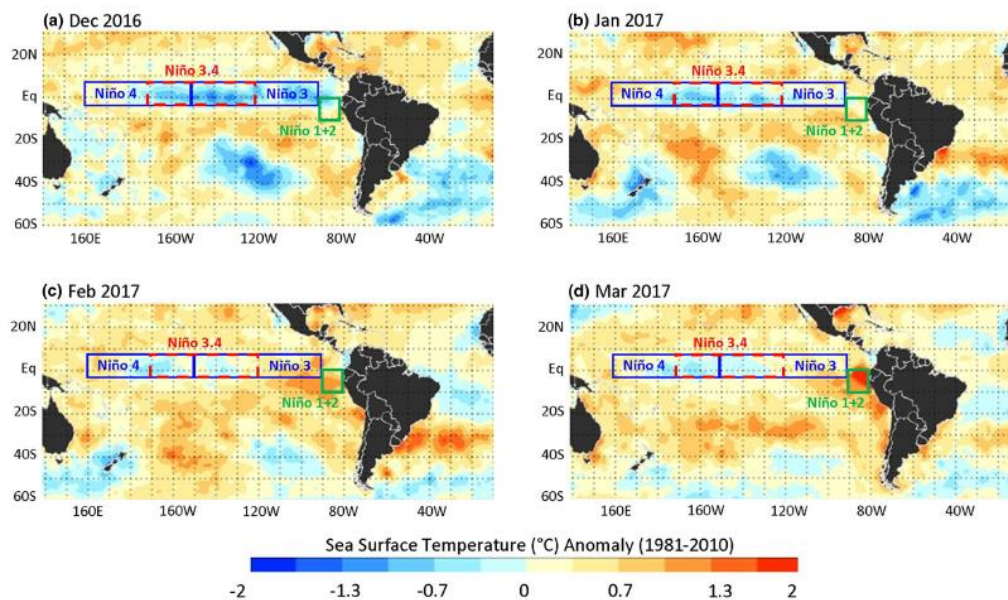


Figura 3: Evolución de la anomalía mensual de la TSM en el Pacífico ecuatorial entre diciembre 2016 y marzo 2017, utilizando datos proporcionados por NOAA/OAR/ESRL PSD

FUENTE: Rodríguez-Morata *et al.* 2018

2.3 Percepción remota de la precipitación

El sensoramiento remoto obtiene información sobre un objeto mediante el análisis de datos adquiridos por un dispositivo (sensor) que no está en contacto con el objeto (remoto) (Lillesand y Keifer, 1994). De esta manera midiendo las ondas electromagnéticas en dominios de los rayos gamma a las microondas, podemos recuperar información sobre los objetos que queremos estudiar. Los sensores que son sensibles a diferentes bandas de longitud de onda en el espectro electromagnético nos dan diferente información sobre la Tierra y su atmósfera. Ceccato y Dinku (2010) indican las longitudes de onda de uso común para la lluvia:

- El infrarrojo termal (longitudes de onda entre 3 y 14 μm)

El sensor infrarrojo termal registra la radiación en longitudes de onda en las que la tierra emite muy fuertemente. La intensidad de la radiación recibida es un buen indicador de la temperatura en la superficie. La imagen infrarroja puede ser vista durante el día y durante la noche. La temperatura de la atmósfera decrece con la altura de una manera regular en toda la tropósfera dependiendo de las condiciones climáticas. La emisión de radiación infrarroja es menor en temperaturas bajas y mayor en temperaturas más altas. Por lo tanto, el sensor de infrarrojos actúa como un termómetro remoto que puede estimar la temperatura de las nubes. Las nubes de tormentas se muestran como muy frías con temperaturas típicamente inferiores a -40°C y algunas veces tan bajas como -80°C . Los cirrus también se identifican como muy frías, las nubes estratiformes o cúmulos inactivas muestran temperaturas más altas entre 0 y -20°C .

- Microondas (longitudes de ondas entre 0.3 a 300 cm ó 3000 y 3000000 μm)

La radiación emitida en longitudes de onda microonda está influenciada por la naturaleza de la superficie emisora y el tamaño de las partículas que atraviesa. Las ondas microondas están fuertemente afectadas por gotas de agua y cristales de hielo en la nube. Este tipo de ondas pueden distinguir entre nubes con gotas lo suficientemente grandes como para producir lluvia y otras nubes. Las frecuencias de MW (ondas microondas) pueden también penetrar las nubes cirrus. Por lo tanto,

son más adecuadas para la detección de lluvia que las frecuencias infrarrojas. Se ha demostrado que las estimaciones de precipitación con microondas tienen una precisión significativamente mejor que las estimaciones de infrarrojos. Sin embargo uno de los problemas que se puede presentar con los sensores microondas es que la resolución espaciales es mucho más gruesa que en otras bandas.

La percepción remota puede dividirse en: Percepción remota pasiva, inferida indirectamente de la radiación infrarroja emitida por las nubes y estimada de la radiación de microondas emitida o dispersada por las partículas de la precipitación, las frecuencias más bajas denominadas “canales de emisiones” miden la precipitación de la energía emitida por las gotas de lluvia y las frecuencias más altas o “canales de dispersión” absorben energía dispersada por partículas de hielo arriba del nivel de congelamiento. También existe la percepción remota activa, estimada de la radiación de microondas retrodispersada transmitida por radares, por ejemplo los satélites de la NASA, TRMM y GPM.

2.4 Datos satelitales GPM y sus productos IMERG

El GPM (Global Precipitation Measurement Mission, por sus siglas en inglés) es una misión designada específicamente para mejorar las mediciones de precipitación, fue construida sobre el éxito del TRMM (Tropical Rainfall Measuring Mission) lanzada por la NASA de los Estados Unidos y JAXA de Japón en 1997, NASA y JAXA exitosamente desarrollaron el observatorio GPM Core o “GPM núcleo” el 28 de febrero del 2014. El observatorio lleva el primer radar espacial de frecuencia-dual DPR, el cual opera en las bandas KuPR y KaPR (13 y 35 GHz, respectivamente) y un sensor de microondas multicanal de escaneo cónico (10-183 GHz), el GMI (Hou *et al.* 2014). Los sensores del GPM extienden el rango de mediciones alcanzados por el TRMM para considerar precipitaciones con intensidades menores a 0.5 mm por hora, detectando también la caída de nieve que representa una fracción significativa de ocurrencia de precipitación en latitudes medias y altas (Mugnai *et al.* 2007; Kulie and Bennartz, 2009).

La misión de satélites GPM esta específicamente diseñada para proporcionar una nueva generación de observaciones de lluvia y nevada, además de mejorar nuestra comprensión del ciclo del agua. Los datos están disponibles en tiempo real (es decir, dentro de las tres horas de observación) para aplicarlos en investigaciones como por ejemplo; previsión de inundaciones, identificación de tormentas, monitoreo de agua dulce, advertencias de deslizamientos de tierra, predicción de cultivos y seguimientos de enfermedades transmitidas por el agua, entre otras. Las mediciones con el GPM están mejorando mediante adquisición de nuevos sensores para medir la precipitación, que trabajan de manera conjunta, de esta forma lograr productos con una mejor resolución temporal y espacial (Huffman *et al.* 2007).

Las estimaciones de precipitación proporcionadas por algoritmos a partir de la información del GPM están caracterizados por:

- Mayor precisión en las estimaciones de precipitación instantánea, particularmente para lluvias ligeras y precipitaciones sólidas en estaciones frías.
- Permite recolectar la información de precipitación de los radiómetros de microondas mediante el uso de temperaturas de brillo intercalibradas y observaciones de hidrometeoros.

El satélite principal de la constelación es el GPM Core o núcleo, el cual se encuentra en órbita no polar de baja inclinación a una altitud de 407 km sobre la superficie. Presenta una cobertura espacial de 16 orbitas al día cubriendo entre las latitudes (65°S-65°N). Junto con la constelación de satélites, tiene un tiempo de revisita de 1-2 horas sobre la tierra y está conformado por los sensores GMI (GPM Sensor de Microondas) y el DPR (radar de precipitación de frecuencia dual). A su vez esta constelación está conformada por 9 satélites: GCOM-W1 (Global Change Observation Mission-JAXA), NOAA18/19, Megha-Tropiques (CNES/ISRO), GPM core Observatory (NASA/JAXA), Suomi NPP (Suomi National Polar-orbiting Partnership-NASA-NOAA), MetOp B/C (Meteorological Operational -EUMESAT), TRMM (Tropical Rainfall Measuring Mission-NASA/JAXA), JPSS-1(Join Polar Satellite System-NOAA) y DMSP (Defense Meteorological Satellite Program-DOP)(Figura 4).



Figura 4: Constelación de satélites que contribuyen con la misión de GPM

FUENTE: Agencia Japonesa de Exploración Aeroespacial (JAXA, 2018)

El sensor GMI fue desarrollado y construido por “The Ball Aerospace and Technology Corporation” sobre un contrato con NASA’s Goddard Space Flight Center (GSFC), es el encargado de captar las imágenes microondas, su cobertura es de -180° a 180° y 65°N a 65°S , presenta un barrido de 885 km y una resolución espacial de 4.4 a 32km (Figura 5). A comparación del sensor del TRMM (TMI) el GMI tiene una resolución espacial más alta (Hou *et al.* 2014).

El sensor DPR fue desarrollado por JAXA y NICT, sus mediciones en tres dimensiones de la estructura de precipitación consisten de una banda (KaPR) y una banda (KuPR) en el radar de precipitación, presenta una cobertura de -180° a 180° , 65°S a 65°N , el barrido de su banda (Ka) es de 120 km y la banda (Ku) 245 km, su resolución espacial es de 5.2 km y resolución vertical 250 metros, este radar opera en las frecuencia 13.6 y 35.5 GHz y a comparación del sensor de precipitación del TRMM (PR) tiene mejor identificación de partículas líquidas de hielo y de precipitación en fase mixta (Hou *et al.* 2014).

La función principal del “Satélite Núcleo” es trabajar como calibrador para todos los datos de observación mediante mediciones simultáneas de instrumentos DPR y GMI, para mejorar la precisión de observación de precipitación del radiómetro de microondas, incluidos los satélites de la constelación (Figura 6).

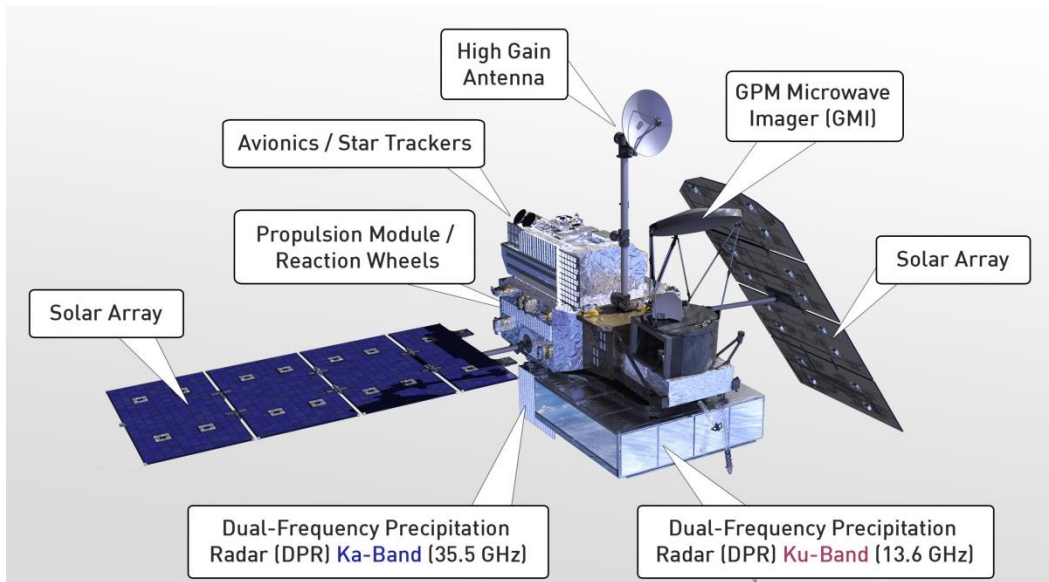


Figura 5: Componentes del satélite GPM-Core o núcleo

FUENTE: NASA

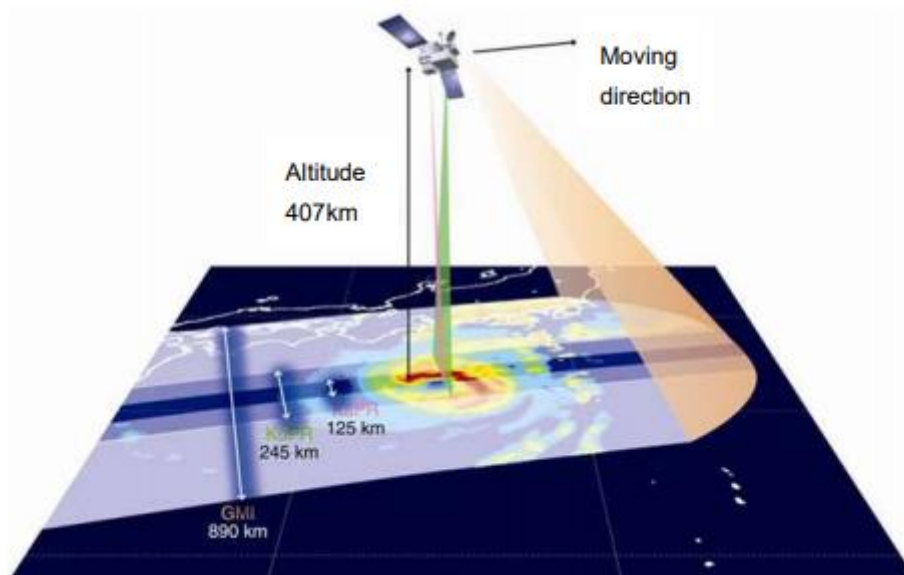


Figura 6: Concepto de observación de precipitación por el observatorio GPM núcleo

FUENTE: Agencia Japonesa de Exploración Aeroespacial (JAXA, 2018)

El Sistema de Procesamiento de Precipitación (PPS) de la NASA, genera un producto con estimaciones de precipitación multi-satélite que pueden ser utilizados para una amplia gama de aplicaciones en tiempo real y áreas de investigación científica. Este producto se llama IMERG (GPM). El cual contiene los datos de precipitación en una cuadrícula de latitud/longitud rectangular de $0.1^{\circ} \times 0.1^{\circ}$ (NASA, 2015). Los datos de IMERG se comenzaron a producir desde el 2014 y los datos en tiempo real se comenzaron a lanzar desde el 2015 (Huffman *et al.* 2014).

El IMERG es el algoritmo unificado de los EE.UU que proporciona el producto de precipitación multi-satélite para el equipo GPM de EE.UU. La precipitación es estimada desde varios sensores de precipitación por satélite con microondas pasivos (PMW), y en donde las precipitaciones que registran los sensores, son calculadas usando la versión 2017 del algoritmo “Perfil de Goddard” GPROF2017, para grillarlos e intercalibrarlos para el GPM y luego combinarlos en el campo de $0.1^{\circ} \times 0.1^{\circ}$ la media horaria. Este algoritmo está destinado a intercalibrar, combinar e interpolar las estimaciones de todos los satélites de precipitación de microondas, junto con las estimaciones de satélites infrarrojos (IR), analizadores de mediciones de precipitación y otros estimadores de precipitación en escalas de tiempo y espacio corto.

El sistema IMERG se ejecuta varias veces para cada tiempo de observación, primero dando una estimación rápida y sucesivamente proporcionando mejores estimaciones a medida que llegan más datos. El producto final o “Final run” utiliza datos mensuales para crear productos de nivel de investigación. IMERG se ejecuta dos veces en tiempo casi real: (1) “Early” producto multi-satélite cuatro horas después del tiempo de observación y (2) “Late” Producto multi-satélite 12 horas después del tiempo de observación. Y una vez que se recibe el análisis mensual de la medición en tierra: “Final”, tres meses después del mes de observación. Los tres recorridos crean productos de media hora a una resolución de $0.1^{\circ} \times 0.1^{\circ}$ (3IMERGHH).

Para facilitar el acceso y la interpretación de los datos GPM, que originalmente fueron distribuidos en formatos HDF4, HDF5, y binario, la NASA generó una nueva versión en formato geotiff que hace que su incorporación a cualquier GIS sea más simple, y se pueda llegar a un mayor número de usuarios de datos. La tasa de precipitación es una cantidad

útil para muchos estudios científicos, los datos GPM almacenan la tasa de precipitación en milímetros por hora; cada tres horas se tiene un producto con los acumulados de a 30 minutos (NASA, 2015). De esta manera, los usuarios en tiempo real siempre tienen la acumulación más actualizada posible con una resolución espacial de 0.1° por píxel (Brito, 2015).

2.5 Espacialización de la precipitación

Según Hu *et al.* (2019), en los últimos años con el desarrollo de ciencias como la meteorología, hidrología y percepción remota se han generado un conjunto técnicas que mejoran la espacialización de la precipitación. La primera de ellas es la interpolación espacial de la precipitación, que es la manera tradicional de transformar la precipitación medida en estaciones y distribuirla en un área, dependiendo de una función matemática, su eficacia está relacionada con la correlación espacial y la densidad de pluviómetros (Qingfang, 2013). La segunda está relacionada con la estimación de precipitación proveniente de la teledetección mediante uso de radares, como el reciente DPR o radar de polarización dual (Zhang *et al.* 2016) y el uso de estimaciones de precipitación por satélite, que abarca coberturas más amplias, además presentan múltiples algoritmos para sus estimaciones (visible e infrarrojo VIS/IR, microondas pasivo PMW, microondas activo AMW y multisensor MPE). El tercero está relacionado con los datos de reanálisis utilizado para registros climáticos. Finalmente, el cuarto está relacionado a la mezcla de fuentes de precipitación, debido a que la red de pluviómetros proporciona información de lluvia puntual y la eficacia de las estimaciones de satélite depende de la variabilidad local, por ello para estimar la precipitación de una manera más acertada se proponen combinaciones o mezclas mediante algoritmos, como por ejemplo el que se basa en la interpolación espacial. En esta investigación se utilizará la técnica de mezcla RIDW o “Regression Inverse Distance Weighting”.

2.6 Métricas estadísticas

Los estadísticos empleados para esta investigación se muestran en resumen en la Tabla 5. Posteriormente, la descripción teórica de cada uno de ellos.

Tabla 5: Cuadro resumen de los estadísticos empleados para el análisis de datos de precipitación

Nombre	Fórmula	Medida Perfecta	Valores	
Coefficiente de correlación de Pearson	$R = \frac{\sum(O - \bar{O})(S - \bar{S})}{\sqrt{\sum(O - \bar{O})^2} \sqrt{\sum(S - \bar{S})^2}}$	1	r (+) Relación directa	r (-) Relación indirecta
Bias relativo	$RB = \left(\frac{\sum(S - O)}{\sum O} \right)$	0	pbias (-) Subestimación	pbias (+) Sobreestimación
Raíz del error cuadrático medio	$RMSE = \sqrt{\frac{\sum(S - O)^2}{N}}$	0	RMSE>0 mayores errores en las estimaciones de precipitación	
Probabilidad de detección	$POD = \frac{H}{(H + M)}$	1	0<POD<0.5 Menor detección de evento de lluvia	0.5<POD<1 mayor detección de evento de lluvia
Ratio de falsa alarma	$FAR = \frac{F}{(H + F)}$	0	0<FAR<0.5 Menor falsas alarmas	0.5<FAR<1 mayor falsas alarmas

FUENTE: Elaboración propia

Coefficiente de correlación de Pearson (R): Es una medida de la relación lineal entre dos variables aleatorias cuantitativas. Está limitado entre -1 y 1; que es, $-1 \leq R_{os} \leq +1$. Si $R_{os} = -1$ hay una perfección por una asociación negativa lineal entre o (dato observado) y s (dato satelital). Similarmente si $R_{os} = +1$ hay una perfección por una asociación positiva lineal (Wilks, 2011). A continuación se muestra la ecuación del coeficiente de correlación de Pearson.

$$R_{os} = \frac{\sum_{i=1}^n (O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2} \sqrt{\sum_{i=1}^n (S_i - \bar{S})^2}}$$

Donde, O =observado, S =satélite, \bar{O} =promedio de observados, \bar{S} =promedio de satelitales.

Bias relativo (RB): También sesgo sistemático, nos proporciona información sobre la tendencia del modelo a sobreestimar o subestimar una variable, por lo que nos cuantifica el error sistemático (Pielke, 1984). El bias es simplemente la relación entre el número de valores predichos y el número de valores verdaderos. Cuando $RB = 0$, nos indica que el evento fue pronosticado igual número de veces que fue observado. Cuando el $RB > 0$, nos indica que el evento fue pronosticado con mayor frecuencia que el observado, lo que se conoce como sobreestimación, en consecuencia un valor de $RB < 0$ nos indica que hay subestimación. A continuación se muestra la ecuación de este estadístico.

$$Relative\ bias(RB) = \frac{\sum_{i=1}^N (S_i - O_i)}{\sum_{i=1}^N O_i}$$

Raíz cuadrada media del error (RMSE): Nos da la medida de las diferencias en promedio entre el valor de precipitación observada O_i y valor estimada por satélite S_i . En otras palabras es la raíz cuadrada de las diferencias individuales entre valores simulados y observados; cuantifica la precisión del modelo (Pielke, 1984). A continuación se muestra la ecuación del estadístico.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N \{O_i - S_i\}^2}$$

Dónde: O_i = valor de precipitación observada, S_i = valor estimada por satélite, N = número de valores analizados.

Tasa de acierto y falsa alarma (POD y FAR): El índice POD o tasa de aciertos, cuenta la proporción de eventos de una clase que fueron pronosticadas como tal. En cambio el índice FAR o ratio de falsa alarma, da la proporción de los pronósticos que fallaron en producirse, esperándose que un buen pronóstico de como resultado valores cercanos a cero (Domínguez y Gonzales, 2012). Se muestran las ecuaciones, donde H representa eventos de precipitación observada y estimadas, F es el número de falsas alarmas, es decir cuando se han observado valores de precipitación que están por debajo del valor límite pero se han

estimado valores de precipitación por encima del valor límite. M es el número de eventos en los que se han estimado precipitación por debajo del límite pero han sido observados por encima del límite, considerando que el valor del límite de precipitación es de 0,5 mm/ tiempo de acumulación.

$$FAR = \frac{F}{H+F}; \quad POD = \frac{H}{H+M}$$

Alguno de estos estadísticos puede ser representado mediante un diagrama de Taylor.

- **Diagrama de Taylor**

Permite resumir los resultados de los estadísticos RMSE, desviación estándar (indica la dispersión de los datos respecto a la media) y correlación, está diseñado para indicar gráficamente cuál de las representaciones o modelos se acerca más a los datos observados (Figura 7).

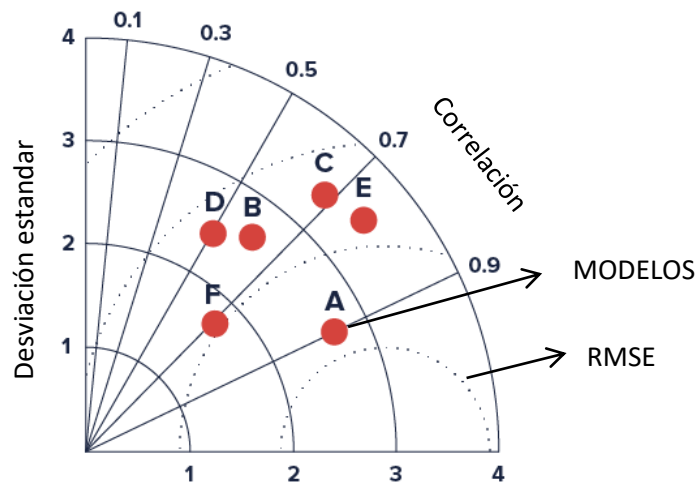


Figura 7: Ejemplo de un diagrama de Taylor

FUENTE: <https://datavizproject.com/data-type/taylor-diagram/>

III. MATERIALES Y MÉTODOS

3.1 Materiales

3.1.1 Zona de estudio

La región Lima se ubica en la región occidental y central del Perú con 34 948 km² y representa cerca del tres por ciento del territorio peruano, limita al norte con Ancash, al este con Huánuco, Pasco y Junín , al sur con Ica y Huancavelica y al oeste con la provincia constitucional del callao y el océano Pacífico, dentro de esta región se encuentra el área de estudio posicionada entre -11.20 y -12.3 de latitud sur y -77.39 y -75.83 de longitud oeste, la zona de estudio está comprendida por dos regiones, costa y sierra, la costa es desértica con pampas áridas y elevaciones de poca altura hasta los 1000 m.s.n.m; y la sierra se caracteriza por su configuración accidentada con estrechas y profundas quebradas donde la Cordillera Occidental de los Andes puede llegar cerca de los 6000 m.s.n.m.

En la zona de estudio se encuentran tres cuencas pertenecientes a la región hidrográfica del pacífico, la altitud en la zona es de 0 a 5000 m.s.n.m. según se puede observar en la Figura 8. Las tres cuencas que se ubican en el área de estudio son: Chillón, Rímac y Lurín.

- Cuenca del Río Rímac

Se ubica en la jurisdicción del departamento de Lima y en menor proporción en el departamento de Junín, enmarcándose en las provincias de Lima, Huarochirí y Yauli, frente a la costa del Perú; Se origina en la vertiente occidental de la Cordillera de los Andes a una altitud máxima de aproximadamente 5508 m.s.n.m.

Las actividades agrícolas y las industriales más importantes del país dependen de sus aguas y se encuentra ubicada entre los paralelos $11^{\circ} 36'52''$ y $12^{\circ} 05'47''$ de latitud sur y $76^{\circ} 11' 05''$ y $77^{\circ} 04'11''$ de longitud oeste (ANA, 2010).

- Cuenca del Río Chillón

Su ubicación abarca el departamento de Lima y provincia del Callao. La cuenca se origina en las inmediaciones del flanco occidental de la cordillera de la Viuda recibiendo sus deshielos. Su ubicación abarca los paralelos $11^{\circ} 20'$ y $12^{\circ} 00'$ latitud sur y $76^{\circ} 20'$ y $77^{\circ} 10'$ de longitud oeste, se caracteriza por presentar precipitaciones medias de 500 mm/año en la zona baja (3000 a 1000 msnm) y junto a la Planicie Costera (0 a 1000 msnm) en el rango de 100 a 200 mm/año. (SENAMHI, 2016).

- Cuenca del Río Lurín

Se encuentra entre los meridianos de $76^{\circ}17'11''$ y $76^{\circ}54'33''$ de longitud oeste y $11^{\circ} 50'31''$ y $12^{\circ}16'34''$ de latitud sur. Es una cuenca que no presenta transvase, se encuentra en la vertiente del pacífico, limita al norte con la cuenca del río Rímac y por el este con la cuenca de Mala, por el sur con la cuenca Chilca y por el suroeste con el Pacífico (ANA, 2004).

Ésta zona fue estudiada con anterioridad por el Servicio Nacional de Meteorología e Hidrología SENAMHI, el cual hizo un estudio de “Análisis regional de frecuencia de precipitación anual para la determinación de mapas de sequías”. Determinando que las zonas menos vulnerables a eventos de sequías con situaciones de deficiencia de lluvia del 80 por ciento son las cuencas altas de Chillón, Rímac y Lurín (Región CHIRILU). Las zonas más vulnerables, son las cuencas medias y bajas con periodos de retorno de 25 a 100 años (Acuña *et al.* 2015).

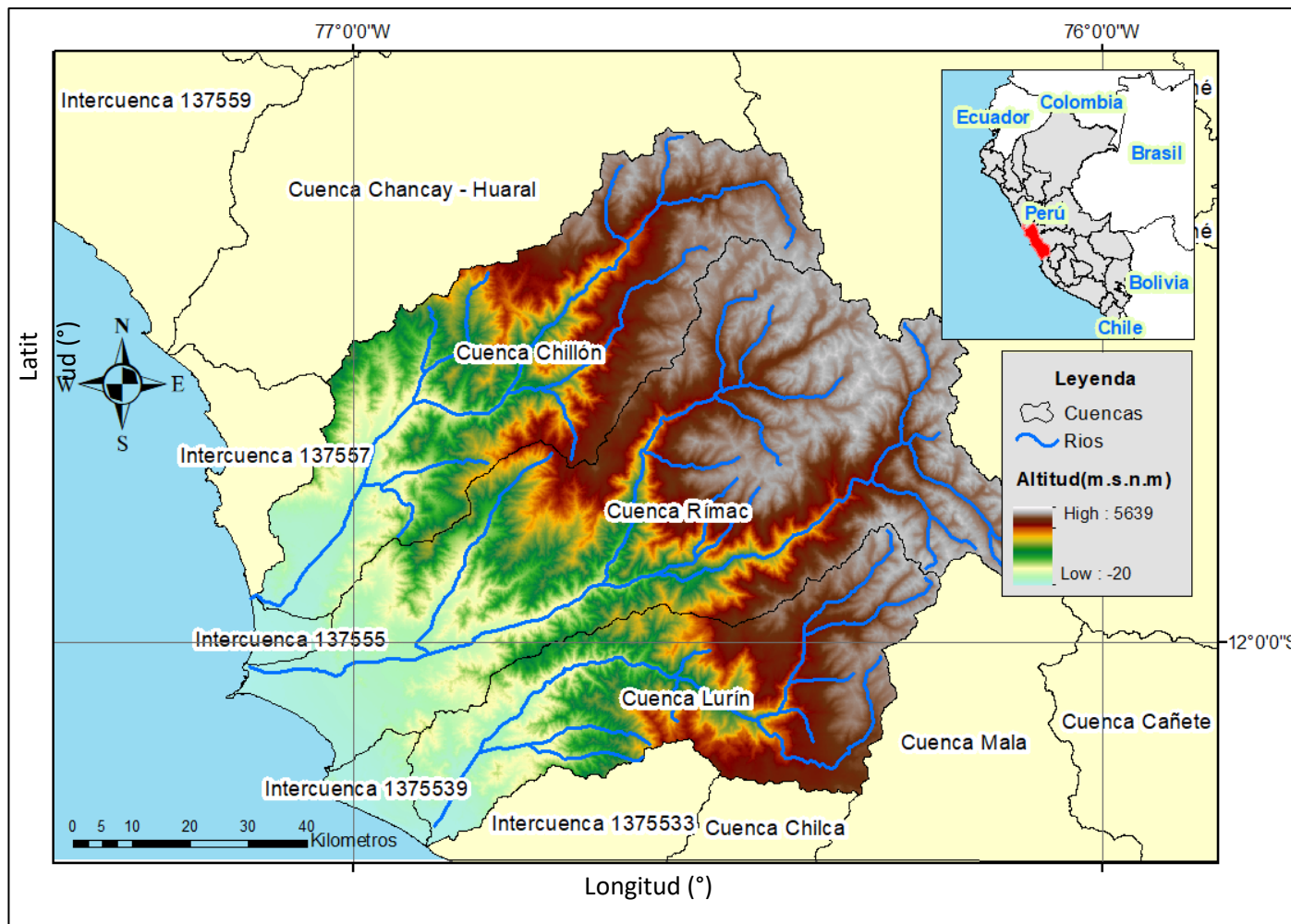


Figura 8: Mapa de la zona de estudio CHIRILU

FUENTE: Elaboración propia

3.1.2 Información observada

- Estaciones convencionales

Los datos de precipitación de estaciones convencionales fueron proporcionados por el Servicio Nacional de Meteorología e Hidrología SENAMHI, se obtuvo información histórica en el periodo del 1 de enero del 2000 al 31 de diciembre del 2018 de 34 estaciones convencionales (Figura 9). Esta data histórica se utilizó solamente para hacer un análisis exploratorio de los datos, para la etapa de validación de productos satelitales se seleccionaron solo algunas estaciones con datos completos en periodos húmedos y secos durante los años 2016, 2017 y 2018. La precipitación registrada por estaciones tipo convencionales es a través de pluviómetros, instrumento que está conformado por un vaso cilíndrico cuya boca de recepción es de 200 centímetros cuadrados para pluviómetros tipo Hellman y 1000 centímetros cuadrados para pluviómetros tipo costa. La precipitación es recolectada en el fondo del recipiente para luego tomar la medición con una probeta graduada en milímetros. Las lecturas con este instrumento se realizan a las horas sinópticas principales. La precipitación registrada en un día se define como la ocurrida entre las 7:00 a.m del día hasta las 7:00 a.m del día siguiente.

- Estaciones automáticas

Los datos de precipitación de estaciones automáticas fueron proporcionados por el Servicio Nacional de Meteorología e Hidrología SENAMHI, se obtuvo información histórica en el periodo del 1 de enero del 2007 al 31 de noviembre del 2018 de 20 estaciones automáticas ubicada en el área de estudio (Figura 9). Los registros en estaciones tipo automáticas es a través de sensores electrónicos que consideran un tipo de sistema de alimentación por intervalos de medición, configurable libremente, entre un segundo y 24 horas en incrementos de un segundo. La información obtenida para este estudio fue horaria, por lo que para tener registros en otros intervalos de tiempo se realizó la acumulación previa en el periodo de tiempo requerido. La precipitación diaria se obtuvo por la acumulación de los registros desde las 7:00 a.m del día a las 7:00 a.m del siguiente día.

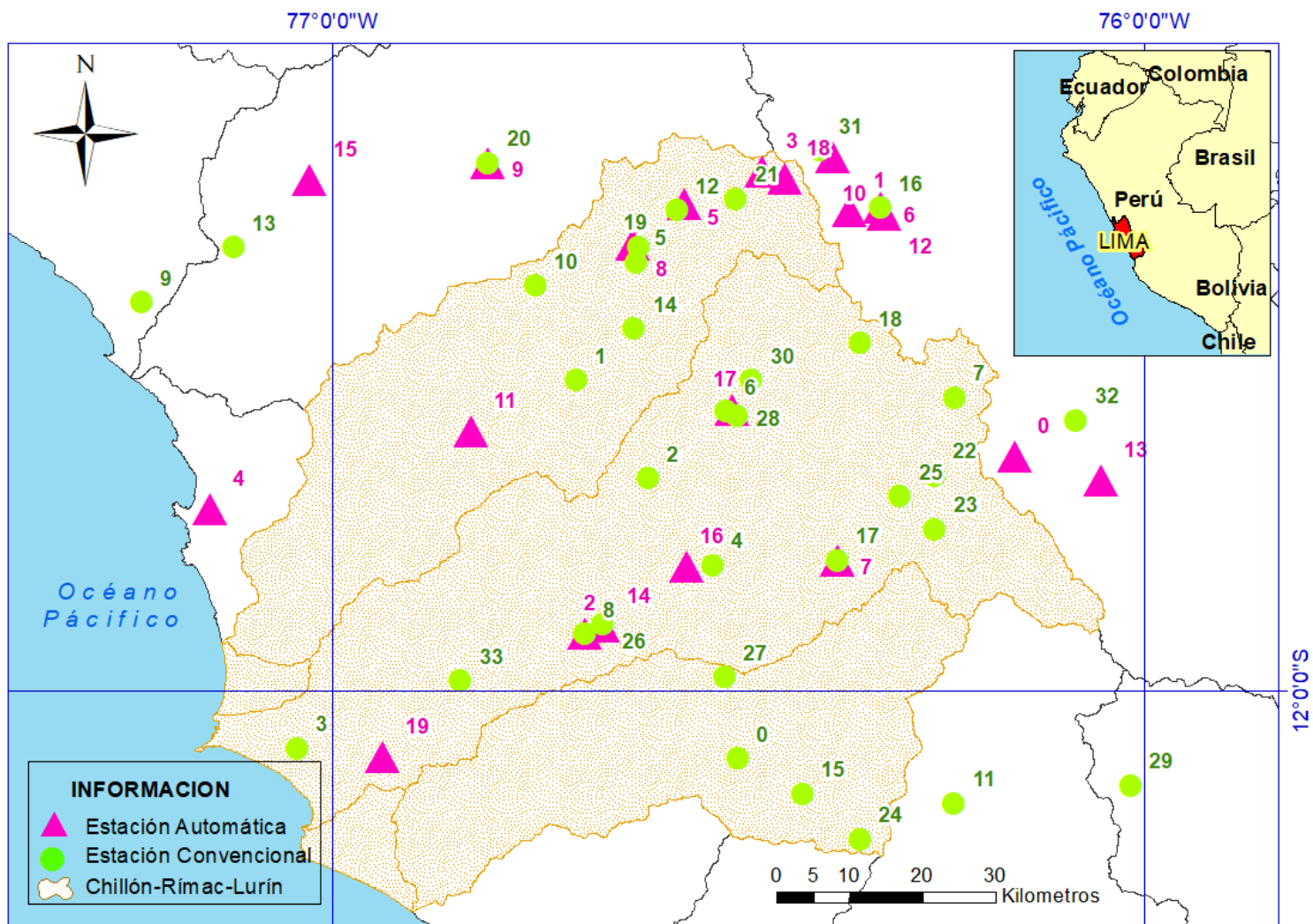


Figura 9: Ubicación de las estaciones pluviométricas convencionales y automáticas en la región CHIRILU

FUENTE: Elaboración propia

3.1.3 Información satelital

En esta investigación se utilizó la información satelital de la misión GPM a través del algoritmo que estima precipitación IMERG en su versión V05, el cual se encarga de intercalibrar, fusionar e interpolar todos los cálculos de precipitación de satélites microondas (PW) junto con las estimaciones de satélites infrarrojos (IR), análisis pluviométricos y otros estimadores a escalas de tiempo y espacio finas. Este sistema se ejecuta varias veces para cada tiempo de observación, primero dando una estimación rápida y sucesivamente proporcionando mejores estimaciones a medida que llegan más datos.

Se trabajó con tres productos de estimación de precipitación proporcionados por IMERG que tienen una resolución espacial de $0.1^{\circ} \times 0.1^{\circ}$ y de resolución temporal de 30 minutos. IMERG es un algoritmo unificado, las estimaciones de precipitación de diversos sensores de microondas pasivo o PMW que se encuentran en los satélites de GPM, se calculan utilizando el algoritmo GPROF2017 para la versión V05 (2018). Estas estimaciones se proporcionan a los algoritmos CMORPH-KF (The Climate Prediction Center Morphing-Kalman Filter) (Joyce *et al.* 2004; Joyce y Xie 2011) del Centro de Predicción Climático CPC y PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) (Sorooshian *et al.* 2000; Hong *et al.* 2004), el CMORPH-KF utiliza las estimaciones PMW e IR para crear estimaciones cada media hora. Más detalle sobre los tres productos utilizados se muestra en la Tabla 6.

Tabla 6: Productos de estimación de precipitación con el algoritmo IMERG

Producto	Tiempo de latencia	Posibles usos
IMERG-Early	4 horas	Usuarios que necesitan información en tiempo real (situaciones de inundaciones o deslizamientos)
IMERG-Late	12 horas	Usuarios que trabajan en la previsión agrícola o vigilancia de sequías
IMERG-Final	2.5 meses	Investigación

FUENTE: Elaboración propia

3.1.4 Equipos

Se emplearon los siguientes equipos de escritorio y programas de cómputo:

Equipos de escritorio:

- Laptop LENOVO, Procesador AMD A9-9420 RADEON R5, memoria RAM de 8Gb.
- Memoria externa Toshiba, 1 Tb
- Materiales de escritorio

Software:

- Lenguaje de programación R – versión 3.5
- Texmaker 5.0.2
- ArcMap – versión 10.24
- Microsoft Office

3.2 Metodología

Para poder cumplir con los objetivos específicos planteados se realizó un resumen que detalla los procedimientos a seguir en el cumplimiento de cada objetivo (Figura 10).

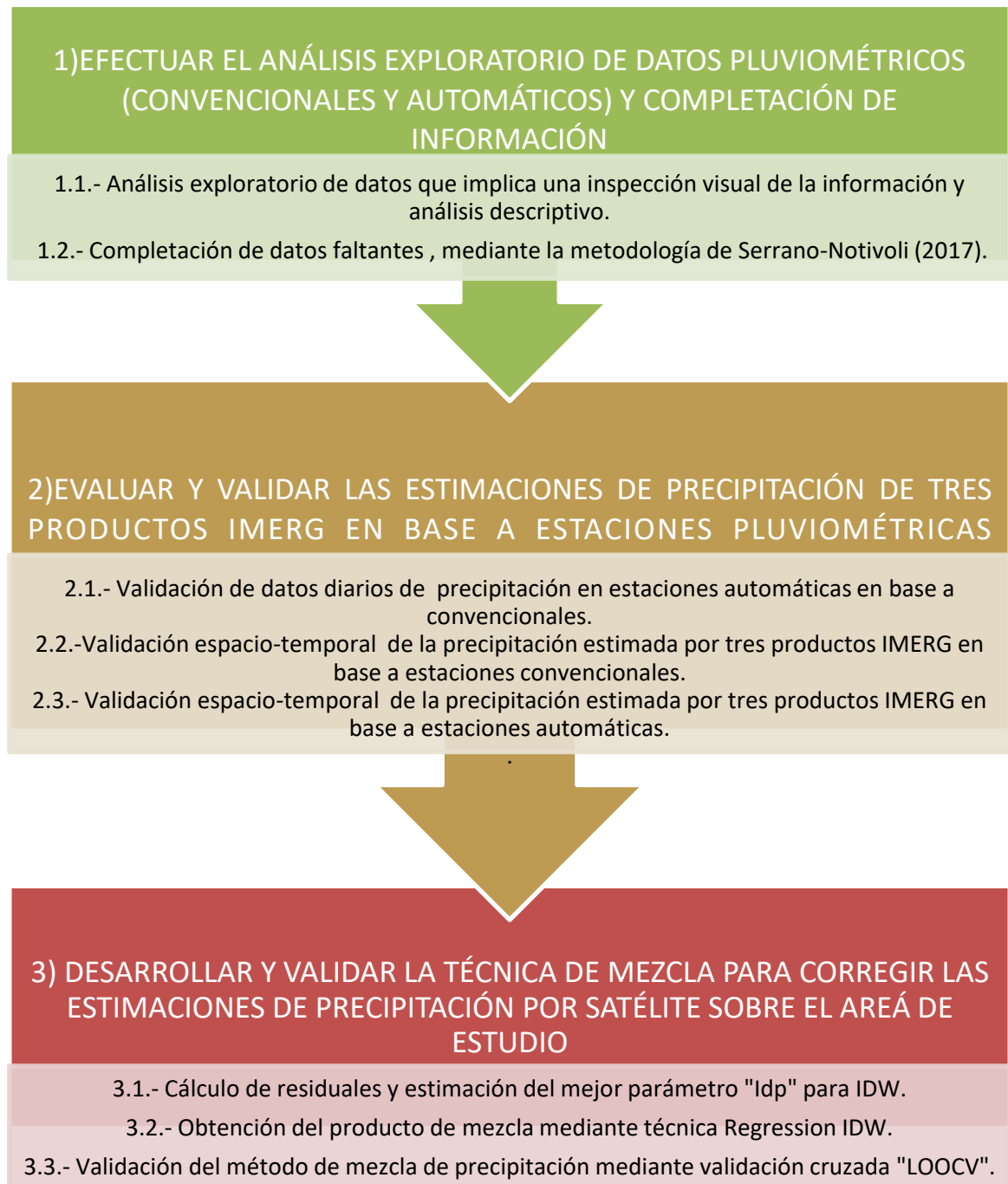


Figura 10: Procedimientos de la metodología empleada.

FUENTE: Elaboración propia

3.2.1 Efectuar el análisis exploratorio de datos pluviométricos (convencionales y automáticos) y completación de la información.

Se entiende como “Análisis Exploratorio de datos o AED” a una técnica que permiten comprender de manera rápida la naturaleza de un conjunto de datos. Fue introducida por Tukey (1977) y se basa en dos criterios: la visualización de datos y las estadísticas de resumen. Para cumplir con el primer objetivo de esta investigación, se realizarán dos pasos. El primer paso en este análisis consistió en evaluar la disponibilidad de información histórica en estaciones convencionales en el periodo 2000-2018 y automáticas 2007-2018, para ello se elaboró un gráfico de datos disponibles por estación y año que permite visualizar aquellas estaciones que cuenta con al menos 95 por ciento de datos diarios en un año. El segundo paso consistió en el análisis de cajas boxplot a nivel mensual, en base al registro histórico. Del análisis anterior se pudo determinar el comportamiento de la precipitación en la zona de estudio, el paso final consistió en utilizar estadísticos descriptivos para el periodo 2016-2018, que permitan descartar estaciones no útiles para el estudio, como aquellas que cuentan con una gran cantidad de datos faltantes o presentan valores atípicos. Las estaciones seleccionadas durante este procedimiento fueron utilizadas en la completación de datos faltantes.

La completación de datos faltantes fue mínima y se realizó solo en algunas estaciones para registros de datos diarios de estaciones pluviométricas convencionales y automáticas, debido a la distribución no uniforme de las estaciones en la zona de estudio, descartar estaciones que tenían algunos días de vacío de información significaba perder información y completar los valores con “0” no era válido para días con precipitación registrada por estaciones vecinas.

Para el procedimiento de completación se utilizó una herramienta denominada librería “reddPrec”, las operaciones matemáticas necesarias para este procedimiento fueron compiladas en código abierto mediante el lenguaje de programación R. Esta librería utiliza un conjunto de funciones para: i) aplicar un control de calidad completo sobre los datos originales de precipitación diaria, marcando datos sospechosos basados en cinco criterios predefinidos, ii) completando los valores faltantes en las series utilizando las 10

observaciones más cercanas para cada día , iii) creando nuevas series y conjuntos de datos grillados (Serrano *et al.* 2017)

Se escogió la herramienta “reddPrec” para la completación de datos faltantes debido a que actualmente existen varios programas que realizan esta función pero son muy pocos los que trabajan a escala diaria. Algunos ejemplos se citan a continuación: HOMER (Mestre *et al.* 2013) solo realiza homogeneización mensual y control de calidad, su adaptación a datos diarios se denomina SPLIDHOM (Mestre *et al.* 2011), sin embargo no incluye métodos para reconstruir datos faltantes, ProClimDB (Stepanek, 2008) aplica procedimientos de homogeneización y completa datos faltantes pero a pesar de ser *software* libre se basa en licencias privadas, otra herramienta muy utilizada es RclimDex realiza un control de calidad detallado mediante EXTRAQC (Aguilar y Prohom, 2011) en datos diarios, pero no llena valores perdidos ni hace reconstrucciones completas.

La librería “reddPrec” crea valores de referencia diarios, denominados (RV) utilizando todos los datos de las estaciones más cercanas, para calcular los RV se hace uso de la regresión logística multivariable o (MLR) el cual los calcula en función a los 10 vecinos más cercanos (NNS) y las variables geográficas y topográficas como covariables. Este método no depende de la longitud de la serie de precipitación y preserva la variabilidad local de la distribución de la precipitación (Serrano *et al.* 2017).

RV se calcula en función de dos valores predichos:

- Una predicción binomial o BP de la probabilidad de que ocurra un día húmedo (usa los 10 vecinos más cercanos).
- Una predicción de magnitud o MP de la precipitación.

Al final RV es determinado por la combinación de MP y BP, utilizando un valor mayor a 0.5 para determinar un día húmedo.

$$RV = \begin{cases} MP \vee BP \geq 0.5 \\ 0 \vee BP < 0.5 \end{cases}$$

Esta librería utiliza estos RV para el control de calidad detectando datos sospechosos y eliminándolos bajo cinco criterios.

Estos criterios son definidos para situaciones generales pero dependen mucho del tipo de clima en la zona de análisis. En el proceso de llenado de datos faltantes se calculan también los RV y estos pueden ser en función a los datos originales o en función a datos que han pasado por el control de calidad. Además no se usan las mismas 10 estaciones todos los días porque se seleccionan los 10 NNS disponibles para ese momento.

Para cumplir con el segundo objetivo de esta investigación se empleará la metodología que se explica a continuación.

3.2.2 Evaluar y validar las estimaciones de precipitación de tres productos IMERG en base a estaciones pluviométricas.

Consistió en evaluar y validar la información de precipitación registrada por estaciones de tipo convencionales y automáticas sobre el área de estudio (región CHIRILU), así como la precipitación estimada por el algoritmo IMERG en sus tres productos (IMERG-Early, IMERG-Late e IMERG-Final) para la versión (V05).

El periodo de análisis elegido comprende tres meses representativos de la estación de verano (enero, febrero y marzo) definida en adelante como “EFM- lluvioso” y tres meses representativos de la estación de invierno (junio, julio y agosto) definida en adelante como “JJA-seco”, los años de análisis fueron: 2016, 2017 y 2018. El criterio de elección de estos periodos trimestrales se basó en el comportamiento climático de la precipitación en la zona de estudio, registrándose en “EFM-lluvioso” los valores más altos de precipitación y en “JJA-seco” los valores más bajos.

La herramienta utilizada para la comparación de las diferentes fuentes de registro y estimación de precipitación (observada y satelital, respectivamente) fue el lenguaje de programación R. Se dividió el análisis en tres etapas: 1) Validación de datos diarios de precipitación en estaciones automáticas en base a convencionales, 2) Validación espacio-temporal de la precipitación estimada por tres productos IMERG en base a estaciones

convencionales, 3) Validación espacio-temporal de la precipitación estimada por tres productos IMERG en base a estaciones automáticas. Los detalles de cada una se describen a continuación:

- Validación de datos diarios de precipitación en estaciones automáticas en base a convencionales.

En base a las estaciones pluviométricas seleccionadas en el objetivo anterior (ANEXO3), se recaudó información de precipitación diaria para el periodo de análisis. Cabe mencionar, que el registro en estaciones automáticas es de una hora por lo que para obtener acumulados diarios se debe hacer una suma de los registros horarios entre las 7:00 a.m del día hasta las 7:00 a.m del día siguiente para obtener la precipitación del día. Además la comparación entre estaciones se llevó a cabo por pares de estaciones convenciones y automáticas, por lo que aquellas estaciones que no se encontraban geográficamente cerca (menos de 2 km de distancia), fueron descartadas. Para el análisis de datos se realizó una inspección visual mediante series de tiempo que permitieron observar el comportamiento de la precipitación en cada par de estaciones, así como el empleo de estadísticos (R o coeficiente de correlación de Pearson, RB o bias relativo, RMSE o raíz del error cuadrático medio, POD o probabilidad de acierto y FAR (probabilidad de no acierto), que permitieron cuantificar la relación entre ambos tipos de datos. En el análisis también se consideró la ubicación de los pares de estaciones en los tipos de climas que se encuentran en la región CHIRILU.

- Validación espacio-temporal de la precipitación estimada por tres productos IMERG en base a estaciones convencionales.

Consistió en realizar una validación entre datos de estimación satelital de tres productos IMERG y datos de precipitación diaria de estaciones convencionales ubicadas en la región CHIRILU en el periodo de análisis. Debido a que los datos de precipitación satelital son descargados desde un FTP (protocolo de transferencia de archivos) en formato “HDF5” fue necesario hacer una conversión a formato “TIF” para poder extraer los valores de precipitación en cada coordenada que corresponde

a una estación convencional, a esta metodología se le conoce como extracción punto-píxel. Las estimaciones de precipitación en raster que se obtienen son de acumulados de cada media hora por el algoritmo IMERG, por lo que fue necesario acumular esta precipitación para obtener precipitaciones diarias. El análisis consistió en una inspección visual mediante series de tiempo para evaluar la variabilidad temporal, el cálculo de estadísticos (R, RB, POD y FAR), gráficos boxplot y gráficos de la función CDF (distribución acumulada de probabilidad), para los periodos “EFM- lluvioso” y “JJA-seco” y un análisis de la variabilidad espacial de los estadísticos R o correlación de pearson y RB o sesgo relativo, considerando la clasificación climática en la zona de estudio.

- Validación espacio-temporal de la precipitación estimada por tres productos IMERG en base a estaciones automáticas.

Al igual que en la validación anterior se extrajo los valores de precipitación satelital mediante la técnica “punto-píxel”. Sin embargo debido a que los registros en estas estaciones son horarios, las acumulaciones permitieron contar con cinco intensidades de precipitación de 1 hora (mm/hr), 3 horas (mm/3hr), 6 horas (mm/6hr), 12 horas (mm/12hr) y 1 día (mm/día). El análisis también se llevó a cabo para los periodos “EFM-lluvioso” y “JJA-seco”, realizando series de tiempo, gráficos boxplot de los estadísticos y gráficos de la función CDF para analizar la variabilidad temporal así como gráficos de radar de los estadísticos considerando las clasificaciones climáticas en la zona y las cinco intensidades de precipitación para el análisis espacial.

Algunas consideraciones sobre los estadísticos empleados en los tres puntos mencionados anteriormente se detallan a continuación.

Coefficiente de correlación de pearson (R): para determinar la significancia estadística del coeficiente de correlación, se usa un valor crítico como referencia para conocer si el R calculado entre la precipitación observada versus la precipitación satelital es significativo bajo un grado de significancia (α) cuando se acepta que existe correlación entre dos variables ($H_0: R = 0$, $H_1: R < 0$ ó $R > 0$) (Fallas, 2012). El valor crítico se elige por la

tabla de valores críticos del coeficiente de correlación de pearson, que establece ciertos valores de acuerdo al grado de significancia y el número de datos de la muestra.

La probabilidad del error para esta investigación está determinada con un 5 por ciento, siendo un umbral convencional. El número de muestras “n” depende de la cantidad de datos, debido a que se trabajó con estimaciones diarias de precipitación para tres meses (periodo EFF-lluvioso o JJA-seco) en un año, el valor mínimo de “n” fue 90. Por lo que se establece que el valor de R calculado debe ser mayor a 0.207 para que sea significativo con un nivel de confianza de 95 por ciento.

Sesgo o bias relativo (RB) y RMSE: el sesgo o bias relativo varía desde valores de $-\infty$ a $+\infty$, indicando subestimación o sobreestimación respectivamente, siendo cero el valor perfecto, por lo que los valores que se aproximan a cero son los mejores valores de RB, cabe mencionar que este estadístico es sensible a valores extremos. Con respecto al RMSE no hay un umbral que defina su significancia estadística, ya que depende del rango en el que varíen los datos, por lo que si la amplitud de los datos es grande, se espera valores de RMSE mayores a la unidad, considerando que mientras más se acerca a cero este estadístico, se consigue menor error en las estimaciones de precipitación.

Finalmente, para poder cumplir con el tercer objetivo de esta investigación se empleará la metodología que se explica a continuación.

3.2.3 Desarrollar y validar la técnica de mezcla para corregir las estimaciones de precipitación por satélite sobre el área de estudio.

Consistió en utilizar la información de precipitación de observaciones en tierra junto con las estimaciones proporcionadas por algoritmos satelitales que estiman la precipitación, para corregir las estimaciones de precipitación por satélite de un producto IMERG, para ello se empleó el producto satelital IMERG que mejor se ajuste a los datos observados analizados en la etapa anterior y estaciones observadas.

La elevada variabilidad espacial de la precipitación implica el uso de un método capaz de analizar y reproducir con mayor confiabilidad el patrón espacial de los campos de

precipitación. En ese sentido, un método de interpolación empleado debe ser eficiente en la combinación de los datos registrados por las estaciones y datos de satélite disponible, además el empleo del método no debe implicar un tiempo computacional largo y debe ser capaz de usar pocos recursos computacionales (Foehn *et al.* 2018).

En esta etapa se utilizó el método “Regression based inverse distance weighting (RIDW)” (Joseph y Kang, 2011). Esta metodología pretende combinar los datos observados y de satélite. Para ello aplica un coeficiente de multiplicación a los datos satelitales (rasters), el coeficiente es resultado de una regresión lineal entre los datos observados y satelitales, con la finalidad de obtener una tendencia o satélite corregido al cual se le sumará los residuales obtenidos de la regresión lineal interpolados por IDW (Inversion distance weighting), este sería el producto final de mezcla. El método RIDW fue probado por Song *et al.* (2014), en su estudio indica que una interpolación con IDW aplicada a los residuales supera a la interpolación con Kriging en términos de exactitud de predicción con mucha menos demanda computacional, para datos de precipitación por radar a escala de una hora.

Regression Inverse Distance Weighting (RIDW): utiliza la información de precipitación satelital y la precipitación registrada por estaciones en la zona de estudio para combinarlas y generar un producto mejorado de precipitación. La implementación de esta metodología se realizó en el lenguaje de programación R. Los procedimientos utilizados en base a Foehn *et al.* (2018) se detallan a continuación:

- a) Transformación de los datos de precipitación observada y satelital por normalización, debido a que la distribución de la precipitación no sigue una distribución normal, se le aplicó el logaritmo neperiano a cada información. El vector de mediciones de precipitación en un periodo t está dado por:

$$g(t) = [g(e_1, t), g(e_2, t), \dots, g(e_{N,t})] \dots \dots \dots (1)$$

Donde N es el número de mediciones de pluviómetros disponibles durante el periodo " t " y $e = (x, y)$ las coordenadas espaciales del vector en un punto dado. Por otro lado las estimaciones de precipitación por satélite en las ubicaciones de las estaciones observadas durante un periodo " t " está dado por:

$$s(t) = [s(e_1, t), s(e_2, t), \dots, s(e_N, t)] \dots\dots\dots(2)$$

- b) Aplicar regresión lineal a los datos observados y satelitales para obtener los residuales y los coeficientes de multiplicación. En Geoestadística, un proceso aleatorio $Z(e, t)$ puede ser modelado como la suma de su partes deterministas $m_z(e, t)$, correspondiente al componente promedio o de tendencia, y un componente residual estocástico $\epsilon(e, t)$, que corresponde a las fluctuaciones locales de la tendencia, así que:

$$Z(e, t) = m_z(e, t) + \epsilon(e, t) \quad \forall e \in D \subseteq \mathbb{R}^2; t \in T \subseteq \mathbb{R} \dots\dots\dots(3)$$

Donde "e" es el vector de coordenada espacial de un punto dado y "T" el dominio temporal. Por lo tanto la Ecuación 3 puede ser sobrescrita para la profundidad de precipitación ρ sobre el dominio entero como:

$$p(e, t) = m_p(e, t) + \epsilon(e, t) \dots\dots\dots(4)$$

La tendencia $m_p(e, t)$ para un vector de coordenadas espaciales "e" es comúnmente modelada como una función lineal de una variable externa que varía suavemente, el satélite.

$$m_p(e, t) = a(t)s(e, t)$$

Donde $a(t)$ son los coeficientes de regresión y $s(e, t)$ es el vector de valores satelitales en el tiempo "t". Los coeficientes $a(t)$ son la pendiente de una regresión lineal y la intercepción de los pares de puntos compuestos por los valores de estaciones en el "eje Y" y los valores de satélite en ese punto en el "eje X", se asume que $a(t)$ es constante espacialmente. Finalmente para calcular los residuales $\epsilon(e, t)$, la tendencia $m_p(e, t)$ es restada del valor observado en la ubicación de la estación.

$$\epsilon(e, t) = g(e, t) - m_p(e, t) = g(e, t) - a(t)s(e, t) \dots\dots\dots(5)$$

- c) Grillar la información, la precipitación del producto satelital se encuentra originalmente en formato raster, el valor en cada grilla de 10kmx10km es extraído

y ubicado espacialmente como puntos espaciales para el dominio de estudio, por lo que se crea una malla con todas las cuadrículas que corresponden al área de estudio.

- d) Estimar el mejor parámetro para la interpolación IDW. Para ello se realizará la validación cruzada sobre los residuales, este procedimiento se aplica para un rango de valores entre 0.8 y 3.5 cada 0.1 considerando en total 28 valores, cada parámetro se puede elegir minimizando la varianza de estimación en cada ubicación, esto funciona mejor según Babak y Deutsch (2008) que aplicar parámetros constantes. Se calcula el valor de MSE obtenido de la interpolación, escogiendo como mejor parámetro aquel que tenga el valor mínimo de MSE o error cuadrado medio ya que el parámetro “p” o “Idp” controla la suavidad de interpolación y el comportamiento local en el espacio(Aybar *et al.* 2019).
- e) Realizar interpolación IDW a los residuales obtenidos, los residuos $\in (s, t)$ se interpolan considerando el mejor parámetro estimado. La interpolación mediante distancia inversa ponderada o IDW calcula los valores en un punto desconocido a través de la combinación ponderada linealmente de un conjunto de puntos cercanos. Esta interpolación, supone que “las cosas que están más cerca son más parecidas, por lo que tiene más influencia sobre el punto a estimar” (Cañada, 2008). Sin embargo la disminución en la relación espacial entre dos puntos no es simplemente proporcional a la distancia. Como resultado, a menudo se usa una función de potencia o exponencial que modifica el peso de la distancia (Lu y Wong 2007). El cálculo de la interpolación por IDW es la siguiente:

$$\hat{P}(s_0, t) = \frac{\sum_{i=1}^N \lambda_i Z(s_i, t)}{\sum_{i=1}^N \lambda_i}; \lambda_i = \frac{1}{d(s_0, s_i)^p}$$

Donde: $\hat{P}(s_0, t)$ = Es la precipitación estimada en un punto s_0 y tiempo t ,
 $Z(s_i, t)$ = Valor de la muestra en el punto s_i y tiempo t , N= es el número total de observaciones i , p= es el coeficiente de potencia, d= distancia entre la localización de la interpolación y la observación.

f) Obtener el producto final

$$\hat{P}_{RIDW}(e_0, t) = m_p(e_0, t) + \hat{\epsilon}_{RIDW}(e_0, t) \quad \forall e_0 \in D \subseteq \mathbb{R}^2 \dots\dots\dots(6)$$

Después de obtener el producto de mezcla, se validó la técnica RIDW empleada por medio de validación cruzada en base a los datos observados de todas las estaciones empleadas en la técnica RIDW. A continuación se detalla el método:

Validación Cruzada dejando uno afuera (LOOCV): Este método se basó en eliminar un dato de la muestra y con el resto de datos realizar la técnica RIDW (utilizando como parámetro de potencia o “p” en la interpolación IDW, el que ya fue determinado en el método RIDW aplicado con anterioridad para calcular el valor en el punto eliminado, es decir, simular que no existe un punto eliminado y calcularlo de nuevo para compararlo con el dato observado (Tomczak, 1998) a esto se le denomina validación cruzada dejando uno afuera o “Leave-one-out cross-validation” (LOOCV, por sus siglas en inglés). Este procedimiento se repite hasta que cada muestra se haya eliminado a su vez (Figura 11), el rendimiento de la técnica se evalúa aplicando estadísticos cuantitativos como la raíz media de los residuos al cuadrado (Davis, 1987; Song y DePinto, 1995), correlación y sesgo.

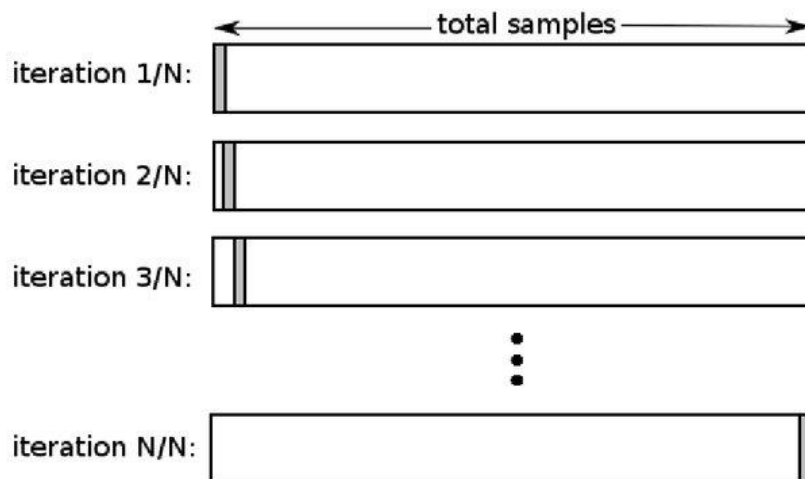


Figura 11: Validación cruzada dejando uno fuera LOOCV.

FUENTE: Nikolaos-kosmas, 2013

IV. RESULTADOS Y DISCUSIÓN

4.1 Efectuar el análisis exploratorio de datos pluviométricos (convencionales y automáticas) y completación de información.

4.1.1 Análisis exploratorio de la información

En la zona de estudio (región CHIRILU) se dispuso de 34 estaciones convencionales y 20 estaciones automáticas ANEXO 1. En la Figura 12 se muestra la disponibilidad de datos diarios de precipitación por estación en el periodo del 2000-2018 (estaciones convencionales) y en la Figura 13 para el periodo 2007-2018 (estaciones automáticas).

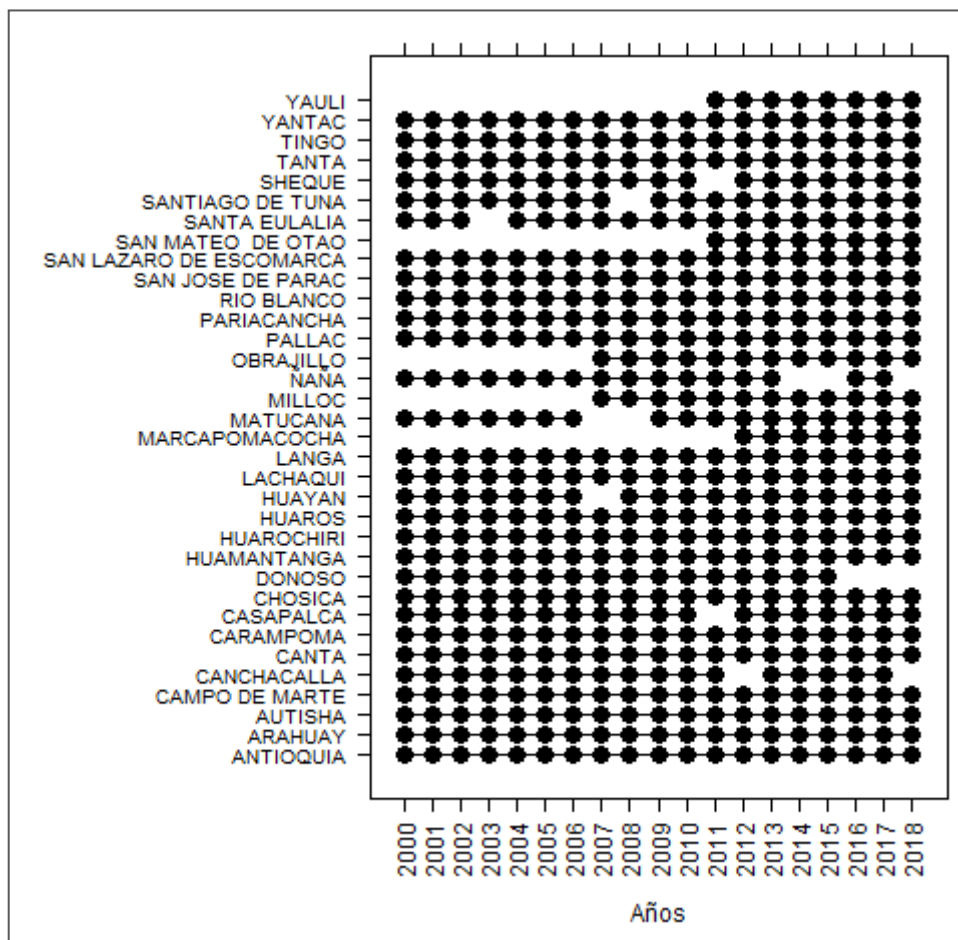


Figura 12: Disponibilidad de estaciones pluviométricas convencional con información para el periodo 2000-2018, los puntos negros indican que al año existieron al menos 95 por ciento de los datos.

FUENTE: Elaboración propia.

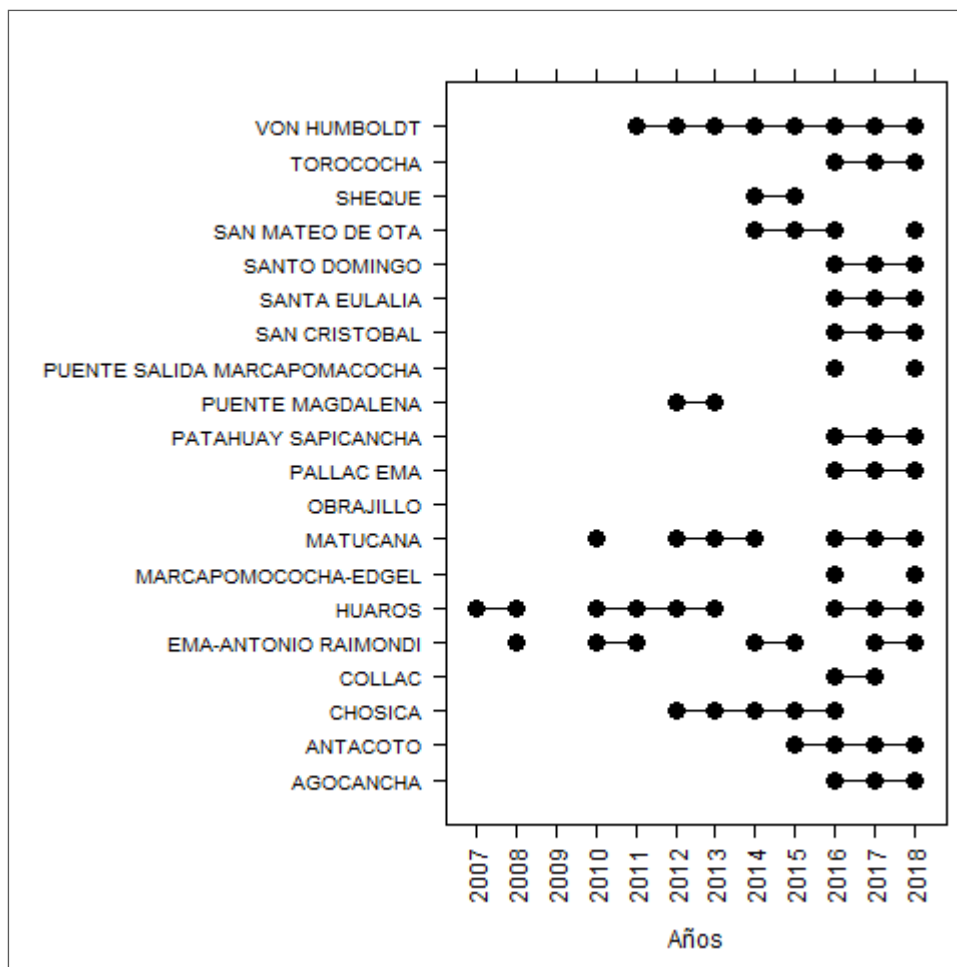


Figura 13: Disponibilidad de estaciones automáticas con información para el periodo 2007-2018, los puntos negros indican que al año existieron al menos 95 por ciento de los datos.

FUENTE: Elaboración propia

De las figuras presentadas se puede resaltar que las estaciones convencionales son las que presentan un amplio registro de datos de precipitación con información diaria, mientras que el registro en estaciones automáticas es notablemente más escaso, considerando también que la información disponible en la mayoría de estas estaciones, es a partir del 2011.

Posterior a la inspección de disponibilidad de información se analizó la variabilidad de la precipitación mensual en cada estación, esto con la finalidad de escoger los periodos de máxima y mínima precipitación (lluvioso y seco, respectivamente). En la Figura 14, se muestra el gráfico boxplot para cuatro estaciones y en el ANEXO 2 la información para todas las estaciones.

Los gráficos boxplot representan los valores promedio mensuales para el periodo de análisis, las cajas representan el 50 por ciento de la distribución de los datos, evidenciando que los trimestres de mayor y menor precipitación son enero, febrero y marzo (periodo lluvioso) y junio, julio y agosto (periodo seco).

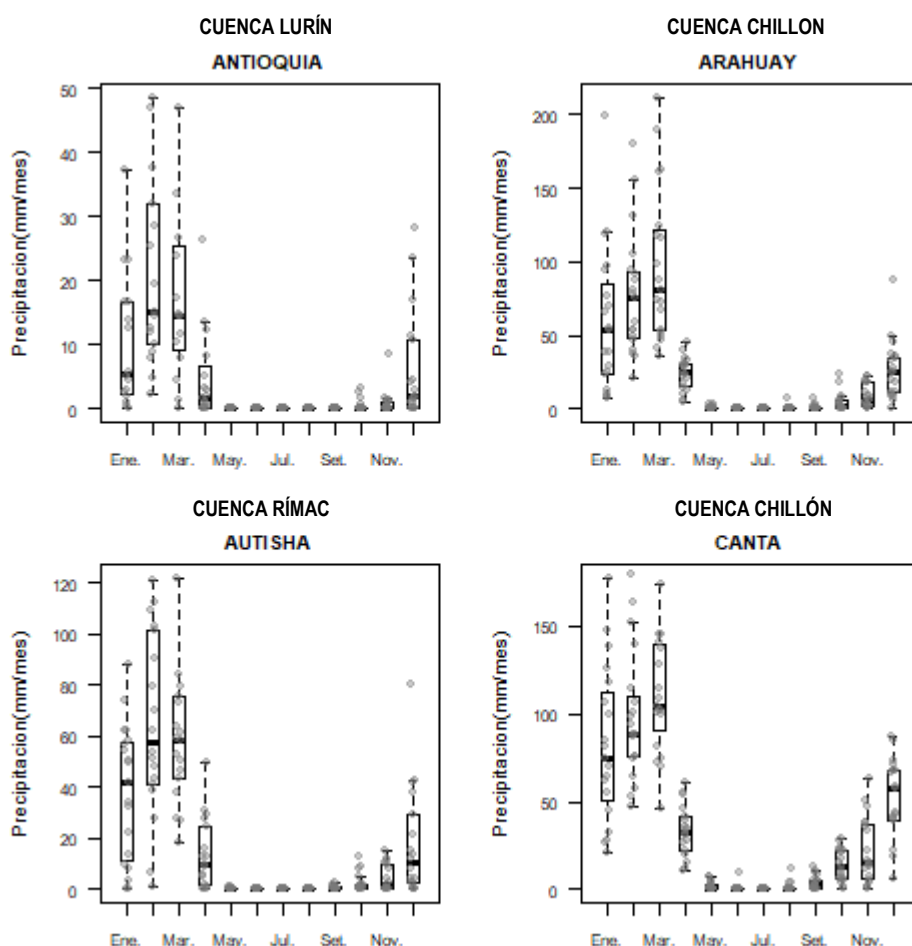


Figura 14: Gráficos boxplot de cuatro estaciones convencionales para acumulados mensuales de precipitación en el periodo 2000-2018.

FUENTE: Elaboración propia

Se realizó también un análisis descriptivo de todas las estaciones involucradas considerando esta vez, solo los datos de precipitación diaria de los periodos húmedos y secos de los años 2016, 2017 y 2018 debido a que la disponibilidad de datos satelitales de productos IMERG para el estudio es en ese periodo.

La Tabla 7, muestra los resultados de los estadísticos descriptivos de las estaciones convencionales para datos diarios, el valor mínimo registrado por todas las estaciones fue

cero, el valor máximo diario fue de 45.1 mm/día en la estación “Huamantanga”, el valor promedio más alto se registró en la estación “Tingo” y fue de 3.24 mm/día, el valor mínimo de desviación estándar se registró en la estación “Campo de marte” , además las estaciones que presentaron mayor cantidad de “NA”, es decir no se registró información, fueron : “Canchacalla”, “Donoso”, “Ñaña” y “Tanta”.

Tabla 7: Estadísticos descriptivos de datos de precipitación diaria en estaciones convencionales.

Estaciones	Mín.	Máx.	Media	Desv.Est	NA(sin dato)
Antioquía	0	9.6	0.575	1.450	12
Arahuay	0	23.4	1.485	3.529	0
Autisha	0	17.4	1.199	2.517	0
Campo de Marte	0	1.9	0.056	0.219	0
Canchacalla	0	36.5	2.661	5.951	184
Canta	0	23.5	1.947	3.919	1
Carampoma	0	31.1	1.966	4.089	0
Casapalca	0	29	2.663	4.526	0
Chosica	0	19.5	0.270	1.535	0
Donoso	0	6.5	0.098	0.509	307
Huamantanga	0	45.1	1.979	4.399	0
Huarocharí	0	33	2.775	5.409	2
Huaros	0	22.80	2.172	4.080	2
Huayan	0	7	0.173	0.489	2
Lachaqui	0	32	2.309	4.383	0
Langa	0	26.9	1.505	3.696	0
Marcapomacocha	0	30.2	3.139	4.979	0
Matucana	0	21.2	1.363	3.032	1
Milloc	0	30.6	2.932	4.979	0
Ñaña	0	5.5	0.395	0.703	59
Obrajillo	0	23.5	1.911	3.957	0
Pallac	0	41.7	1.767	4.228	0
Pariacancha	0	27.0	2.468	4.559	0
Rio blanco	0	27.9	2.263	4.215	8
San José de Parac	0	25.2	2.620	4.538	0
San Lázaro	0	34.7	2.520	5.508	0
San mateo de Otao	0	21.7	1.623	3.508	0
Santa Eulalia	0	21.6	0.300	1.679	0
Santiago de tuna	0	42.8	2.023	4.667	1
Sheque	0	30.5	1.846	3.900	0
Tanta	0	21.8	2.366	4.875	31
Tingo	0	32	3.242	5.484	1
Yantac	0	15.3	3.221	4.128	0
Yauli	0	28.4	2.718	3.932	0

FUENTE: Elaboración propia

La Tabla 8, muestra los resultados de los estadísticos descriptivos para las estaciones automáticas. El valor mínimo registrado por todas las estaciones fue cero , el valor máximo diario fue de 63.1 mm/día en la estación “Puente Magdalena”, el valor promedio más alto se registró en la estación “San Cristóbal” y fue de 3.864 mm/día, el valor mínimo de desviación estándar se registró en la estación “Ema-Antonio Raimondi”, además las estaciones que presentaron mayor cantidad de “NA”, es decir no se registró información, fueron : “Chosica”, “Collac”, “Ema-Antonio Raimondi”, “Obrajillo”, “Marcapomacocha”, “Puente Magdalena”, “Puente salida Marcapomacocha”, “San mateo de Otao” y “Sheque”.

Tabla 8: Estadísticos descriptivos de datos de precipitación diaria en automáticas

Estaciones	Mín.	Máx.	Media	Desv.Est	NA(sin dato)
Agocancha	0	25.6	2.523	4.209	0
Antacoto	0	23.5	2.147	3.852	0
Chosica	0	16.2	0.184	1.141	46
Collac	0	25.2	2.341	4.319	182
Ema Antonio R.	0	2.2	0.014	0.124	74
Huaros	0	21.7	1.585	3.289	2
Marcapomacocha	0	27.0	3.058	5.191	163
Matucana	0	35.6	1.380	3.235	0
Obrajillo	0	25.1	2.732	4.733	344
Pallac Ema	0	33.5	1.471	3.614	0
Patahuay S.	0	22.1	2.555	4.116	0
Puente magdalena	0	63.1	0.655	4.261	298
Puente salida M.	0	22.1	2.447	3.810	32
San Cristóbal	0	33.4	3.864	5.459	1
Santa Eulalia	0	32.0	0.366	2.091	6
Santo domingo	0	13.1	0.104	0.688	0
San mateo de otao	0	21.0	0.990	2.639	183
Sheque	0	17.5	1.849	3.709	446
Torococha	0	23.1	2.622	4.329	8
Von Humboldt	0	5.40	0.104	0.395	2

FUENTE: Elaboración propia

Producto del análisis descriptivo realizado, se clasificaron las estaciones pluviométricas de acuerdo a tres categorías: “Bueno”, “Regular” y “Malo” (Figura 15).

Cuando la cantidad de datos perdidos en una estación es muy grande, el método empleado para la completación de datos puede no ser tan confiable y aumentar el error, es por ello que en esta investigación, se prefirió trabajar solo con las estaciones que no tenían una gran cantidad de valores “NA”. Por lo que finalmente se dispusieron de 30 estaciones convencionales y 11 estaciones automáticas ANEXO 3.

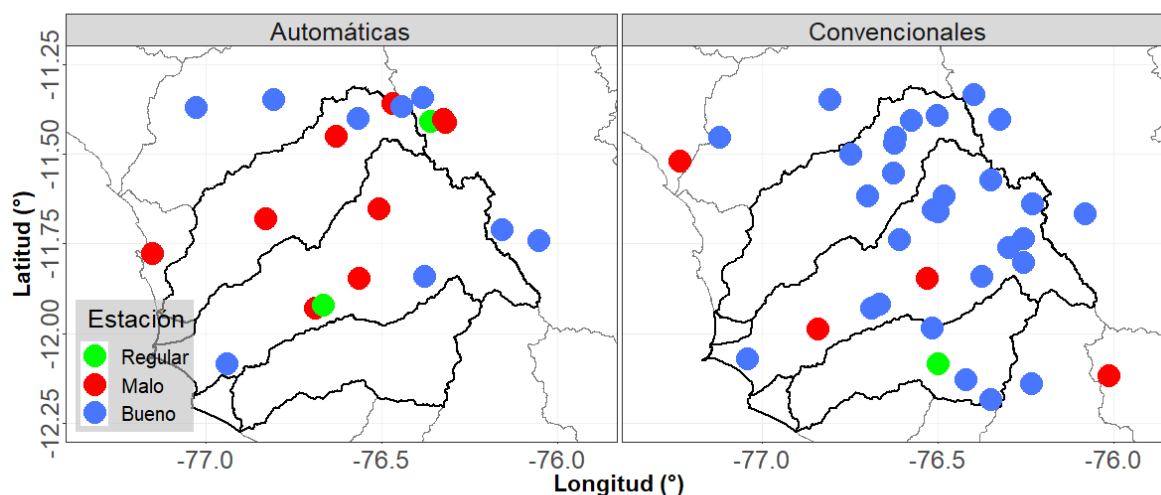


Figura 15: Mapa de estaciones por categorías de estación en la región CHIRILU

FUENTE: Elaboración propia

4.1.2 Completación de datos

El proceso de completación de datos de series de precipitación diaria se realizó mediante el uso de un paquete en el lenguaje de programación R, denominado “reddPrec” que permite trabajar control de calidad, completación de datos y grillado de datos en series de precipitación diarias. Se utilizó la función “gapFilled” para la completación de datos faltantes, para poder ejecutar esta función se requirió contar previamente con un archivo en formato “.xlsx” que contenga las series de tiempo seleccionadas en el análisis exploratorio y un archivo contenedor de la siguiente información: ID o “identificador de las estaciones”, latitud y longitud en coordenadas UTM y altitud en “m.s.n.m”. El archivo con las series de tiempo de precipitación contiene información de 30 estaciones convencionales y 11 estaciones automáticas con datos para los periodos lluviosos y secos del 2016 al 2018. Los resultados del proceso de completación de datos diarios se guardan en un archivo de formato “txt”. En la Figura 16, se muestran las series de tiempo para estaciones convencionales que fueron completadas (9 de 30 estaciones) los segmentos rojos en las series corresponden a los valores resultado del proceso de completación, la estación “Antioquia” fue la que presentaba mayor cantidad de datos faltantes, 2.2 por ciento (12 de 547) y en las estaciones “Canta”, “Matucana” y “Santiago de tuna” y “Tingo” fue de 0.2 por ciento (1 de 547).

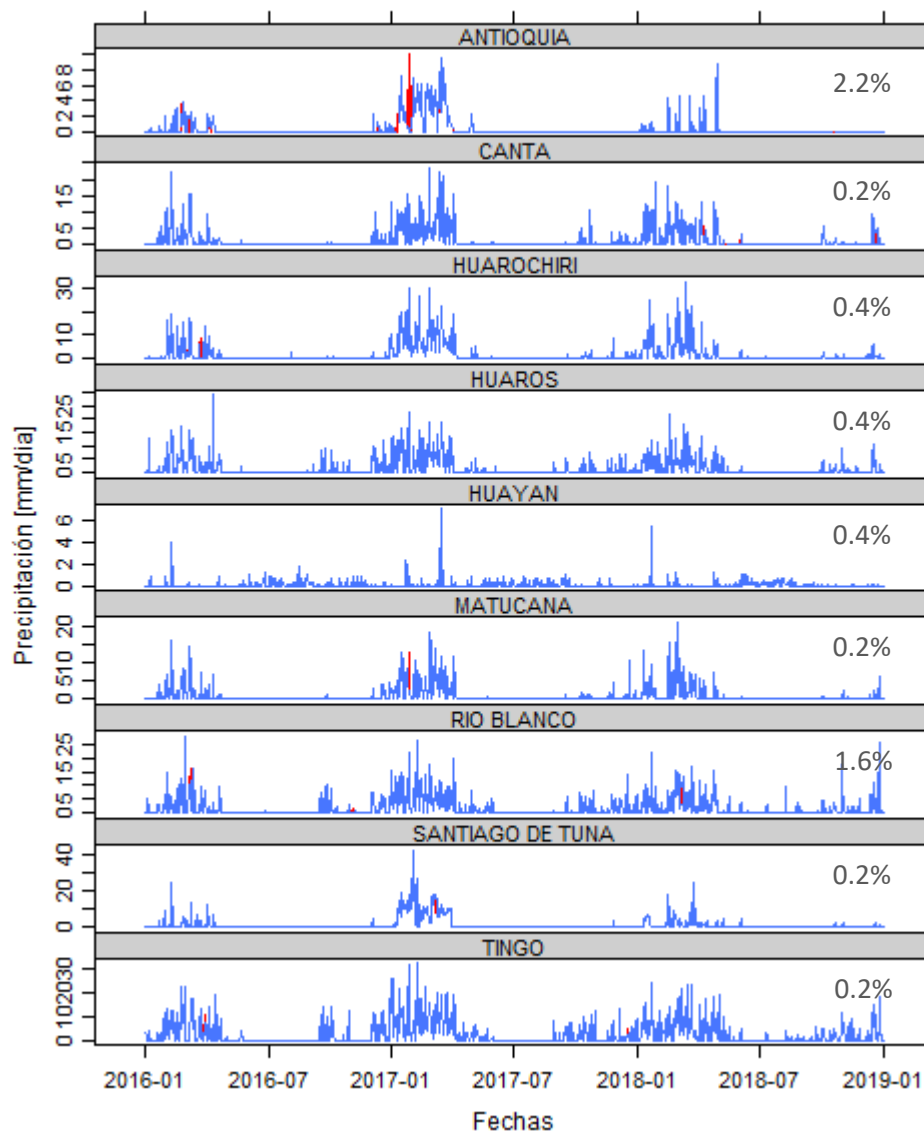


Figura 16: Estaciones pluviométricas convencionales que pasaron el proceso de completación de datos.

FUENTE: Elaboración propia

En la Figura 17, se muestran las series con datos completos para estaciones automáticas, siendo completadas 5 de 11. Las estaciones “Santa Eulalia” y “Torococha” fueron las que presentaban mayor cantidad de datos faltantes (1.1 y 1.5 por ciento, respectivamente), mientras que el resto de estaciones se completaron menos del 0.4 por ciento del total de 547 datos.

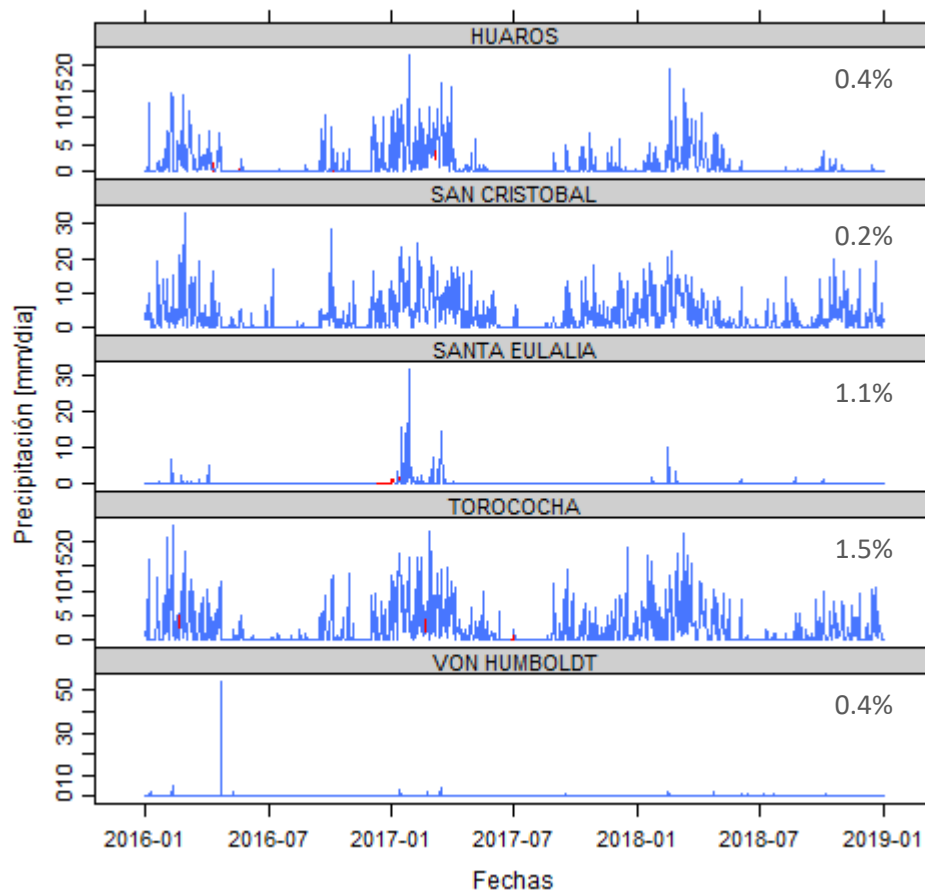


Figura 17: Estaciones pluviométricas automáticas que pasaron el proceso de completación de datos

FUENTE: Elaboración propia

Las estaciones utilizadas en el proceso de completación de datos no fueron alteradas por la función de control de calidad “qcPrec” del paquete “reddPrec”, ya que se analizó previamente esta función con los datos, pero los resultados eliminaban una gran cantidad de datos de las series originales, por lo que se podía perder información. El motivo principal de esta decisión fue que el análisis de calidad de dicha función establece cinco criterios estándar para detectar valores sospechosos, sin embargo tales criterios podrían no ser adecuados para todas las regiones climáticas, por lo que serían necesarios umbrales diferentes en algunos contextos y debido a que no se cuenta con esta información y a que luego del análisis exploratorio no se encontraron valores atípicos extremos, se decidió realizar la completación de datos con las series originales.

4.2 Evaluar y validar las estimaciones de precipitación de tres productos IMERG en base a estaciones pluviométricas.

4.2.1 Validación de datos diarios de precipitación en estaciones automáticas en base a convencionales.

Se realizó la validación de estaciones automáticas que se encuentran en la misma ubicación o cercana a estaciones convencionales. Se dispuso de seis pares de estaciones con información completa para el periodo 2016-2018. Debido a que en la mayoría de los casos, estaciones automáticas y convencionales presentan el mismo nombre de estación, se colocó una sola etiqueta para cada par de estaciones a excepción del par “Yantac-Patahuay Sapicancha” (Figura 18).

En las series de tiempo se observa un comportamiento similar en ambos tipos de estación, el ciclo anual de la precipitación está correctamente caracterizado, los valores registrados en la mayoría de estaciones convencionales para algunas fechas tienden a ser mayores que en estaciones automáticas, sin embargo ocurre lo contrario en el caso de “Santa Eulalia” y “Matucana” a pesar de encontrarse en coordenadas muy cercanas.

Además, se calculó el grado de correspondencia entre ambos tipos de medición, ya que se esperaría que las estaciones automáticas y convencionales al ubicarse en un mismo espacio presenten registros de lluvia muy similares. Este análisis implicó el uso de estadísticos cuantitativos como el coeficiente de correlación de Pearson o R , el sesgo o Bias relativo RB, la raíz del error cuadrático medio o RMSE, así como estadísticos cualitativos como la probabilidad de acierto o POD y probabilidad de falsa alarma FAR. Los resultados se muestran en la Tabla 9, el par de estaciones “Pirca” presentaron el valor más alto de correlación ($R=0.98$), las estaciones “Santa Eulalia” y “Matuca” automáticas sobreestiman al valor registrado por sus pares convencionales, la estación automática “Patahuay Sapicancha” presenta el valor de raíz cuadrático del error medio más alto con respecto a su par convencional. Finalmente, los estadísticos cualitativos indicaron que las estaciones automáticas son capaces de detectar eventos de lluvia y no lluvia con la misma precisión que las estaciones convencionales.

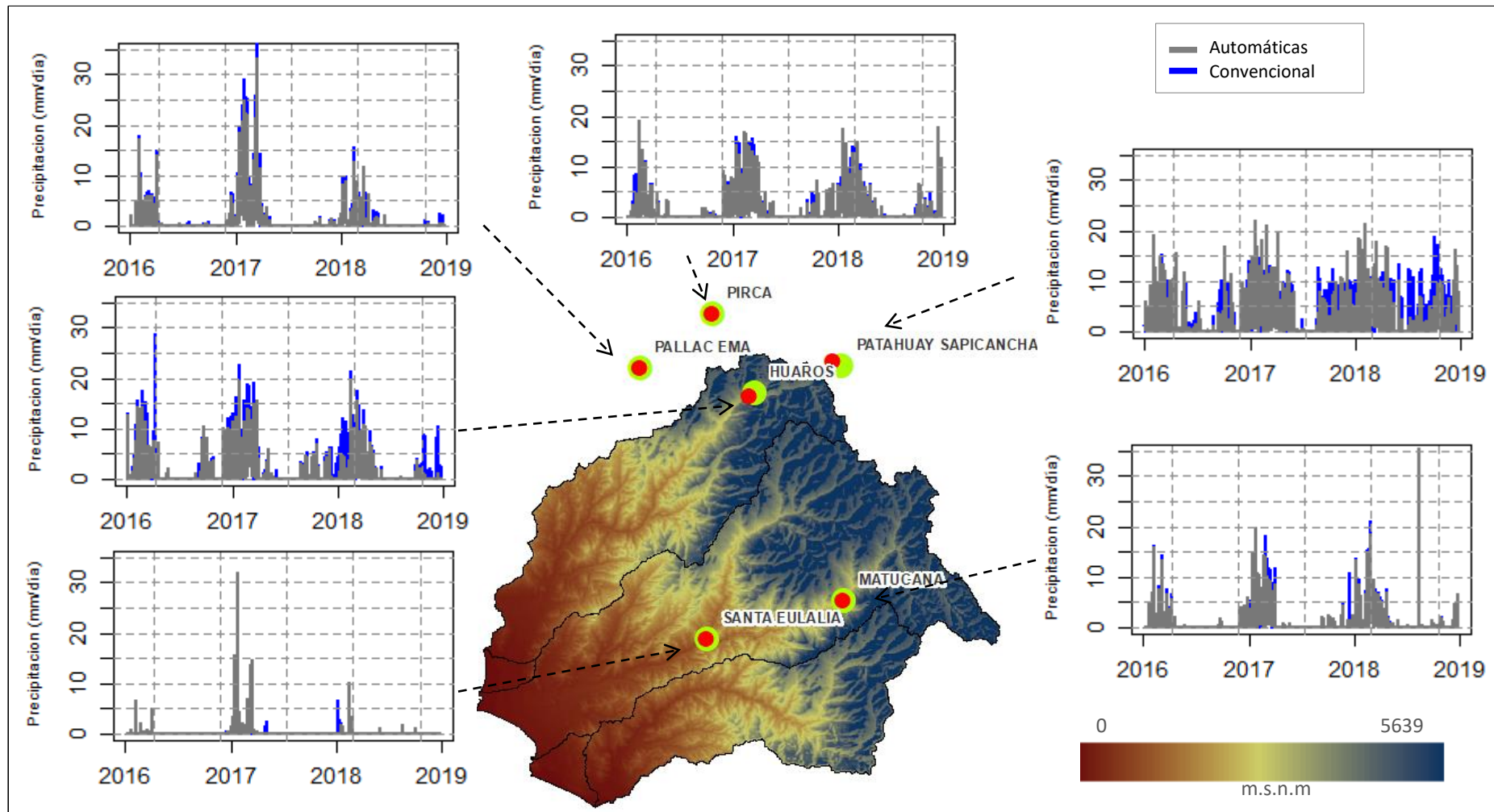


Figura 18: Pares de estaciones pluviométricas convencionales y automáticas en la región de estudio CHIRILU.

FUENTE: Elaboración propia

Tabla 9: Resultado de estadísticos entre pares de estaciones convencionales y automáticas

PAR DE ESTACIONES	R	RB	RMSE	POD	FAR
Huaros	0.81	-0.27	2.04	0.79	0.20
Matucana	0.86	0.01	1.29	0.97	0.22
Santa Eulalia	0.94	0.22	0.54	0.78	0.38
Pallac	0.97	-0.16	0.90	0.90	0.19
Pirca	0.98	-0.08	0.60	0.96	0.07
Yantac-Patahuay	0.58	-0.29	3.72	0.89	0.19

FUENTE: Elaboración propia

La variación de los valores de correlación de Pearson entre los pares de estaciones es explicada por las características climáticas que se encuentran en la región de estudio (Figura 19). Los pares de estaciones que se ubican dentro de una misma clasificación climática, aumentan positivamente la relación entre sus valores (mayor correlación) a medida que se encuentran a mayor altitud, tal es el caso del par “Matucana” que se ubica a 2417 m.s.n.m y “Pirca” a 3342 m.s.n.m (Tabla 10). Las clasificaciones climáticas que se encuentran en la zona de estudio se diferencian también por la elevación, siendo “E(d)B’1H3” o “CC4”, la de menor altitud, por ello el par de estaciones “Santa Eulalia” presenta menor correlación a comparación de las estaciones “Pallac” y “Pirca” que se ubican en clasificaciones climáticas con características más lluviosas, por lo tanto de mayor elevación. Por otro lado la distancia entre los pares de estaciones también tiene influencia en la relación de sus registros de precipitación, por lo que el par “Yantac-Patahuay” que se encuentran a una distancia de 2km aproximadamente, presenta baja correlación y mayor subestimación.

Tabla 10: Altura y clasificación climática entre los pares de estaciones

PARES	Altura Automáticas	Altura Convencional	Clasificación
HUARIOS	3585	3569	C(o,i,p)B’3H3
MATUCANA	2417	2417	C(o,i,p)B’2H3
SANTA EULALIA	970	970	E(d)B’1H3
PALLAC	2367	2367	C(o,i,p)B’2H3
PIRCA	3342	3342	C(o,i,p)B’3H3
PATAHUAY-YANTAC	4535	4617	B(i)D’H3

FUENTE: Elaboración propia

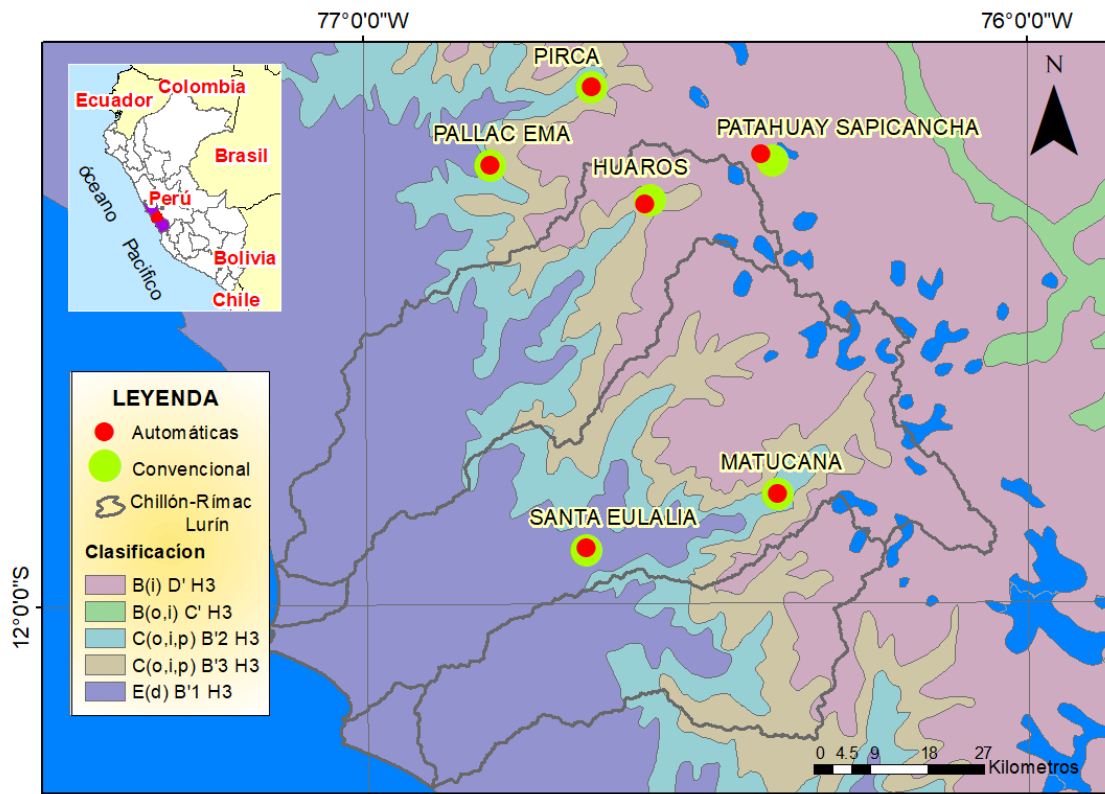


Figura 19: Mapa de pares de estaciones pluviométricas automáticas y convencionales sobre la clasificación climática

FUENTE: Elaboración propia

4.2.2 Validación espacio-temporal de la precipitación estimada por tres productos IMERG en base a estaciones convencionales.

Se realizó un análisis comparativo entre los datos de precipitación en intensidades de (mm/día) registradas por estaciones convencionales (30 puntos de observación) y datos extraídos de tres productos satelitales en tiempo real IMERG-Early, IMERG-Late e IMERG-Final, para dos periodos; enero, febrero y marzo (EFM- lluvioso) y junio, julio y agosto (JJA-seco), los resultados obtenidos se muestran en la Tabla 11, los cuales representan el promedio de los valores estadísticos para todas las estaciones. En la Figura 20, se representa las series de tiempo de las estaciones que obtuvieron los valores más altos de correlación en cada periodo, en la estación “Carampoma” ($r_{IMERG-early} = 0.64, r_{IMERG-late} = 0.64, r_{IMERG-final} = 0.66$) para el periodo lluvioso y en la estación “Langa” ($r_{IMERG-early} = 0.95, r_{IMERG-late} = 0.96, r_{IMERG-final} = 0.96$) en el periodo seco.

Tabla 11: Estadísticos entre los datos de precipitación diaria del producto satelital y el registrado por el pluviómetro convencional, para dos periodos

Producto	EFM - lluvioso					JJA - seco				
	R	RB	RMSE	POD	FAR	R	RB	RMSE	POD	FAR
IMERG-E	0.40	1.47	6.27	0.87	0.29	0.48	16.17	1.33	0.53	0.85
IMERG-L	0.42	1.13	6.12	0.86	0.28	0.50	18.79	1.42	0.52	0.86
IMERG-F	0.42	2.94	9.51	0.88	0.29	0.62	20.08	1.93	0.59	0.86

FUENTE: Elaboración propia

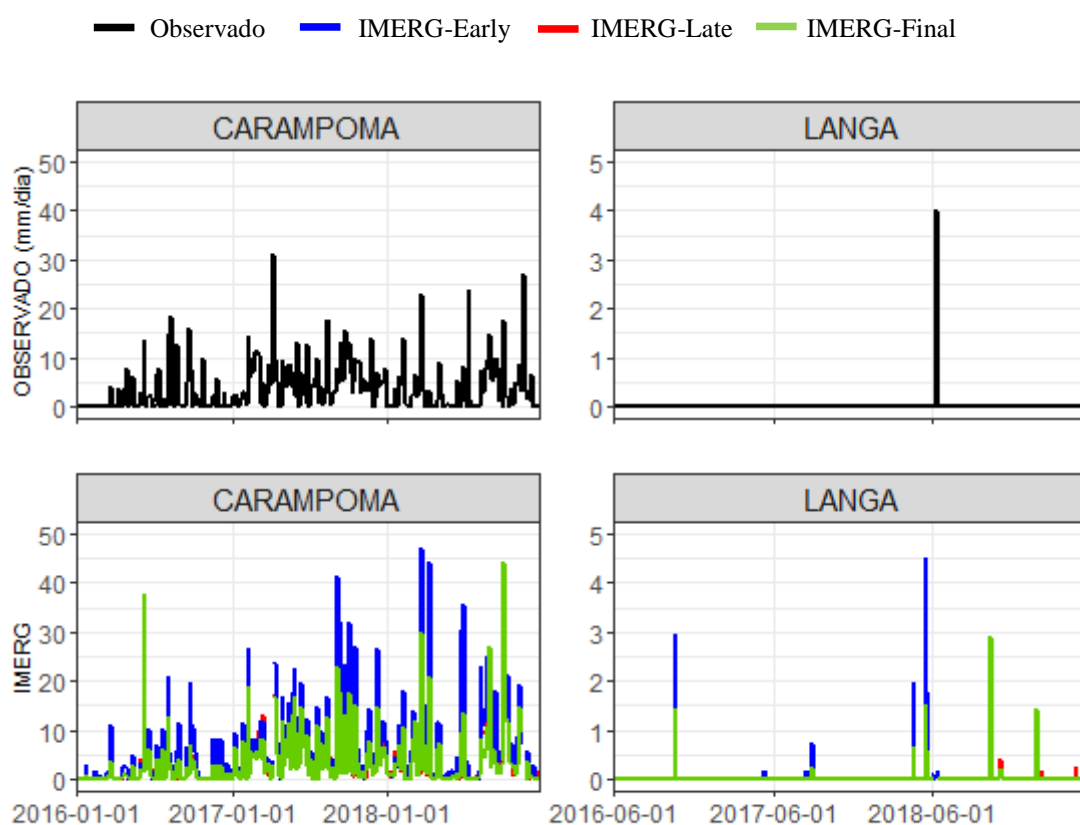


Figura 20: Series de tiempo de las estaciones con mejores estadísticos para la época EFM-lluvioso y JJA-seco en estaciones convencionales.

FUENTE: Elaboración propia

La información completa de series temporales entre datos observados convencionales y los tres productos satelitales, en el ANEXO 4.

Los resultados de los estadísticos cuantitativos (R y RB) y categóricos (POD y FAR) aplicados a cada estación pluviométrica convencional para datos diarios en los periodos EFM-lluvioso y JJA-seco de los años 2016, 2017 y 2018 se presentan en la Figura 21.

Puede observarse en general que existe mayor dispersión para el producto IMERG-Final, principalmente en periodos secos; el que, por otra parte, presenta un valor de R promedio mayor en los dos periodos a comparación de los otros dos productos satelitales, además un valor de sesgo o bias relativo en promedio mayor, por lo que se convierte en el producto satelital que más sobreestima la precipitación. Por otro lado, la tasa de cierto de presentarse un evento de lluvia es alta en los tres productos, siendo mayor en periodos lluviosos, mientras que la probabilidad de detectar eventos de no lluvia es más favorable en periodos lluviosos que secos.

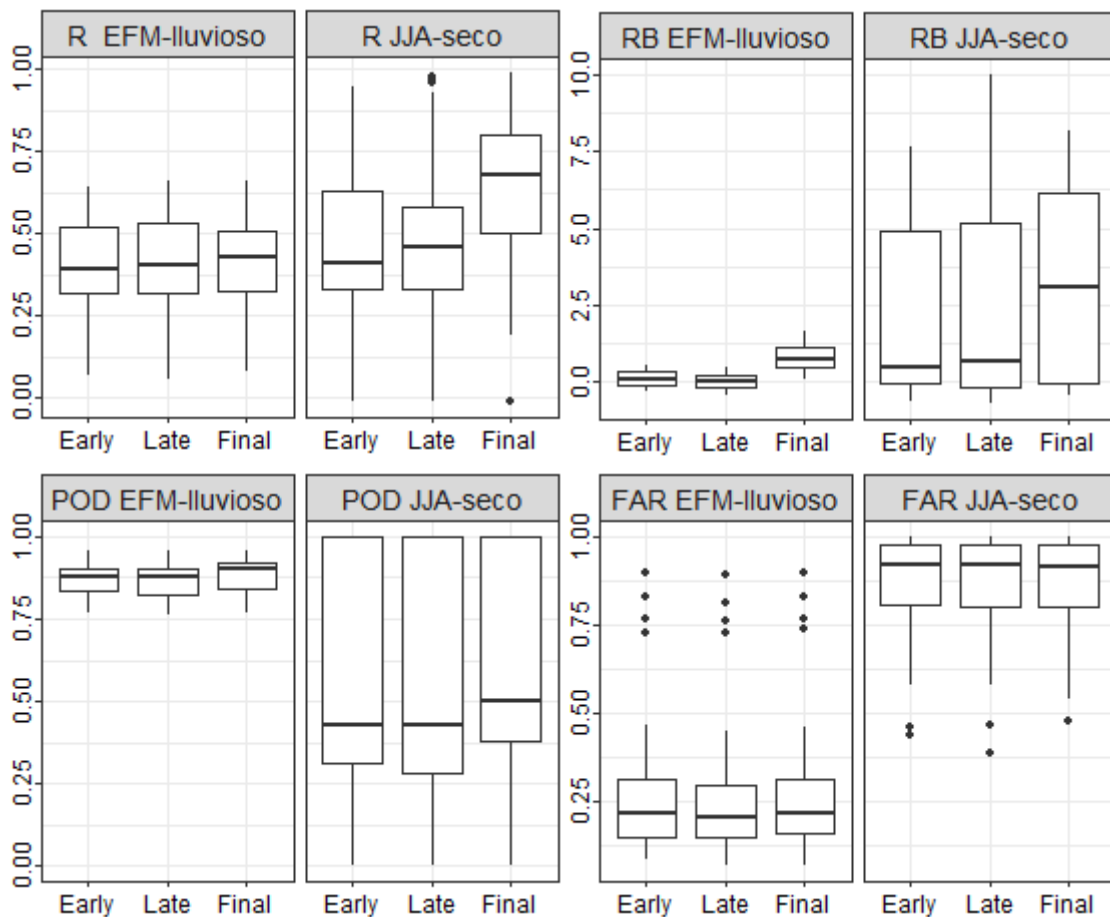


Figura 21: Boxplot de los estadísticos aplicados a todas las estaciones pluviométricas convencionales por periodo lluvioso y seco

FUENTE: Elaboración propia

En la Figura 22, se evalúa la variación espacial de la correlación de pearson (R) y el bias relativo (RB) dentro de la región de estudio CHIRILU, considerando los cinco tipos de climas presentes (CC0= “Semifrígido, lluvioso y deficiencia de lluvia en invierno”, CC1= “frio, lluvioso, lluvia deficiente en invierno”, CC2= “semiseca, semifría con deficiencia de lluvias en invierno”, CC3= “semiseca, templada con deficiencia de lluvias en invierno”, CC4= “desértica, semicálido y deficiencia de lluvias en todas las estaciones del año”). Los estadísticos mostrados son el resultado de la relación entre el valor observado por estaciones convencionales y los tres productos IMERG a paso diario, para los periodos EFM-lluvioso y JJA-seco.

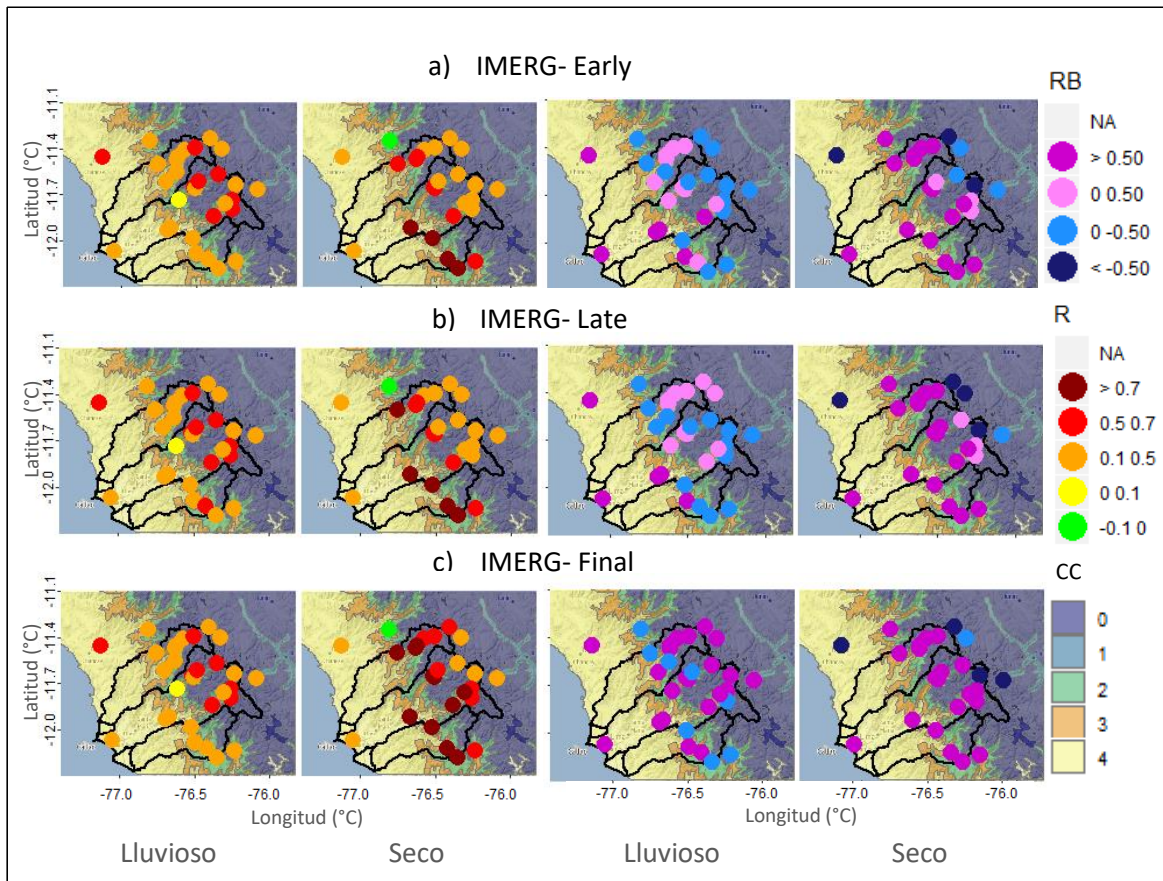


Figura 22: Mapa de análisis espacial de los resultados estadísticos de correlación de pearson “R” (izquierda) y bias relativo “RB (mm)” (derecha) para el periodo lluvioso y seco, entre los datos observados y los tres productos del algoritmo IMERG. La clasificación climática está representada por “CC”.

FUENTE: Elaboración propia

Al analizar la variación espacial de la correlación de Pearson se encuentran valores más altos ($R > 0.7$) en estaciones que pertenecen a clasificaciones climáticas con características frías y de lluvias en verano (CC0, CC2 y CC3), solo se encontró una estación con relación lineal negativa entre sus valores ($-0.1 < R < 0$) en época seca, siendo este un patrón frecuente en los tres productos satelitales. Además, el comportamiento espacial del sesgo o bias relativo en periodos lluviosos, corrobora la sobreestimación de los datos observados para el producto satelital IMERG-Final en más del 90 por ciento de estaciones y en los otros dos productos, hay una marcada subestimación en regiones topográficamente elevadas. En periodos secos la sobreestimación aumenta en estaciones con clasificación CC0 y CC2 que presentan características de deficiencia de lluvia en invierno.

Las funciones de distribución acumuladas estimadas para los tres productos de precipitación satelital y los datos observados en estaciones convencionales para intensidades de lluvia en (mm/día), se muestran en la Figura 23.

El ajuste de la función de distribución acumulada de los tres productos satelitales presentan diferencia relativamente grandes entre sus CDF's en periodos lluviosos. El CDF de IMERG-Final muestran altas tasas de lluvia, por lo que existe una tendencia a sobreestimar al valor observado, mientras que la diferencia entre productos IMERG-Early e IMERG-Late es mínima, sin embargo con tasas menores a 10 mm/día (87 por ciento de probabilidad), existe un ligero cambio con subestimaciones imperceptibles, por lo que en general estos dos productos se acerca más a los valores observados. Por el contrario en periodos secos, la sobreestimación es evidente en los tres productos satelitales, encontrándose mayor diferencia con el producto IMERG-Final. Además, para días con precipitación igual a 0 mm los tres productos se ajustan al valor observado.

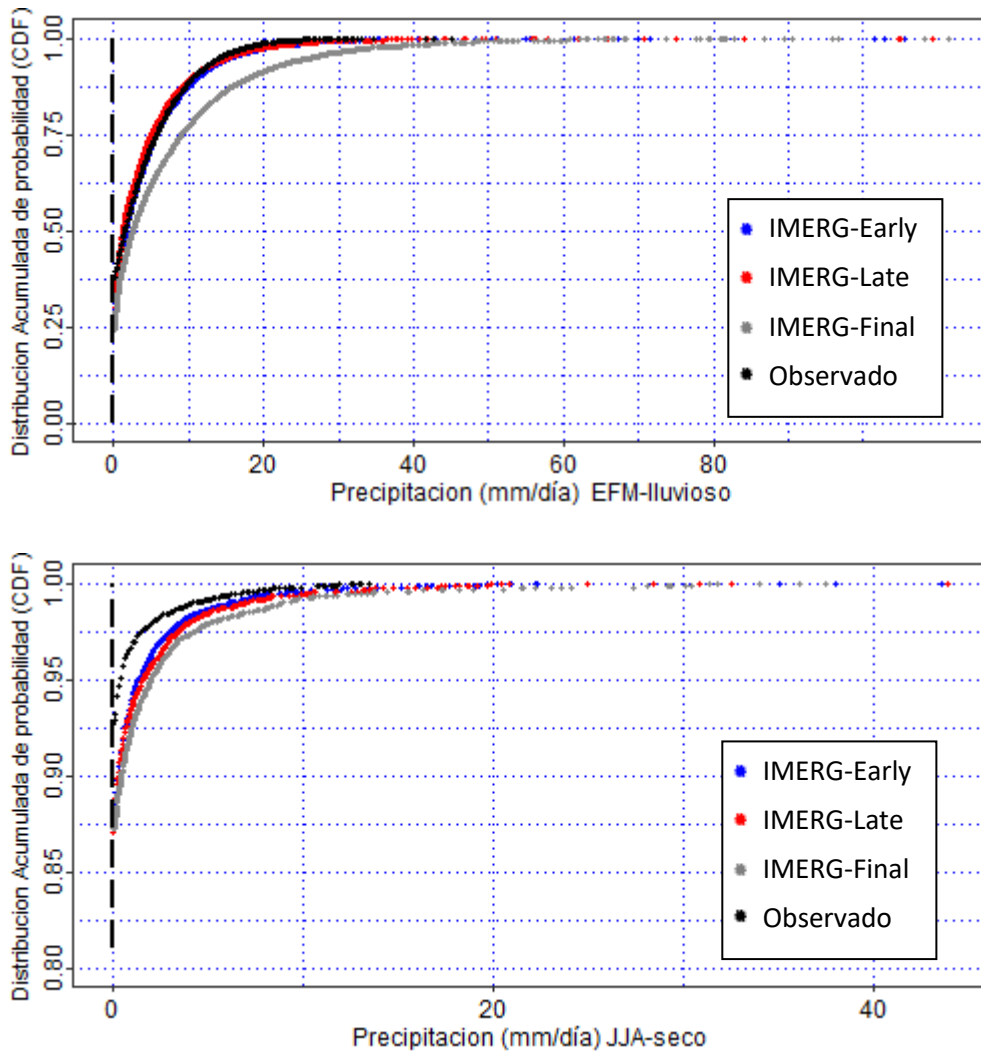


Figura 23: Función de Distribución Acumulada de Probabilidad (CDF) entre datos de precipitación observada convencional y los tres productos del algoritmo IMERG para dos periodos.

FUENTE: Elaboración propia

4.2.3 Validación espacio-temporal de la precipitación estimada por tres productos IMERG en base a estaciones automáticas.

El uso de estaciones pluviométricas automáticas permitió analizar los productos satelitales en escalas temporales menores a un día. En la Tabla 12, se detallan los resultados promedio de los estadísticos aplicados a cada producto satelital de precipitación para los periodos secos y lluviosos de los años 2016, 2017 y 2018 en 11 estaciones automáticas, considerando cinco rangos temporales de ocurrencia de precipitación.

La correlación de pearson R disminuye considerablemente cuando se analizan tasas de lluvia de duración una hora, esto se incrementa aún más en periodos secos. No obstante IMERG-Late tiene mejor respuesta en tiempos cortos de lluvia, debido a que presenta menor sobreestimación de los valores registrados en estaciones automáticas, así como menor raíz del error cuadrático medio (4.39 mm/12 hr). Con respecto a la probabilidad de falsas alarmas de la ocurrencia de un evento de no lluvia, este aumenta a medida que el tiempo de lluvia disminuye. En periodos secos, hay una evidente subestimación para rangos temporales menores a un día, así como la probabilidad de acierto de lluvia no es buena y las falsas alarmas se incrementan.

Tabla 12: Estadísticos entre los datos de precipitación diaria del producto satelital y el registrado por el pluviómetro automático, para dos periodos y cada tiempo de acumulación.

EFM - lluvioso						JJA - seco				
Tiempo 1 hr										
Producto	R	RB	RMSE	POD	FAR	R	RB	RMSE	POD	FAR
IMERG-E	0.11	8.61	0.80	0.34	0.85	0.01	-0.72	0.41	0.02	0.89
IMERG-L	0.11	7.12	0.75	0.36	0.86	0.02	-0.73	0.40	0.02	0.88
IMERG-F	0.10	14.68	1.36	0.35	0.87	0.03	-0.71	0.43	0.02	0.86
Tiempo 3 hr										
IMERG-E	0.16	8.74	1.75	0.45	0.79	0.02	-0.72	0.96	0.03	0.84
IMERG-L	0.16	7.20	1.69	0.46	0.80	0.03	-0.73	0.95	0.03	0.82
IMERG-F	0.14	14.88	3.09	0.46	0.81	0.03	-0.71	1.05	0.03	0.81
Tiempo 6 hr										
IMERG-E	0.17	8.74	2.78	0.54	0.75	0.01	-0.72	1.56	0.04	0.80
IMERG-L	0.17	7.20	2.75	0.55	0.76	0.02	-0.73	1.56	0.05	0.77
IMERG-F	0.15	14.88	4.98	0.56	0.77	0.03	-0.71	1.73	0.05	0.74
Tiempo 12 hr										
IMERG-E	0.21	8.74	4.41	0.66	0.69	0.01	-0.72	2.60	0.07	0.73
IMERG-L	0.20	7.20	4.39	0.66	0.70	0.01	-0.73	2.60	0.08	0.72
IMERG-F	0.17	14.88	7.93	0.67	0.71	0.03	-0.71	2.89	0.07	0.68
Tiempo 24 hr										
IMERG-E	0.35	2.24	5.83	0.85	0.33	0.24	3.84	1.77	0.40	0.70
IMERG-L	0.40	1.88	5.68	0.87	0.32	0.28	4.07	1.79	0.38	0.71
IMERG-F	0.41	4.53	10.26	0.88	0.33	0.54	13.30	1.82	0.47	0.71

FUENTE: Elaboración propia

Dos estaciones pluviométricas automáticas (Matucana en periodo lluvioso y Santo domingo en periodo seco) fueron escogidas en la Figura 24 para representar el comportamiento temporal de la precipitación registrada diaria y los productos de precipitación satelital IMERG en sus tres productos, las series de tiempo para el resto de estaciones automáticas se muestran en el ANEXO 5.

Se observa que el producto IMERG-Early registró picos altos de lluvia en toda la serie temporal para el periodo lluvioso, alcanzando como valor máximo 45 mm/día en enero del 2017, mientras que los valores observados no registraron tasas tan altas de precipitación, por lo que el producto IMERG-Final fue el que se ajustó más a los valores reales. En periodos secos hay un comportamiento similar en ambos tipos de datos, sin embargo algunos picos son sobreestimados por los productos satelitales (precipitación entre 1 a 5 mm/día) para algunas fechas, mientras que el valor observado indica que las intensidades fueron menores a 2 mm/día.

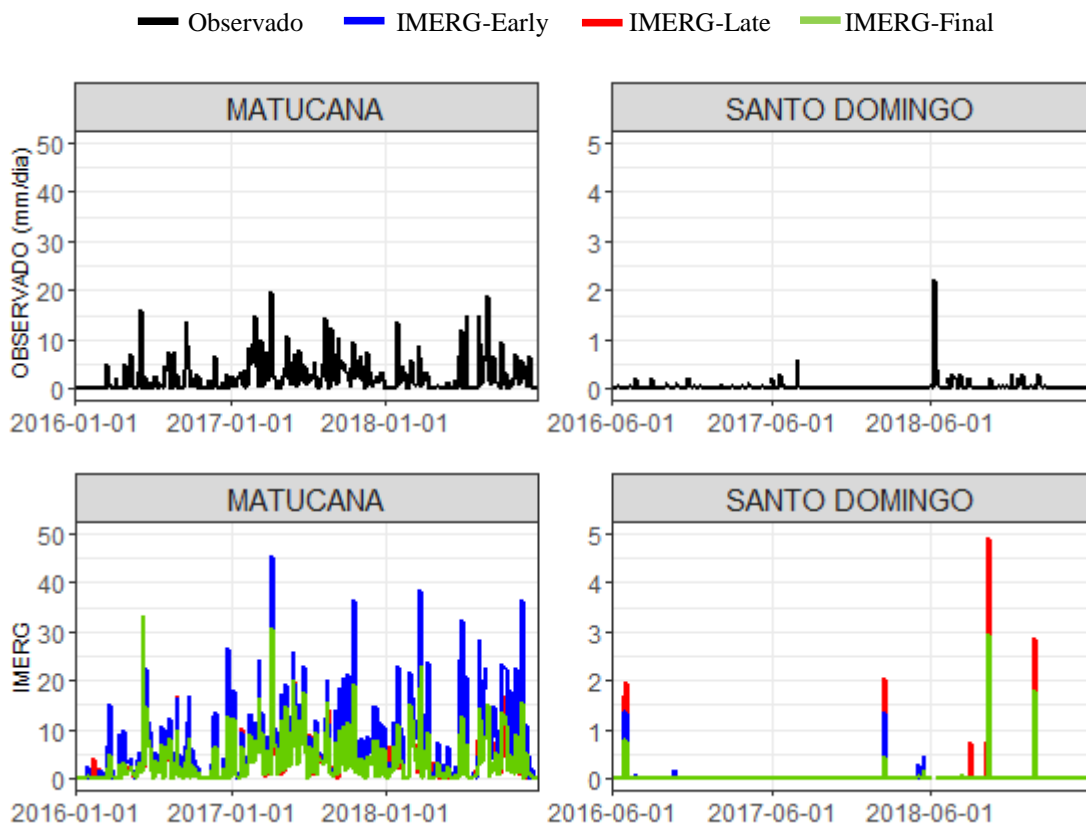


Figura 24: Series de tiempo de las estaciones con mejores estadísticos para la época EFM-lluviosa y JJA-seco en estaciones automáticas.

FUENTE: Elaboración propia

Los estadísticos cuantitativos (R y RB) y categóricos (POD y FAR) aplicados en todas las estaciones pluviométricas automáticas para datos diarios en los periodos EFM-lluvioso y JJA-seco de los años 2016, 2017 y 2018 se presentan en la Figura 25.

Existe mayor dispersión para el producto IMERG-Final, principalmente en periodos secos; el que, por otra parte, presenta una media de R mayor en época lluviosa y seca a comparación de los otros dos productos satelitales, además un valor de sesgo o bias relativo en promedio mayor, por lo que se convierte en el producto satelital que más sobreestima la precipitación. Por otro lado, la tasa de cierto de presentarse un evento de lluvia es alta en los tres productos, siendo mayor en periodos lluviosos, mientras que la probabilidad de detectar eventos de no lluvia es más favorable en periodos lluviosos que secos, no encontrándose mucha diferencia entre los tres productos.

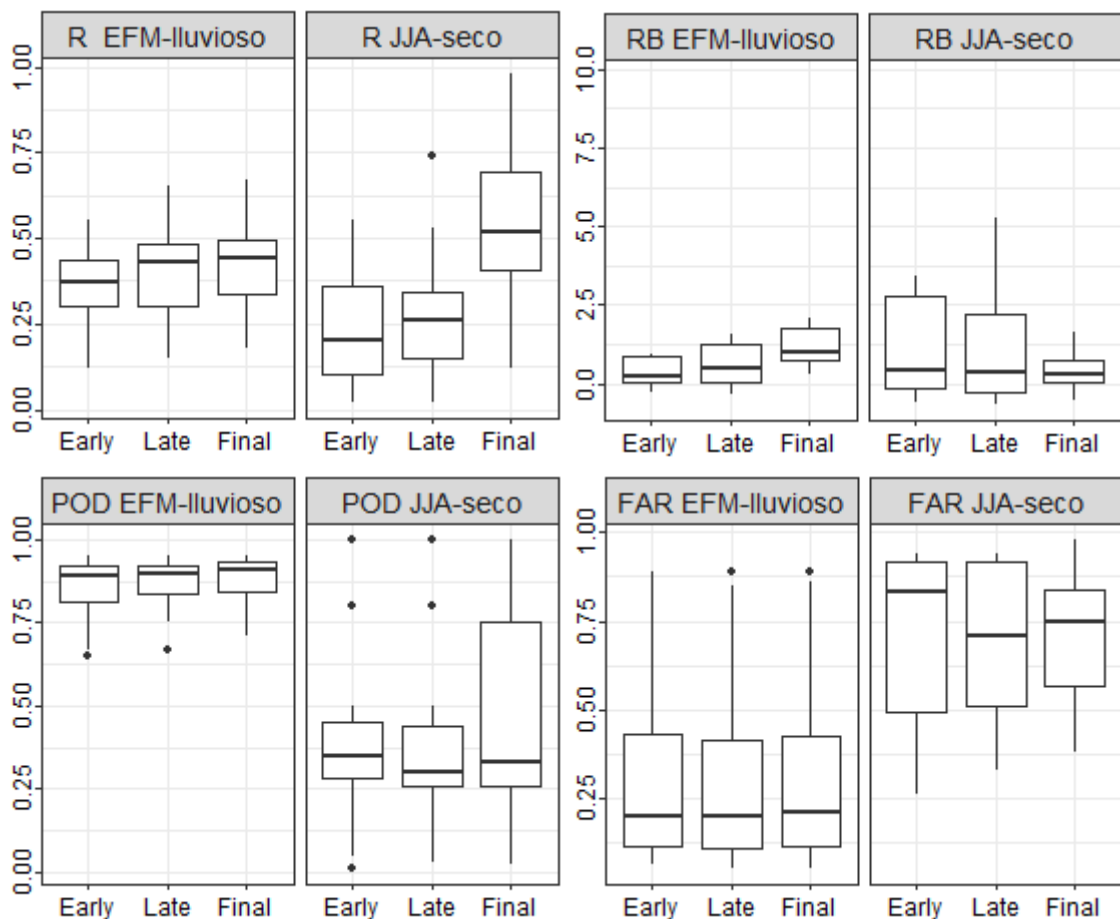


Figura 25: Boxplot de los estadísticos aplicados a todas las estaciones pluviométricas automáticas por periodo lluvioso y seco.

FUENTE: Elaboración propia

Al encontrarse en el análisis temporal, que los estadísticos cuantitativos y categóricos son menos favorables para tasas de lluvia en tiempos cortos (menores a 24 horas), se realizó un gráfico radar que permita evaluar la eficiencia de ellos, de acuerdo a las características de la clasificación climática de la región de estudio CHIRILU (Figura 26). Considerando que las estaciones automáticas solo se ubican en cuatro clasificaciones climáticas, los valores mostrados son un promedio de los resultados estadísticos en las estaciones que pertenecían a una misma clasificación.

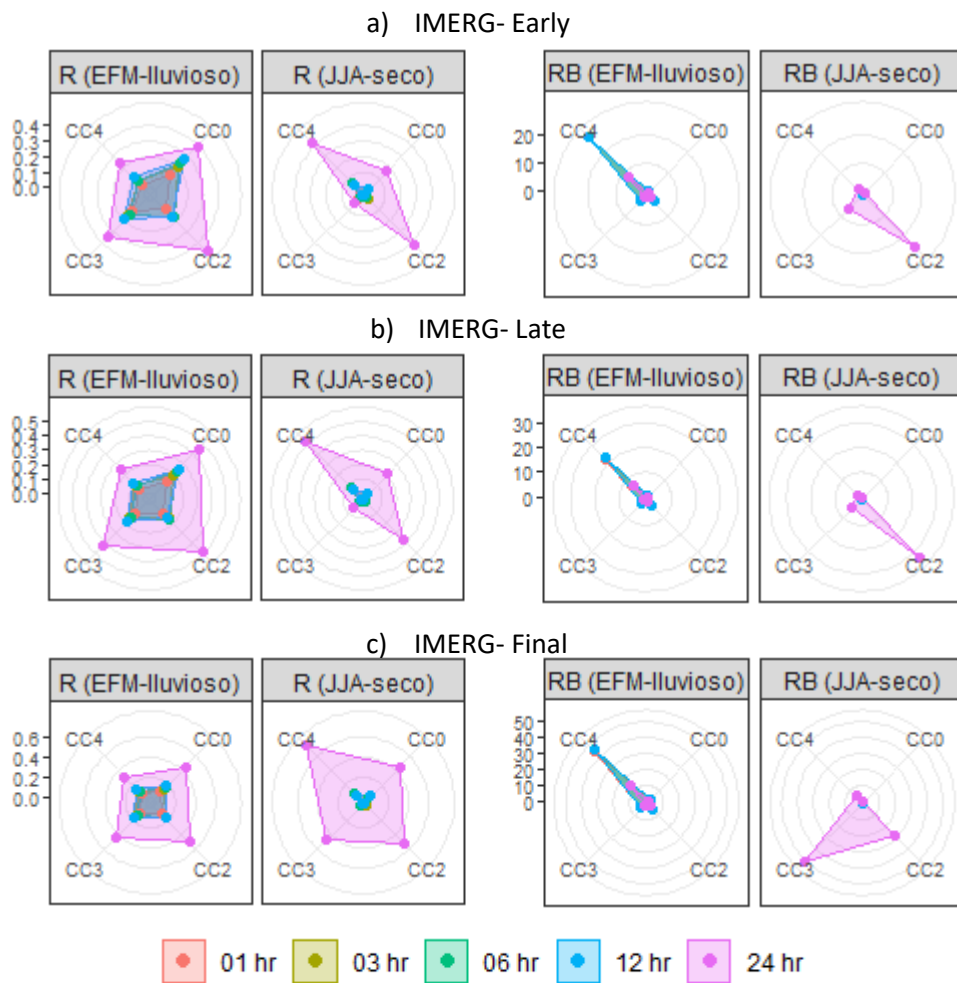


Figura 26: Resultados estadísticos de la validación, coeficiente de correlación de pearson (R) y el sesgo relativo o bias (RB) sobre cuatro tipos de climas para cinco tiempos de acumulación de precipitación

FUENTE: Elaboración propia

El valor de R tiene una tendencia a mejorar en la clasificación climática CC2 (clima semiseco y semifrío, lluvias en verano) para periodos lluviosos en las tres versiones de IMERG, en periodos secos mejora para la CC4 (clima desértico, semicálido y deficiencia de lluvias en todas las estaciones del año). Al evaluar el sesgo o bias relativo RB, la subestimación y sobreestimación mejora para las clasificaciones (CC0, CC2 y CC3) en periodos lluviosos y (CC0 y CC4) en periodos secos.

El ajuste de la función de distribución acumulada también fue evaluada (Figura 27) considerando los tres productos de IMERG.

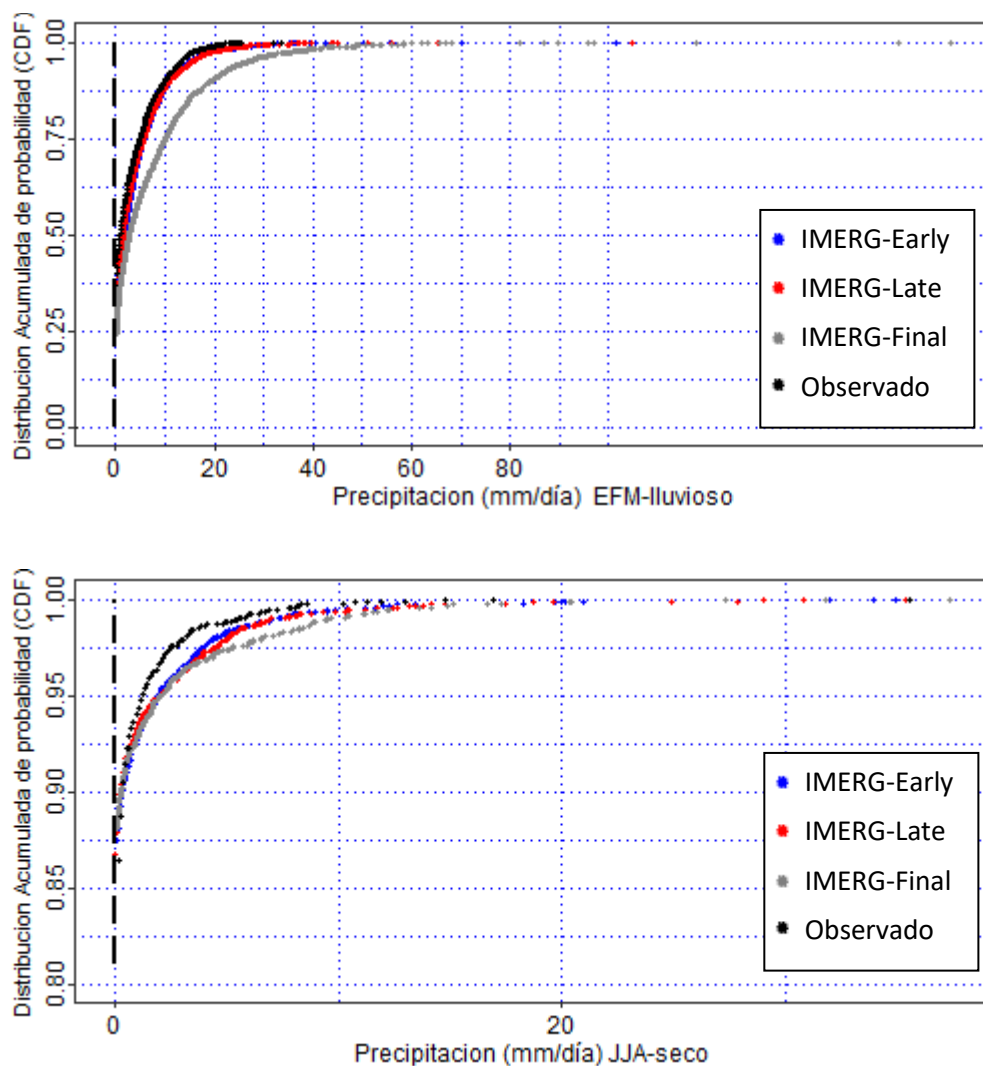


Figura 27: Función de Distribución Acumulada de Probabilidad (CDF) entre datos de precipitación observada convencional y los tres productos del algoritmo IMERG para dos periodos.

FUENTE: Elaboración propia

El CDF de IMERG-Final muestra altas tasas de lluvia, por lo que existe una tendencia mayor a sobreestimar al valor registrado por estaciones automáticas, mientras que la diferencia de los productos IMERG-Early e IMERG-Late es mínima en época lluviosa. En época seca el producto que menos sobreestima es IMERG-Early.

Cabe mencionar que las inexactitudes de las estimaciones de lluvia por productos satelitales IMERG se pueden producir por: a) los algoritmos de sensores de microondas pasivos (PMW) estiman la lluvia dependiendo de los valores de emisividad, sobre la superficie terrestre la emisividad es variable y depende del tipo de suelo y la vegetación. b) la diferencia en los sesgos de precipitación entre diferentes regiones se debe al tipo de nubosidad, en la zona de estudio, las clasificaciones climáticas CC0 y CC2 que se encuentran en elevaciones por encima de los 3500 m.s.n.m desarrollan nubosidades de tipo convectivas, mientras que en la clasificación CC4 es de tipo estratiforme. c) la sobreestimación encontrada en este estudio en el producto IMERG-Final puede estar relacionado por el sesgo que estaría generando su fuente de calibración mensual con datos observados en tierra (GPCC, Centro de Climatología de la Precipitación Global) debido a que la información de la cantidad de estaciones que utiliza este producto sobre un área determinada no es de libre acceso, por lo que una baja densidad de las mismas podría no ser representativo para la zona de estudio.

4.3 Desarrollar y validar la técnica de mezcla para corregir las estimaciones de precipitación por satélite sobre el área de estudio.

4.3.1 Cálculo de residuales y estimación del mejor parámetro I_{dp}

Después de analizar la variación espacio-temporal de la precipitación en base a datos observados en estaciones automáticas-convencionales y compararlo con las estimadas por satélite, se determinó para este análisis usar información de precipitación diaria de estaciones tanto automáticas como convencionales, pero descartando aquellas que se encuentran muy próxima (dentro de una misma grilla del satélite que equivale a 10 kilómetros) por lo que finalmente se utilizaron 33 estaciones pluviométricas. Además, en vista de los resultados obtenidos en la etapa anterior se decidió trabajar la técnica de mezcla o merging solo con el producto IMERG-Late en tres periodos de lluvia (2016 ,2017 y 2018), debido a que este producto sobreestima menos la precipitación diaria, siendo no tan grande la diferencia de sus resultados de correlación de pearson con el producto IMERG-Final y su información de estimación de lluvia (mm/30 minutos) se encuentra disponible en tiempo real, cada 12 horas.

La técnica RIWD de mezcla de producto satelital IMERG-late con los datos observados, se realizó en el lenguaje de programación R, el código empleado se encuentra en el ANEXO 6. Este procedimiento tiene un tiempo de ejecución de aproximadamente de una hora y 30 minutos, por la cantidad de datos empleados (271 días de precipitación, 8 943 iteraciones). Como primer paso, se determinó normalizar la información observada(en puntos espaciales) y satelital(en Raster's) para cada día en todos los puntos de medición, mediante logaritmo natural para que se asemejen a una distribución normal, aplicado este procedimiento se procedió a extraer la información de satélite en cada punto de medición y se aplicó una función lineal para obtener los valores denominados residuales, que son resultado de la diferencia entre el valor observado de la función lineal menos el valor observado real, en el segundo paso se definió la grilla espacialmente de 10 kilómetros de acuerdo a la resolución del satélite. El tercer paso tuvo como objetivo determinar el parámetro de potencia I_{dp} que es usado en la ecuación de interpolación IDW, este parámetro se escogió para cada día en base a una validación cruzada del método Kriging probando para 28 valores en un rango de 0.8 a 3.5, el mejor parámetro fue escogido al

obtenerse el menor error cuadrático medio, los resultados de esta prueba aplicada a cada periodo lluvioso se muestra en la Figura 28, donde se observa que el valor de 0.3 fue escogido con más frecuencia.

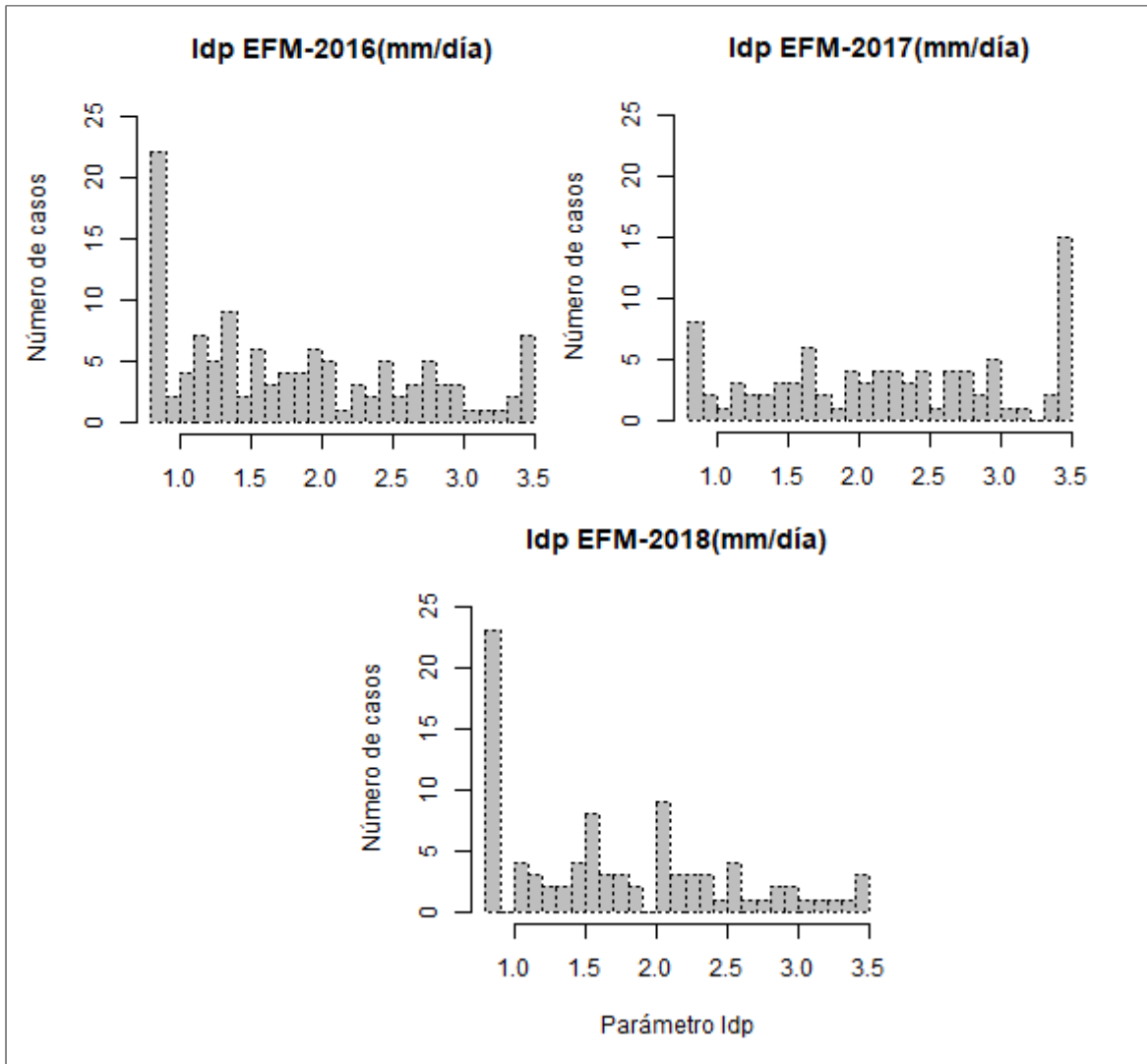


Figura 28: Histogramas de los parámetros I_{dp} empleados en la interpolación RIDW para cada periodo lluvioso.

FUENTE: Elaboración propia

4.3.2 Obtención de producto de mezcla mediante técnica Regression IDW

En el último paso se interpolaron los residuales por el método IDW utilizando para cada día el parámetro I_{dp} escogido, la interpolación tuvo como resultado un raster al cual se le sumo el raster de tendencia es decir, el raster satelital original de precipitación diaria fue multiplicado por el coeficiente que representa la pendiente de la recta lineal y sumado por el coeficiente que representa la intercepción, finalmente se aplica la inversa del logaritmo natural más uno “log1p” al raster de mezcla (desnormalización). En la Figura 29, se representa el raster de tendencia (obtenido por regresión lineal) y residual obtenido para un día (12 de Febrero del 2017), considerado un día lluvioso.

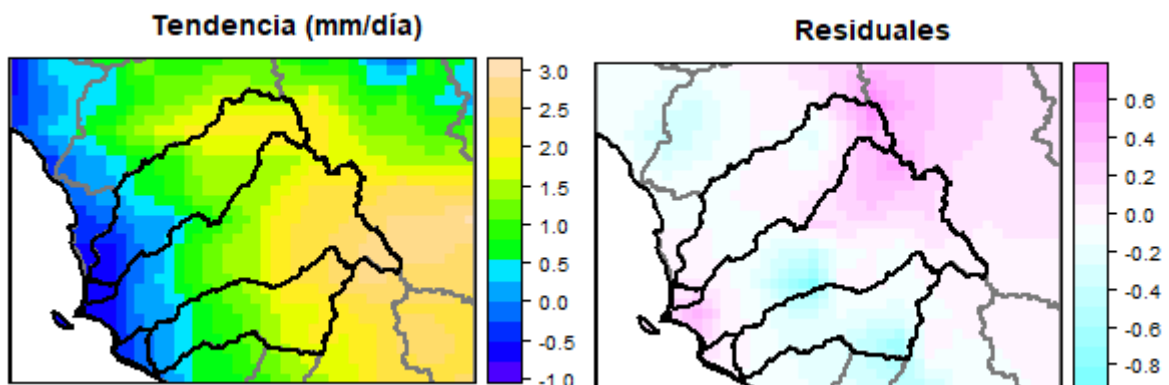


Figura 29: Mapa de valores observados calculados por la función lineal (izquierda) y mapa de valores residuales de la función lineal (derecha).

FUENTE: Elaboración propia

Para evaluar la relación entre los valores de precipitación observada con respecto a la estimación de satélite IMERG-late sin corregir (puntos celestes) y con respecto a la estimación de satélite IMERG-Late corregido por la técnica de mezcla (puntos negros), se utilizaron diagramas de dispersión para cada época lluviosa (Figura 30) y se representaron sus valores de coeficiente de determinación.

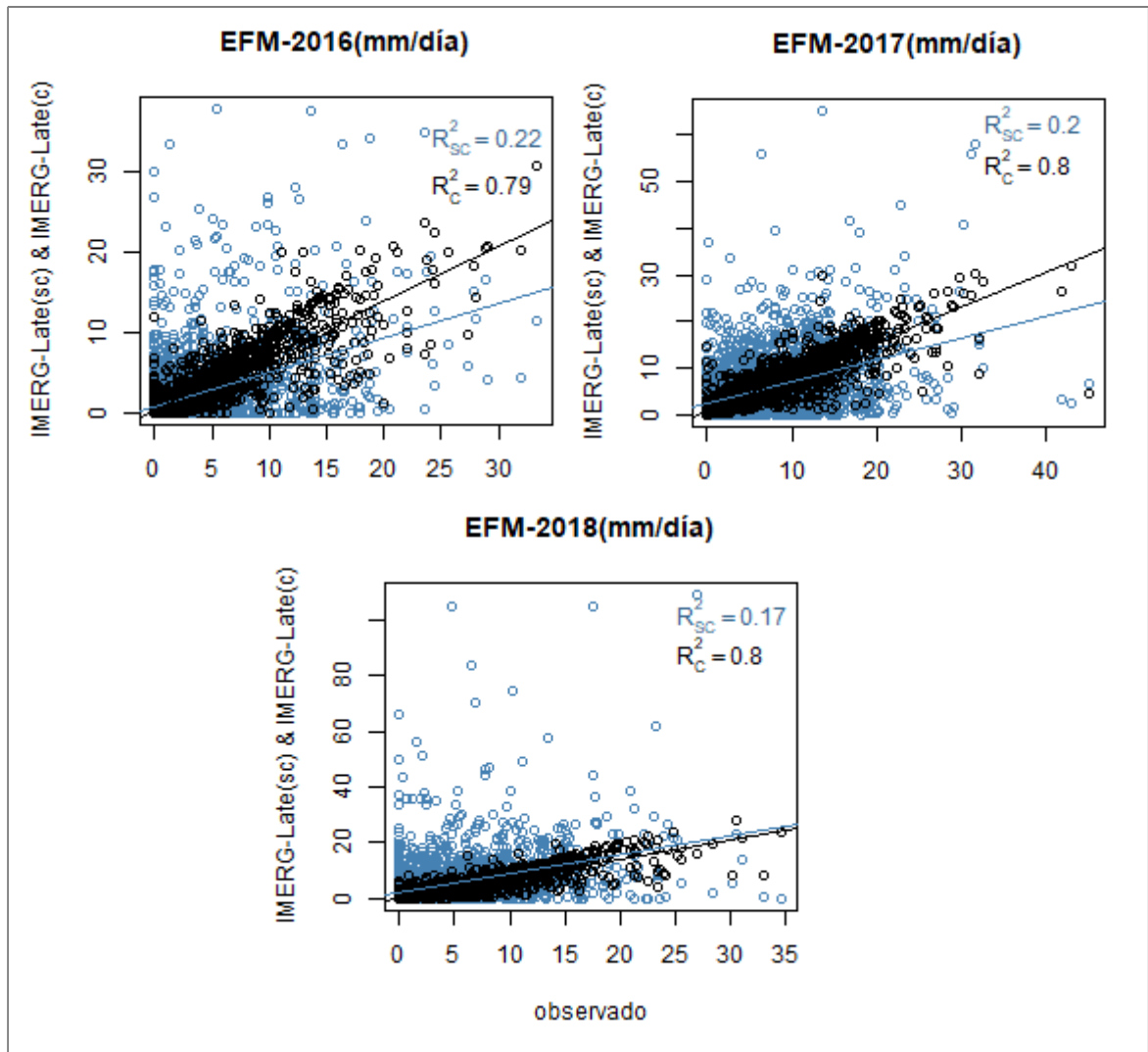


Figura 30: Gráficos de dispersión entre los datos observados de estaciones pluviométricas, producto satelital IMERG-Late y IMERG-Late corregido.

FUENTE: Elaboración propia

Se puede apreciar que el coeficiente de determinación en los tres periodos lluviosos es considerablemente mejor cuando se evalúa la relación entre los datos observados y el satélite IMERG-late corregido, por lo que existe mayor dependencia entre estas variables.

En la Figura 31, se hace una comparación de los valores de precipitación estimados por IMERG-late en comparación al producto de mezcla o IMERG-late corregido para un día. Donde se aprecia una clara disminución en las intensidades de precipitación (mm/día) en el producto de mezcla.

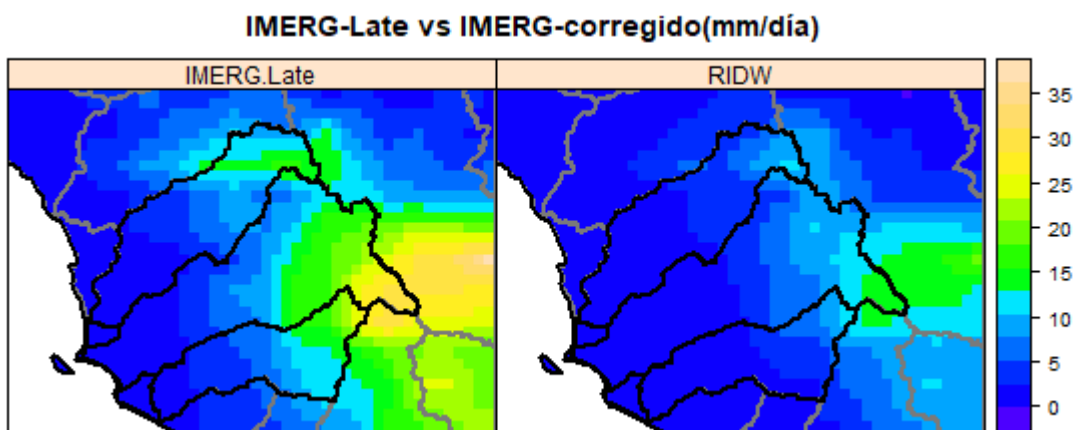


Figura 31: Mapa del producto satelital IMERG-late (izquierda) el producto satelital corregido por técnica de mezcla o merging (derecha)

FUENTE: Elaboración propia

Análisis de tres eventos de precipitación durante “El Niño Costero”

En adición, se realizó un análisis para tres eventos de precipitaciones que ocasionaron el desborde de los ríos pertenecientes a las cuencas de estudio, ocurrido durante “El Niño Costero 2017”. El primer evento se registró entre el 10 y 11 de Marzo debido a las precipitaciones reportadas el 10 de Marzo, se produjo el desborde del río Chillón, afectando viviendas y puentes cercanos, el principal distrito afectado fue “Santa Rosa de Quives” en Canta (INDECI, 2017). En la Figura 32, se detallan los registros de pluviómetros en la parte media y alta de la cuenca, la estación “Huaros” reportó un valor de 14.4 mm para ese día, el producto satelital estimó 21.6 mm y el producto de mezcla por RIDW estimó 11.0 mm/día, para este evento IMERG-late sobreestimó la lluvia en la parte media y alta de la cuenca, lo cual fue corregido por RIDW. El segundo evento fue reportado entre el 14 y 15 de Marzo, las precipitaciones originaron el desborde del río Rímac, perjudicando el tránsito en la carretera central, afectando a personas que viven cerca de la ribera del río, así como el corte de agua en la ciudad. En la Figura 33, se muestra el registro de lluvias para este evento, en la estación “Tingo” se reportó 18.6 mm, el producto satelital estimó 10.6 mm y el producto de mezcla RIDW 10.7 mm, la sobreestimación de los datos observados en la cuenca media y alta fue evidente, mientras que RIDW se acerca más a los valores reales. Finalmente, el tercer evento se reportó también el 14 de Marzo cuando se produjo el huayco y desborde del río Lurín, afectando en Pachacámac y Cieneguilla, IMERG-Late subestimó la lluvia para este día, lo cual fue corregido por la técnica de mezcla RIDW, como se muestra en la (Figura34).

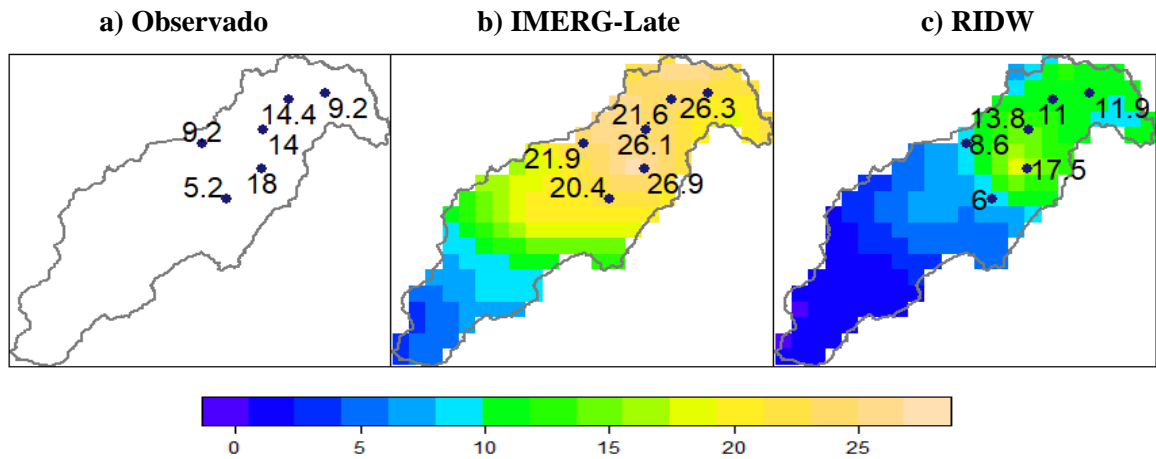


Figura 32: Precipitaciones (mm/día) para el 10 de Marzo del 2017 sobre la cuenca Chillón

FUENTE: Elaboración propia

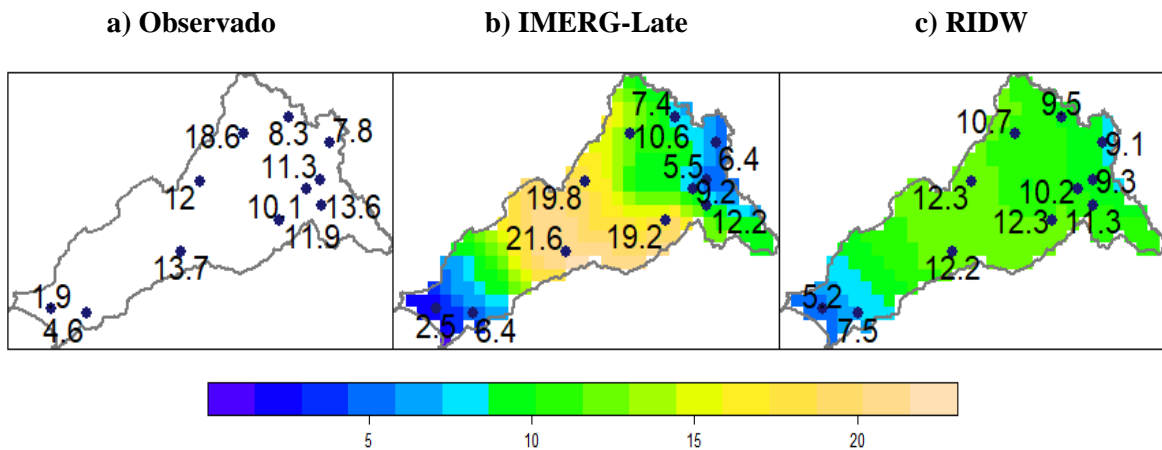


Figura 33: Precipitaciones (mm/día) para el 14 de Marzo del 2017 sobre la cuenca Rímac

FUENTE: Elaboración propia

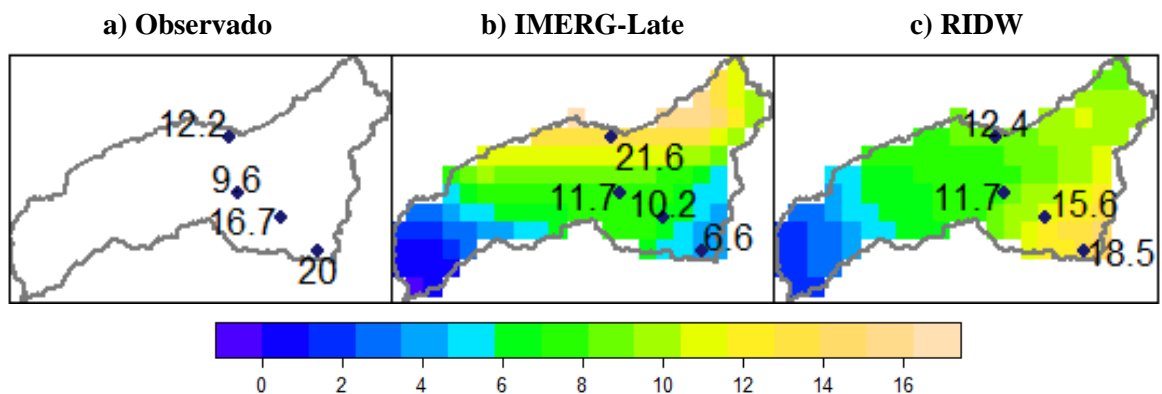


Figura 34: Precipitación (mm/día) para el 14 de Marzo del 2017 sobre la cuenca Lurín

FUENTE: Elaboración propia

Las precipitaciones desarrolladas durante “El Niño costero 2017” fueron captadas por los satélites de la misión GPM mediante su algoritmo IMERG, proporcionando información de lluvia cada media hora, en las figuras anteriores se realizaron acumulaciones para 24 horas, en esta escala espacial el producto satelital detectó los eventos de lluvia, aunque con menor o mayor intensidad. Las estimaciones por RIDW permiten tener información espacial actualizada de precipitación en un tiempo de aproximadamente 12 horas. Cabe recalcar que no se realizó un análisis para periodos JJA de la técnica de mezcla RIDW debido a que en la zona de estudio durante estos periodos la precipitación diaria es mínima e incluso con valores cero, para lo cual no sería significativo emplear el método.

4.3.3 Validación del método de mezcla mediante validación cruzada "LOOCV".

Finalmente se realizó la validación de la técnica de mezcla o merging empleada RIDW. Para cada día analizado se retiró el valor de lluvia de una estación de las 33, con las restantes se aplicó RIDW considerando el valor I_{dp} de ese día ya establecido en el procedimiento anterior, de la mezcla obtenida se extrajo el valor en la ubicación (píxel) de la estación retirada, el cual fue comparado con el valor observado para ese día, el método se repitió para cada estación por lo que se obtuvo una serie de tiempo que fue comparada estadísticamente con los valores observados.

En la Tabla 13, se presenta como ejemplo los resultados para cuatro estaciones, cada una de ellas pertenece a una clasificación climática: San Cristóbal (CC0), San Mateo (CC2), Matucana (CC3), Santa Eulalia (CC4) en los periodos lluviosos del 2016, 2017 y 2018. Para el análisis del producto satelital corregido, se empleó LOOCV o “Validación cruzada dejando uno afuera”. Se obtuvo mejores correlaciones para el producto de mezcla, el sesgo o bias se redujo, así como el RMSE en +/- 2 mm principalmente en estaciones por debajo de los 2000 m.s.n.m. El resultado del RMSE para todas las estaciones se muestra en el ANEXO 7, considerando las clasificaciones climáticas, en casi todas las estaciones se observa una disminución del RMSE a excepción de las estaciones: “Yauli”, “Pallac”, “Patahuay Sapicancha” y “Marcapomacocha” para el año 2016 y la estación “Carampoma” para el 2018. Al evaluar las estimaciones diarias (generadas por la técnica de mezcla) en estas estaciones se encontraron valores atípicos para algunos días que ocasionaron el aumento en el RMSE, esto se originó porque el prescindir del valor de lluvia en una de

estas estaciones puede generar un cambio significativo en la formulación de la ecuación de regresión lineal que permite obtener los residuales para el método RIDW, por lo que la estimación en ese punto se ve afectado y generaría mayores sesgos entre los datos observados y estimados por la técnica de mezcla, que aumentan el error en el cálculo de los estadísticos.

Tabla 13: Estadísticos para cuatro estaciones entre los datos observados y el producto satelital IMERG-late sin corregir (SC) y la validación cruzada de IMERG-late corregido (C).

Estación	Observado vs. IMERG-Late(SC)			Observado vs. LOOCV de IMERG-Late(C)		
	R	RB	RMSE	R	RB	RMSE
EFM-2016						
San Cristóbal	0.53	-0.45	6.12	0.64	-0.29	5.36
San Mateo	0.53	-0.15	3.36	0.82	-0.06	2.00
Matucana	0.65	0.51	3.41	0.73	0.45	2.36
Santa Eulalia	0.90	3.51	1.56	0.93	1.02	0.31
EFM-2017						
San Cristóbal	0.40	-0.24	6.74	0.49	-0.22	5.63
San Mateo	0.49	0.13	5.22	0.81	-0.03	2.69
Matucana	0.41	0.39	5.46	0.73	0.32	3.20
Santa Eulalia	0.46	0.29	4.4	0.92	-0.16	2.37
EFM-2018						
San Cristóbal	0.27	-0.31	6.89	0.47	-0.56	6.39
San Mateo	0.30	0.55	5.86	0.46	0.18	3.98
Matucana	0.71	0.49	3.86	0.81	-0.02	2.51
Santa Eulalia	0.58	12.29	5.88	0.88	1.47	0.65

FUENTE: Elaboración propia

Los resultados mostrados anteriormente son respaldados por la variación espacial de los estadísticos, la correlación R (Figura 35) y el sesgo relativo RB (Figura 36) mejora, con valores mayores a 0.5 con un nivel de confianza del 95 por ciento y se reducen la sobreestimación principalmente en los periodos lluviosos del 2017 y 2018. La variación del RB y R con respecto a la altura se muestra en el ANEXO 8, se observa que la correlación de pearson aumentó en más del 90 por ciento de las estaciones para todas las elevaciones, incluso en algunas estaciones $R \approx 0.9$, mientras que las sobreestimaciones redujeron su valor para estaciones de menor altura. Estos resultados se atribuyen a que los satélites tienden a estimar mejor la precipitación en zonas donde se generan lluvias por convección, mientras que en zonas con formaciones estratiformes como las que se encuentran por debajo los 1000 m.s.n.m hay mayor sesgo en la cantidad de lluvia estimada.

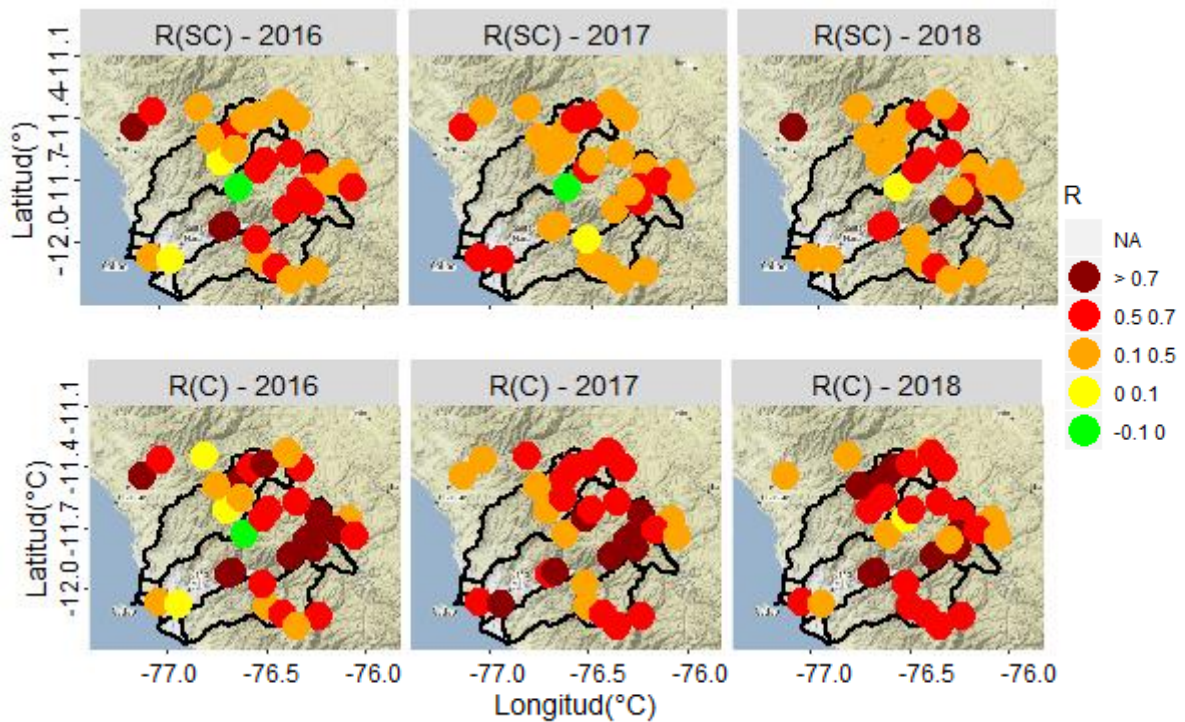


Figura 35: Mapa de análisis espacial de correlación de pearson R entre los datos observados vs producto satelital IMERG-Late sin corregir “R (SC)” y la validación cruzada de IMERG-Late corregido “R(C)”.

FUENTE: Elaboración propia

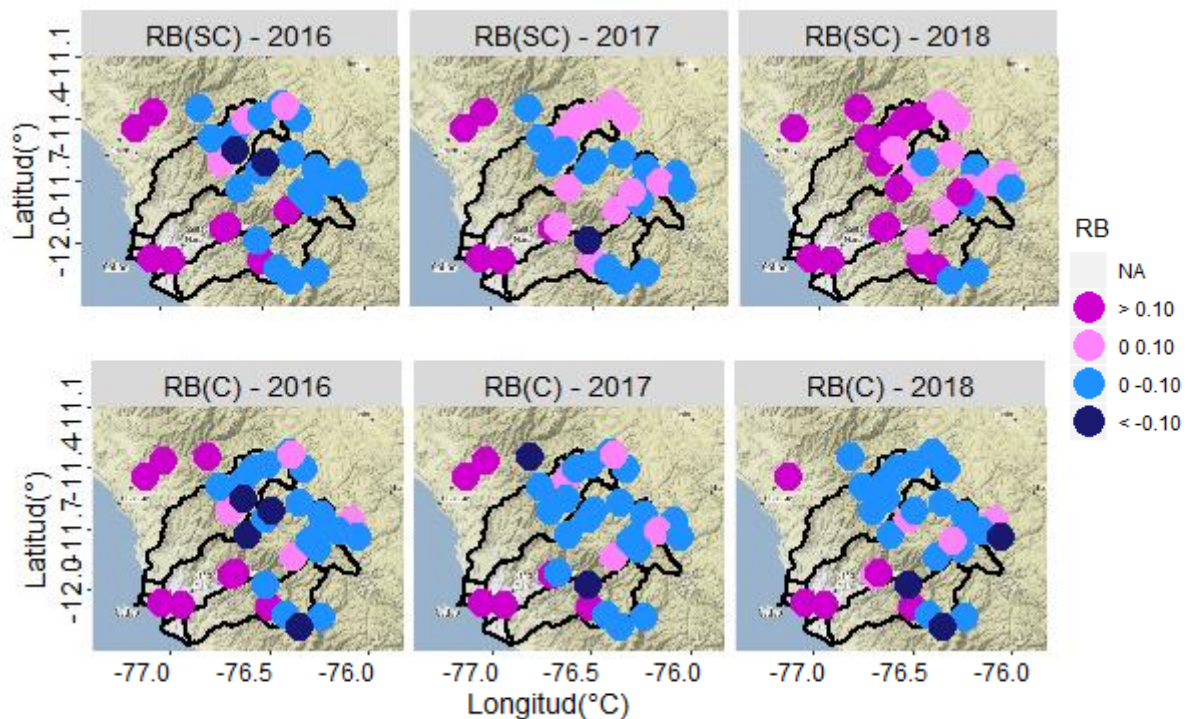


Figura 36: Mapa de análisis espacial de sesgo o bias relativo RB (mm) entre los datos observados vs producto satelital IMERG-Late sin corregir “RB (SC)” y la validación cruzada de IMERG-Late corregido “RB(C)”

FUENTE: Elaboración propia

El último análisis se realizó con el diagrama de Taylor (Figura 37) el cual presenta en resumen, la validación cruzada dejando uno afuera o LOOCV de IMERG-late corregido por técnica merging, presenta mejor bondad de ajuste con una correlación mayor al producto satelital no corregido ($R \cong 0.6$) con un nivel de confianza del 95 por ciento, menor desviación estándar entre 2 y 4 mm/día lo cual indica menor dispersión de los datos con respecto a la media, así como reducción de la raíz del error cuadrático medio entre 1 a 3 mm/día. En el ANEXO 9, se muestra un esquema para usar la técnica en otras zonas.

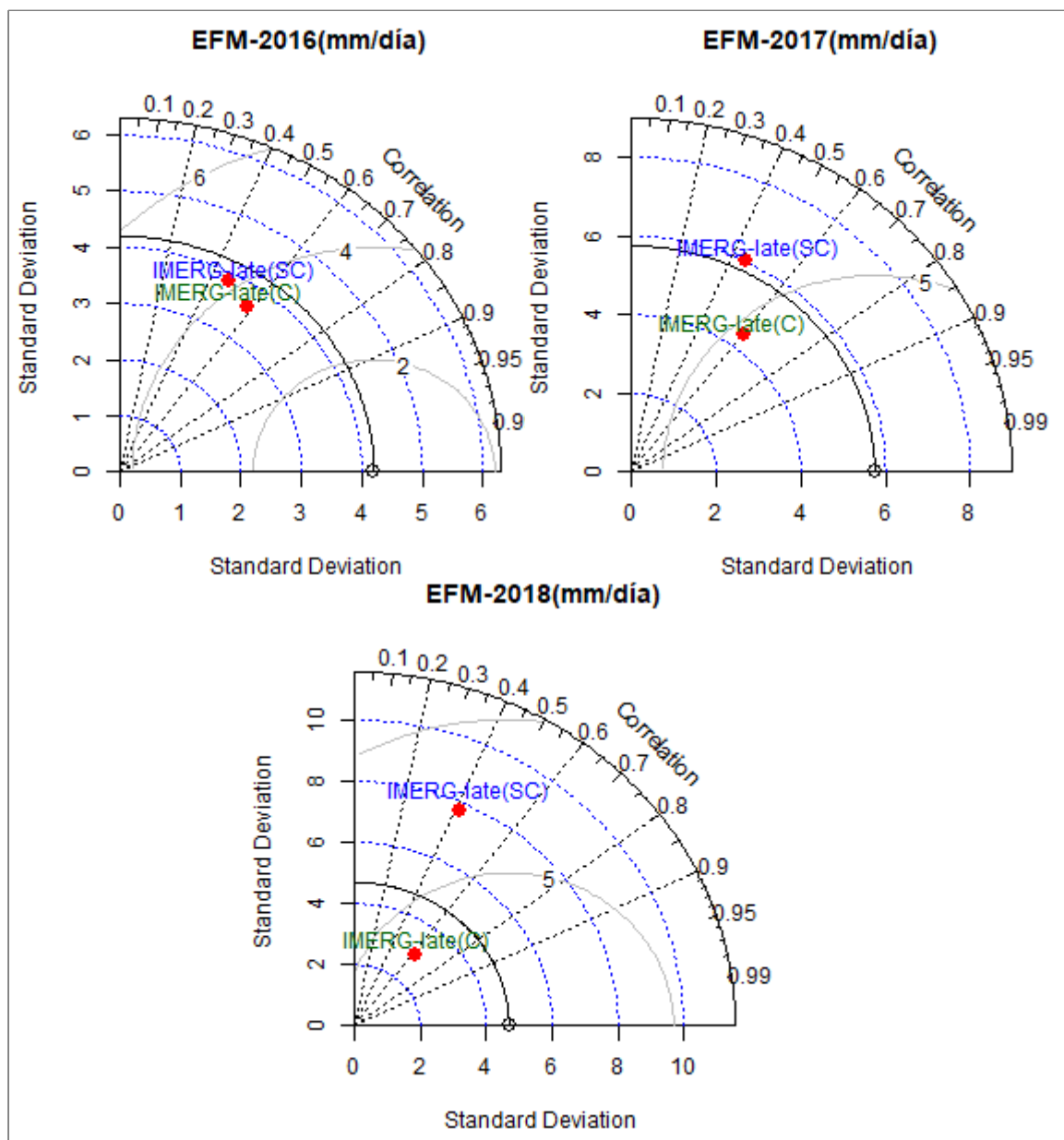


Figura 37: Diagramas de Taylor para evaluar la eficiencia de los producto IMERG-Late sin corregir (SC) y IMERG-Late (C) validado por LOOCV.

FUENTE: Elaboración propia

V. CONCLUSIONES

Conclusiones según los objetivos específicos planteados

- **De efectuar el análisis exploratorio de datos pluviométricos (convencionales y automáticas) y completación de información**, 34 estaciones convencionales fueron analizadas solo cuatro se descartaron por presentar vacíos grandes de información, en el caso de las estaciones automáticas de las 20 se descartaron nueve. El proceso de completación de datos se llevó a cabo utilizando como covariables las variables geográficas y topográficas, completando en total once estaciones convencionales y nueve automáticas, con menos del 2.2 por ciento del total de datos.
- **De evaluar y validar las estimaciones de precipitación de tres productos IMERG en base a estaciones pluviométricas**, con un nivel de confianza del 95 por ciento se obtuvieron coeficientes de correlación significativas ($R > 0.58$) en la validación de estaciones automáticas para datos diarios de precipitación, sesgos menores a 29 por ciento principalmente en estaciones con elevaciones por encima de los 2500 m.s.n.m., por lo que se determinó incluir estaciones automáticas en la validación de las estimaciones satelitales. De la validación de los tres productos IMERG, IMERG-Late resultó ser el mejor producto de estimaciones de precipitación satelital. Con respecto a las intensidades de precipitación analizadas, quedó demostrado que se debe utilizar IMERG-Late para estimaciones diarias (mm/día).
- **Del desarrollo y validación de la técnica de mezcla para corregir las estimaciones de precipitación por satélite sobre el área de estudio**, el producto de mezcla por RIDW (IMERG-late(C)) es capaz de detectar con mayor precisión las estimaciones de lluvia observada en las cuencas CHIRILU que el producto IMERG-late (SC). La validación LOOVC permitió evaluar la independencia de los

resultados estadísticos en las estaciones, determinando la importancia de cinco estaciones en el área de estudio. Además, quedó demostrado por medio de los estadísticos que la técnica de mezcla RIDW mejora las estimaciones de precipitación por el producto IMERG-Late, con correlaciones en promedio de 0.4 a 0.6 (en 20 por ciento) así como disminución en promedio del RMSE de 1 a 3 mm/día. Por lo que si la medición de precipitación en una estación pluviométrica en la zona CHIRILU para un día determinado es de por ejemplo 12 mm/día y la estimación satelital 15 mm/día, el producto satelital corregido estimará entre 12 y 14 mm/día aproximadamente.

VI. RECOMENDACIONES

- Se recomienda, para el uso de los productos satelitales GPM-IMERG a escala diaria, realizar una validación previa con la información de datos en estaciones pluviométricas convencionales y automáticas disponibles, para evaluar la confiabilidad de ambas fuentes de información en diferentes escalas temporales.
- Se recomienda usar IMERG-Late en la región Lima en sistemas de alerta temprana, principalmente para estudios relacionados a las cuencas Chillón, Rímac y Lurín, por su mejor desempeño sobre otros productos GPM-IMERG y su disponibilidad en tiempo real.
- El uso de la técnica de mezcla de Regresión por inverso de la distancia (RIDW) demostró buenos resultados para la mejora del producto satelital IMERG-Late a escala diaria, sin embargo en estudios a escalas mensuales o anuales de precipitación, se recomienda emplear técnicas geoestadísticas en la interpolación de los residuales que permitan capturar el grado de correlación espacial entre los datos, por ejemplo la técnica “Conditional Merging” utiliza la interpolación Kriging, la técnica RK que utiliza la regresión lineal como modelo de ajuste pero interpola los residuales por Kriging simple, entre otros.
- Los datos satelitales como son los productos GPM-IMERG validados sobre un área de estudio, pueden ser empleados en modelos atmosféricos de predicción numérica como WRF o ECMW, en la asimilación de sus datos, ya que están disponibles en tiempo casi real y presenta mejoras en su resolución espacial y temporal.
- Analizar el uso de diferentes fuentes de estimaciones de precipitación satelital en tiempo real, como por ejemplo los que se basan en la información de temperatura del tope de las nubes (Hidroestimadores GOES) que presentan mejor resolución espacial y temporal.

VII. REFERENCIAS BIBLIOGRÁFICAS

- Acuña, J; Felipe, O; Fernandez, C. 2015. Análisis regional de frecuencia de precipitación anual para la determinación de mapas de sequías en las cuencas Chillón, Rímac, Lurín y Alto Mantaro. Revista Peruana Geo-Atmosférica RPGA (4):93-108.
- Aguilar, E; Prohom, M. 2011. EXTRAQC Centre for climate change and servei meteorologic de Catalunya. España. Disponible en www.c3.urv.cat/data.html.
- ANA (Autoridad Nacional del Agua, Perú). 2003. Estudio integral de los recursos hídricos en la cuenca del río Chillón (en línea). Lima, Perú, Consultado 18 dic. 2018. Disponible en <http://repositorio.ana.gob.pe/handle/ANA/1725>.
- ANA (Autoridad Nacional del Agua, Perú). 2004. Estudio hidrológico de la cuenca del río Lurín: informe final (en línea). Lima, Perú, Consultado 18 dic. 2018. Disponible en <http://repositorio.ana.gob.pe/handle/ANA/1691>.
- ANA (Autoridad Nacional del Agua, Perú). 2010. Estudio hidrológico y ubicación de la red de estaciones hidrométricas en la cuenca del río Rímac. Repositorio Institucional ANA (en línea). Lima, Perú, Consultado 17 dic. 2018. Disponible en <http://repositorio.ana.gob.pe/handle/ANA/5>.
- Aybar, C; Fernández, C, Huerta, A; Lavado, W; Vega, F; Felipe-Obando, O. 2019. Construction of high-resolution gridded Rainfall Dataset for Peru from 1981 to the present day. Hydrological Sciences Journal. De próxima publicación.
- Babak, O; Deutsch, CV. 2008. Statistical approach to inverse distance interpolation (en línea). Stochastic Environmental Research and Risk Assessment 23(5):543-553. Consultado 14 en. 2019. Disponible en [doi:10.1007/s00477-008-0226-6](https://doi.org/10.1007/s00477-008-0226-6).
- Bedoya, D. 2017. El Niño costero, el fenómeno más brutal que golpeó Lima. (en línea). El Comercio, Lima, Perú; 24 dic. Consultado 01 ene. 2019. Disponible en <https://elcomercio.pe/lima/sucesos/nino-costero-fenomeno-brutal-golpeo-lima-noticia-482836>.

- Brito, D. 2015. Precipitación Actual mediante datos GPM, Programa Regional de Empleo de Información Satelital para la Producción Agrícola BPR-CONAE (en línea), 4p.
- Brunet-Moret, Y. 1979. Homogénéisation des précipitations. Cahiers ORSTOM. Serie Hydrol. 16:3-4.
- Cañada, R. 2008. Sistemas y análisis de la información geográfica. España, Editorial Alfaomega. 928p.
- Ceccato, P; Dinku, T. 2010. Introduction to remote sensing for monitoring rainfall, temperatura, vegetation and Institute for Climate and Society (en línea). IRI Technical Report 10-04. Consultado 17 ene. 2019. Disponible en <https://iri.columbia.edu/docs/publications/Introduction%20to%20remote-sensing-environmental%20monitoring.pdf>
- COEN (Centro de Operaciones de Emergencia Nacional, Perú). 2017. Precipitaciones pluviales en las provincias del departamento de Lima (en línea). Lima, Perú. 39 p. Informe de emergencia N° 462. Consultado 15 ene. 2019. Disponible en <https://reliefweb.int/sites/reliefweb.int/files/resources/20170412173959-2.pdf>
- Davis, BM. 1987. Uses and abuses of Cross-validation in geostatistics. *Mathematical Geology*, 19(3):241-248.
- Domínguez, D; Gonzáles, MH. 2012. Variabilidad interanual de la precipitación estival en el centro oeste de Argentina, XI Congreso Argentino de Meteorología (CONGREGMET XI)(5, 2012, Mendoza, Argentina). Mendoza, Argentina. 14p.
- Fallas, J. 2012. Correlación lineal, midiendo la relación entre dos variables. Universidad para la cooperación internacional. 30p.
- Foehn, A; Hernández, JG; Schaeffli, B; de Cesare, GD. 2018. Spatial interpolation of Precipitation from multiple rain gauge networks and Weather radar data for Operational applications in Alpine catchments (en línea). *Journal of Hydrology* (563):1092-1110. Consultado 10 ene. 2019. Disponible en <https://doi.org/10.1016/j.jhydrol.2018.05.027>.
- Garreaud, RD. 2001. Pronóstico de la convección en el altiplano sudamericano: empleo del modelo regional Eta/Cptec. *Meteorológica* 26, N°1 y 2 25 – 38.
- Garreaud, RD; Vuille, M; Compagnucci, R; Marengo, J. 2009. Present-day South American climate. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 281(3-4): 180-195.

- Garreaud, RD. 2018. A plausible atmospheric trigger for the 2017 coastal El Niño (en línea). *International Journal Climatol* (38):296-302. Consultado 11 feb. 2019. Disponible en <https://doi.org/10.1002/joc.5426>
- Hartigan, J.A; Wong M.A. 1979. Algorithm AS 136: a K-means clustering algorithm. *Journal of the Royal Statistical Society .Serie C*.28 (1):100-108
- Hou, AY; Kakar, RK; Neek, S; Azarbarzin, AA; Kummerow, CD; Kojima, M; Oki, R; Nakamura, K; Iguchi, T. 2014. The global Precipitation Measurement misión (en línea). *Boletín de la Sociedad Americana de Meteorología*. Consultado 12 dic. 2018. Disponible en doi:10.1175/BAMS-D-13-00164.1.
- Hong, Y; Hsu, K; Gao, X; Sorooshian, S. 2004. Precipitation estimation from remotely sensed information using an artificial neural network – cloud classification System. *Journal of applied Meteorology* (43):1834–1852.
- Hu, Q; Li, Z; Wang, L; Huang, Y, Wang, Y; Li, Lingjie. 2019. Rainfall Spatial estimations: a review from Spatial interpolation to multi-source data merging (en línea). *Water* 11(3):579. Consultado 1 abr. 2019. Disponible en <https://doi.org/10.3390/w11030579>.
- Huffman, G; Bolvin, D; Nelkin, E; Wolff, D; Adler, R; Gu, G;Hong, Y; Bowman, K; Stocker, E. 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales (en línea). *Journal Hydrometeorology* (8):38–55, Consultado 17 feb. 2019. Disponible en doi:10.1175/JHM560.1.
- Huffman, G; Bolvin, D; 2014. Documento técnico: Real-Time TRMM Multi-Satellite Precipitation Analysis Data Set Documentation. NASA 51p.
- INDECI (Instituto Nacional de Defensa Civil, Perú).2017. Consultado 17 Ago. 2017. Disponible en www.indeci.gob.pe/objetos/alerta/MjE0NQ==/20170817171954.pdf.
- Joseph, VR; Kang, L. 2011. Regression-based Inverse Distance Weighting with Applications to computer experiments. *Technometrics* 53(3):254-265. Consultado 10 ene. 2019. Disponible en doi:10.1198/tech.2011.09154
- Joyce, R; Janowiak, J; Arkin, P; Xie, P.2004. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution (en línea). *Journal Hydrometeoroly* (5):487–503.

- Consultado 14 oct. 2018. Disponible en doi: 10.1175/1525-7541(2004)0052.0.CO;2.
- Joyce, R; Xie, P. 2011. Kalman filter–based CMORPH(en línea). *Journal Hydrometeorology* (12):1547–1563. Consultado 21 nov. 2018. Disponible en doi:10.1175/JHM-D-11-022.1.
 - Kulie, M; Bennartz, R. 2009. Utilizing spaceborne radars to retrieve dry snowfall. *Journal of applied Meteorology and Climatology*, 48:2564–2580p.
 - Lavado, W.S; Ronchail, J; Labat, D; Espinoza, J.C ; Guyot, J.L. 2012. Basin scale analysis of rainfall and runoff in Peru (1969-2004): Pacific, Titicaca and Amazonas drainages. *Hydrological Sciences Journal*, 57(4):625-642.doi:10.1080 / 02626667.2012.
 - Lillesand, MT; Kiefer, WR. 1994. Remote sensing and image interpretation (en línea). New York, John Wiley & Sons. *Revista Geológica* 132(02):248. doi:10.1017/s001675680001204.
 - Lu, GY; Wong, DW. 2008. An adaptive inverse-distance weighting spatial interpolation technique (en línea). *Computers & Geosciences*, 34(9):1044-1055. Consultado 17 dic. 2018. Disponible en doi:10.1016/j.cageo.2007.07.010.
 - Martens, B; Cabus, P; De Jongh, I; Verhoest, N.E.C. 2013. Merging weather radar observations with ground-based measurements of rainfall using and adaptative multicuadric surface fitting algorithm(en línea). *Journal of hydrology* 500:84-96.consultado 16 abr. 2019. Disponible en doi: 10.1016 / j.jhydrol.2013.07.011.
 - Mestre, O; Gruber, C; Prieur, C; Caussinus, H; Jourdain, S. 2011. SPLIDHOM: a method for homogenization of daily temperatura Observation (en línea). *Journal of applied meteorology and climatology* 50(11):2343-2358. Consultado 7 ene. 2019. Disponible en <http://dx.doi.org/10.1175/2011JAMC2641.1>.
 - Mestre, O; Domonkos, P; Picard, F; Auer, I; Robin, S; Lebarbier, E; Bohm, R; Aguilar,E; Guijarro, J; Vertachnik, G; Klancar, M; Bubuisson, B; Stepanek, P. 2013. HOMER:Homogenisation software in R-methods and applications (117):47-67.
 - Mugnai, A. 2007. Snowfall measurements by proposed European GPM Mission. *Measuring Precipitation from Space: EURAINSAT and the future*. Turquia, Springer-Verlag, 655-674.

- NASA (National Aeronautics and Space Administration, Estados Unidos). 2015. Reformatting the multi-satellite precipitation estimates (IMERG and TMPA) for display in Geographic Information Systems (GIS) (en línea, sitio web). Consultado 22 feb. 2019. Disponible en URL: [//pps.gsfc.nasa.gov](http://pps.gsfc.nasa.gov)
- Nashwan, M; Shahid, S; Wang, X. 2019. Assesment of satellite-based Precipitation measurement products over the hot desert climate of Egypt (en línea). Remote sensing 11(5):555. Consultado 20 abr. 2019. Disponible en <https://doi.org/10.3390/rs11050555>.
- Nikolaos-Kosmas, C. 2013. Comparison of statistical methods for genomic signature extraction. Tesis Ing. Chania, Grecia. Techical University of crete department of electronic and computer engineering digital signal & image processing lab. 89p.
- Pareja, J. 1936. El clima del Perú (en línea). Revista de la Universidad Católica N°29:645-655. Consultado 9 abr. 2019. Disponible en <http://creativecommons.org/licenses/by-nc-nd/2.5/pe/>
- Qingfang, H. 2013. Rainfall spatial estimation using multi-source information and its hydrological application. Tesis Ph.D. Beijing, China, Tsinghua University.
- Pielke, RA. 1984. Mesoscale Meteorological Modeling. Academic Press, Orlando. 760p.
- Ramírez, IJ; Briones, F. 2017. Understanding the El Niño Costero of 2017: The Definition problem and challenges of climate forecasting and disaster responses(en línea). International Journal of Disaster Risk Science 8(4):489-492. Consultado 16 abr. 2019. Disponible en doi:10.1007/s13753-017-0151-8.
- Rau, P; Bourrel, L; Labat, D; Melo, P; Dewitte, B; Frappart, F; Lavado, W; Felipe, O. 2016. Regionalization of rainfall over the Peruvian Pacific slope and coast. International Journal of Climatology 37(1):153-158. doi:10.1002/joc.4693.
- Rodríguez-Morata, C; Díaz, HF; Ballesteros-Canovas, JA; Rohrer, M; Stoffel, M. 2018. The anomalous 2017 coastal El Niño event in Peru. Climate Dynamics. Doi: 10.1007/ s00382-018-4466-y.
- SENAMHI (Servicio Nacional de Meteorología e Hidrología, Perú). 2002. Mapa de clasificación climática del Perú (en línea). Lima, Perú.9 p. Consultado 15 dic. 2018. Disponible en <https://debconsulting.weebly.com/peruacute-clasif-climat-senamhi---wt.htm>.

- SENAMHI (Servicio Nacional de Meteorología e Hidrología, Perú). 2016. Estudio de vulnerabilidad climática de los recursos hídricos en las cuencas de los ríos Chillón, Rímac, Lurín y parte alta del Mantaro. Repositorio Institucional ANA (en línea). Lima, Perú, 250 p. Consultado 15 ene. 2019. Disponible en <http://repositorio.ana.gob.pe/handle/ANA/2088>
- Serrano-Notivoli, R; de Luis, M; Beguería, S. 2017. An R package for daily Precipitation climate series reconstruction (en línea). *Environmental modelling & software* (89):190-195. Consultado 7 ene. 2019. Disponible en <https://doi.org/10.1016/j.envsoft.2016.11.005>.
- Song, J; DePinto, JV. 1995. A GIS-based Data Query System. Presented at the International Association for Great Lakes Research (LAGLR) Conference, Windsor, Ontario.
- Song, JJ; Kwon, S; Lee, G. 2014. Regression-based inverse distance Weighting (RIDW) for radar Rainfall estimation. *Corea* 844-845.
- Sorooshian, S; Hsu, K; Gao, X; Gupta, H; Imam, B; Braithwaite, D. 2000. Evaluation of PERSIANN system satellite–based estimates of tropical Rainfall (en línea). *Bulletin of the American Meteorological Society* (81):2035–2046. Consultado 11 set. 2018. Disponible en [doi:10.1175/1520-0477\(2000\)0812.3.CO;2](https://doi.org/10.1175/1520-0477(2000)0812.3.CO;2).
- Stepanek, P. 2008. ProClimDB-Software for processing climatological datasets CHMI, regional office Brno (en línea, sitio web). Disponible en www.climahom.eu/ProcData.html
- Takahashi, K. 2017. The El Niño phenomenon: Global vs coastal (Fenómeno El Niño: Global vs “Costero”). *Generacion de informacion y monitoreo del Fenómeno El Niño-NBoletín técnico*. Instituto Geofísico del Perú.Ministerio del Ambiente, Perú. 4(4).
- Tomczak, M. 1998. Spatial interpolation and it uncertainty using automated anisotropic inverse distance Weighting (IDW) cress validation-Jackknife approach. *Journal of Geographic Information and decisión Analysis* 2(2):18-30.
- Tukey, JW. 1977. *Exploratory data analysis*. Pearson. 712p.
- Waylen, P; Poveda, G. 2002. El Niño-Southern Oscillation and aspects for western South American hydro-climatology. *Hydrol. Process.* 16:1247-1260.
- Wilks, D. 2011. *Statistical methods in the atmospheric sciences*. Academic press.70p.

- Woodman, R; Takahashi, K. 2014. ¿Por qué no llueve en la costa del Perú (salvo durante El Niño)? Boletín Técnico “Generación de modelos climáticos para el pronóstico de la ocurrencia del Fenómeno El Niño” Instituto Geofísico del Perú. Lima, Perú. N. °2014-01.
- Zhang, J .; Howard, K .; Langston, C .; Kaney, B .; Qi, Y .; Tang, L .; Grams, H .; Wang, Y .; Gallos, S .; Martinaitis, S .; *et al.* 2016. Estimación de precipitación cuantitativa de sensores múltiples multiradar (MRMS): Capacidades iniciales de operación. *Toro. American Meteorological Society* (97):621–638. <https://doi.org/10.1175/BAMS-D-14-00174.1>

VIII. ANEXOS

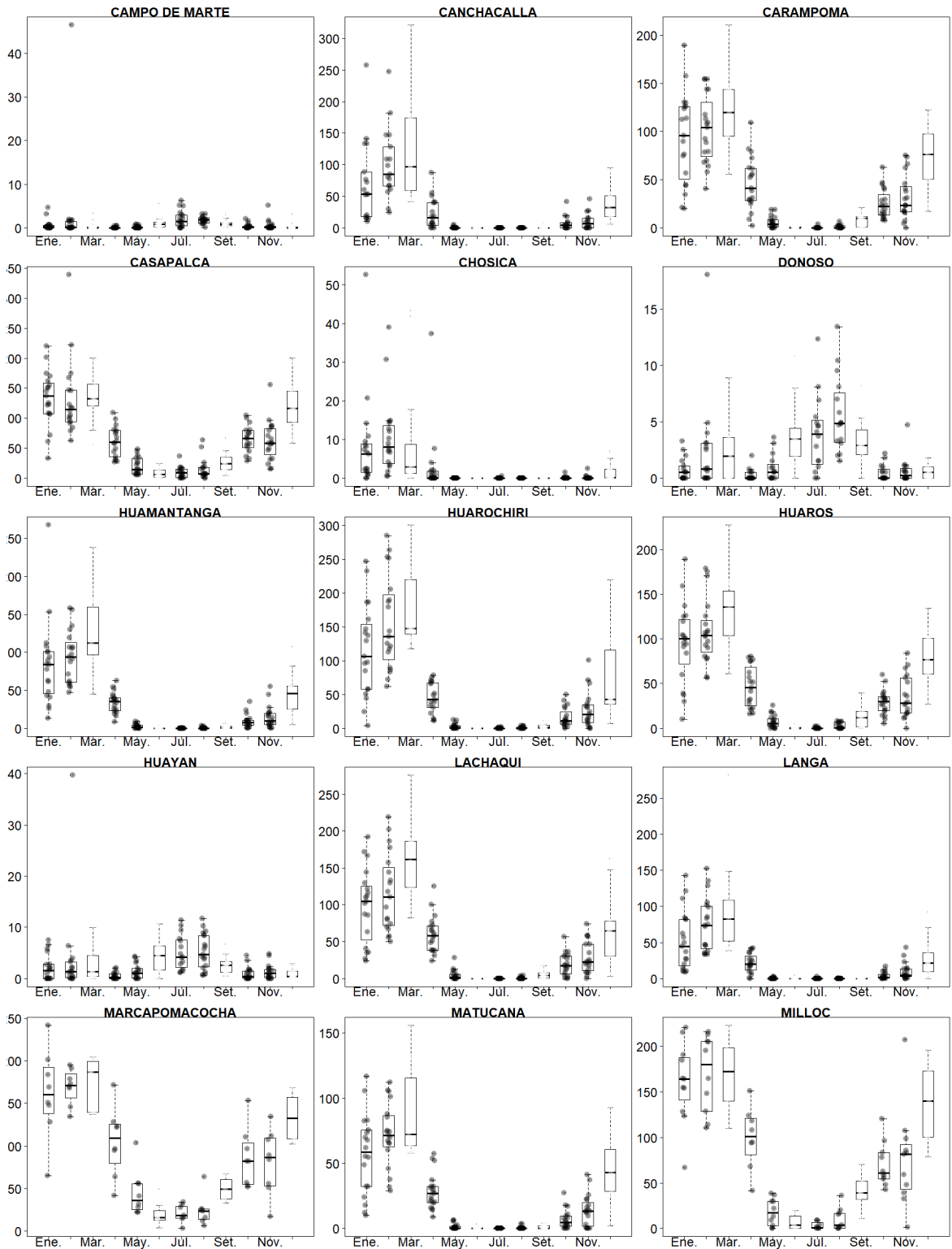
ANEXO 1: Información de estaciones pluviométricas disponibles

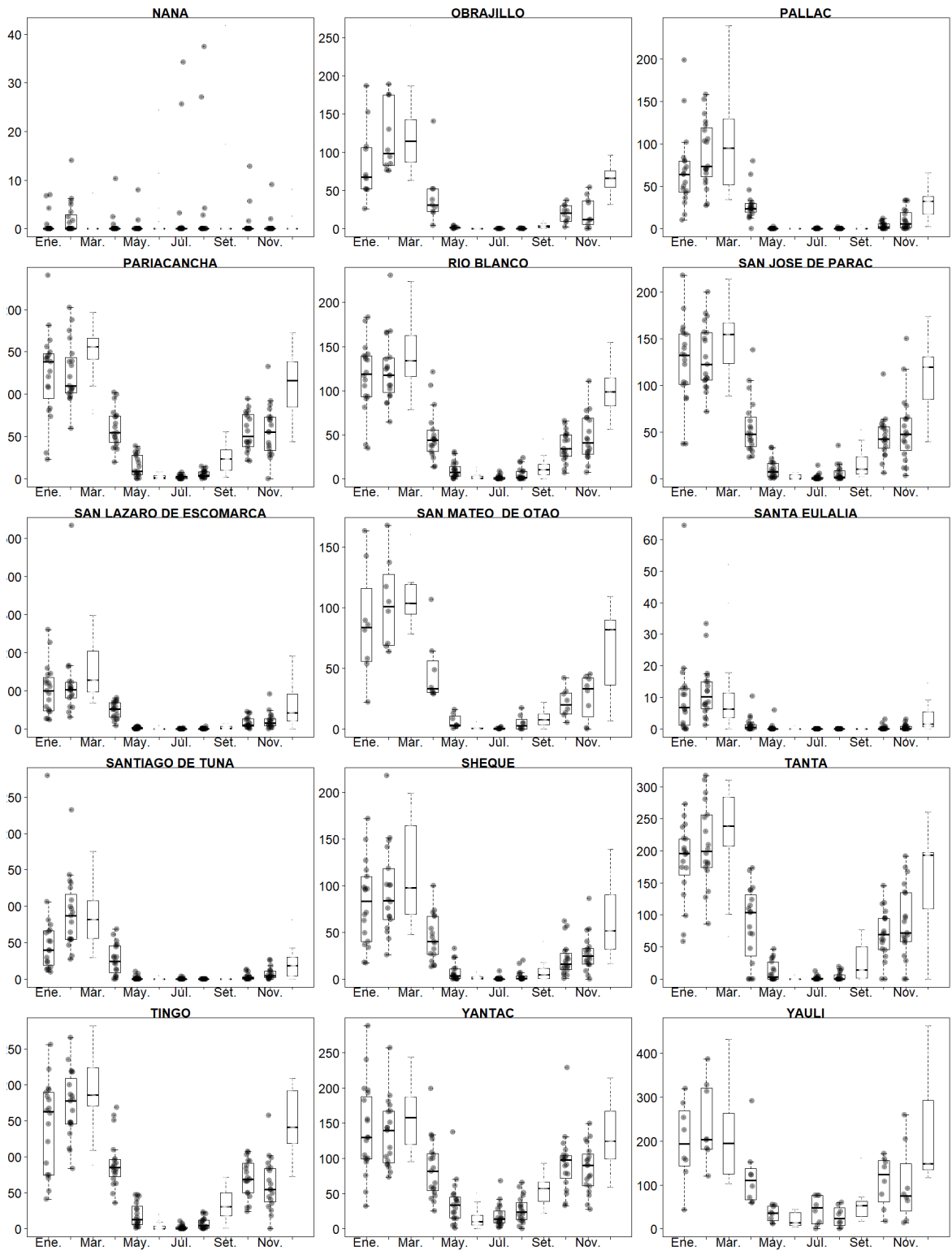
ID	CODIGO	NOMBRE	LONGITUD(°)	LATITUD(°)	ALTURA (msnm)	TIPO
1	112124	ANTIOQUIA	-76.500	-12.083	1839	Convencional
2	111057	ARAHUAY	-76.700	-11.617	2800	Convencional
3	111077	AUTISHA	-76.611	-11.738	2181	Convencional
4	112171	CAMPO DE MARTE	-77.043	-12.071	123	Convencional
5	111058	CANCHACALLA	-76.531	-11.845	2400	Convencional
6	111026	CANTA	-76.626	-11.471	2818	Convencional
7	111091	CARAMPOMA	-76.515	-11.655	3424	Convencional
8	111114	CASAPALCA	-76.233	-11.638	4233	Convencional
9	111060	CHOSICA	-76.690	-11.930	867	Convencional
10	111025	DONOSO	-77.236	-11.521	127	Convencional
11	111085	HUAMANTANGA	-76.750	-11.500	3392	Convencional
12	112055	HUAROCHIRI	-76.234	-12.139	3120	Convencional
13	111089	HUAROS	-76.576	-11.407	3569	Convencional
14	111022	HUAYAN	-77.122	-11.453	346	Convencional
15	111088	LACHAQUI	-76.628	-11.553	3670	Convencional
16	111028	LANGA	-76.421	-12.127	2863	Convencional
17	111028	MARCAPOMACOCHA	-76.325	-11.404	4447	Convencional
18	111027	MATUCA	-76.378	-11.839	2417	Convencional
19	111144	MILLOC	-76.350	-11.571	4384	Convencional
20	111023	ÑAÑA	-76.842	-11.988	543	Convencional
21	111159	OBRAJILLO	-76.622	-11.453	2468	Convencional
22	111083	PALLAC	-76.808	-11.349	2367	Convencional
23	111067	PARIACANCHA	-76.503	-11.394	3842	Convencional
24	111061	RIO BLANCO	-76.259	-11.735	3503	Convencional
25	111093	SAN JOSE DE PARAC	-76.258	-11.801	3829	Convencional
26	112126	SAN LAZARO DE ESCOMARCA	-76.350	-12.183	150	Convencional
27	111175	SAN MATEO DE HUANCHOR	-76.301	-11.760	3506	Convencional
28	111086	SANTA EULALIA	-76.667	-11.918	970	Convencional
29	111092	SANTIAGO DE TUNA	-76.517	-11.983	2924	Convencional
30	111062	SHEQUE	-76.502	-11.661	3188	Convencional
31	112130	TANTA	-76.017	-12.117	4323	Convencional
32	111066	TINGO	-76.483	-11.617	3980	Convencional

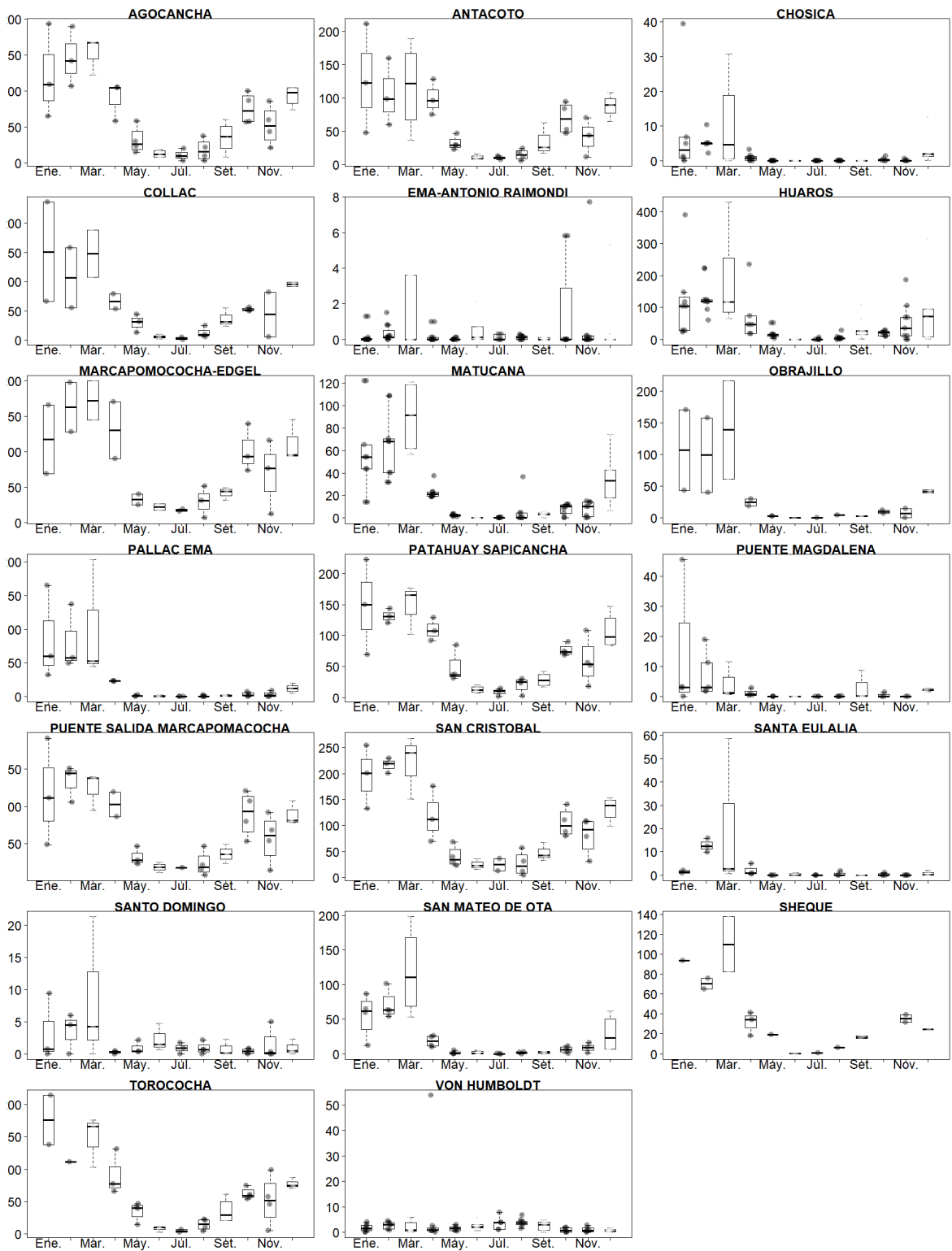
33	111076	YANTAC	-76.400	-11.333	4617	Convencional
34	111099	YAULI	-76.083	-11.667	4141	Convencional
1	472BF518	AGOCANCHA	-76.159	-11.711	4733	Automática
2	47278214	CHOSICA	-76.690	-11.930	867	Automática
3	472BD3F4	COLLAC	-76.470	-11.360	4493	Automática
4	472A218A	EMA-ANTONIO RAIMONDI.	-77.151	-11.776	47	Automática
5	4725F170	HUAROS	-76.567	-11.400	3585	Automática
6	47E8C3E8	MARCAPOMACOCHA - EDEGEL	-76.325	-11.404	4447	Automática
7	472CE45A	MATUCANA	-76.378	-11.839	241	Automática
8	47E9F488	OBRAJILLO	-76.630	-11.450	2710	Automática
9	47E31628	PALLAC EMA	-76.808	-11.349	2367	Automática
10	47E9E7FE	PATAHUAY SAPICANCHA	-76.383	-11.342	4535	Automática
11	47E22148	PUENTE CALLANTAMA	-76.657	-11.223	2740	Automática
12	4721725E	PUENTE MAGDALENA	-76.830	-11.680	935	Automática
13	472BE66E	PUENTE SALIDA MARCAPOMACOCHA	-76.320	-11.413	4435	Automática
14	472C07A8	SAN CRISTOBAL	-76.053	-11.740	4710	Automática
15	472E24B0	SANTA EULALIA	-76.667	-11.920	970	Automática
16	47E214D2	SANTO DOMINGO	-77.028	-11.370	620	Automática
17	472EC742	SAN MATEO DE OTAO	-76.564	-11.847	3513	Automática
18	472CD1C0	SHEQUE	-76.508	-11.653	3209	Automática
19	472C2144	TOROCOCHA	-76.443	-11.368	4431	Automática
20	472AC278	VON HUMBOLDT	-76.939	-12.082	247	Automática

FUENTE: Elaboración propia

ANEXO 2: Gráficos de boxplot de estaciones convencionales y automáticas en la región de estudio CHIRILU de precipitación promedio (mm/mes) para el registro histórico.







FUENTE: Elaboración propia

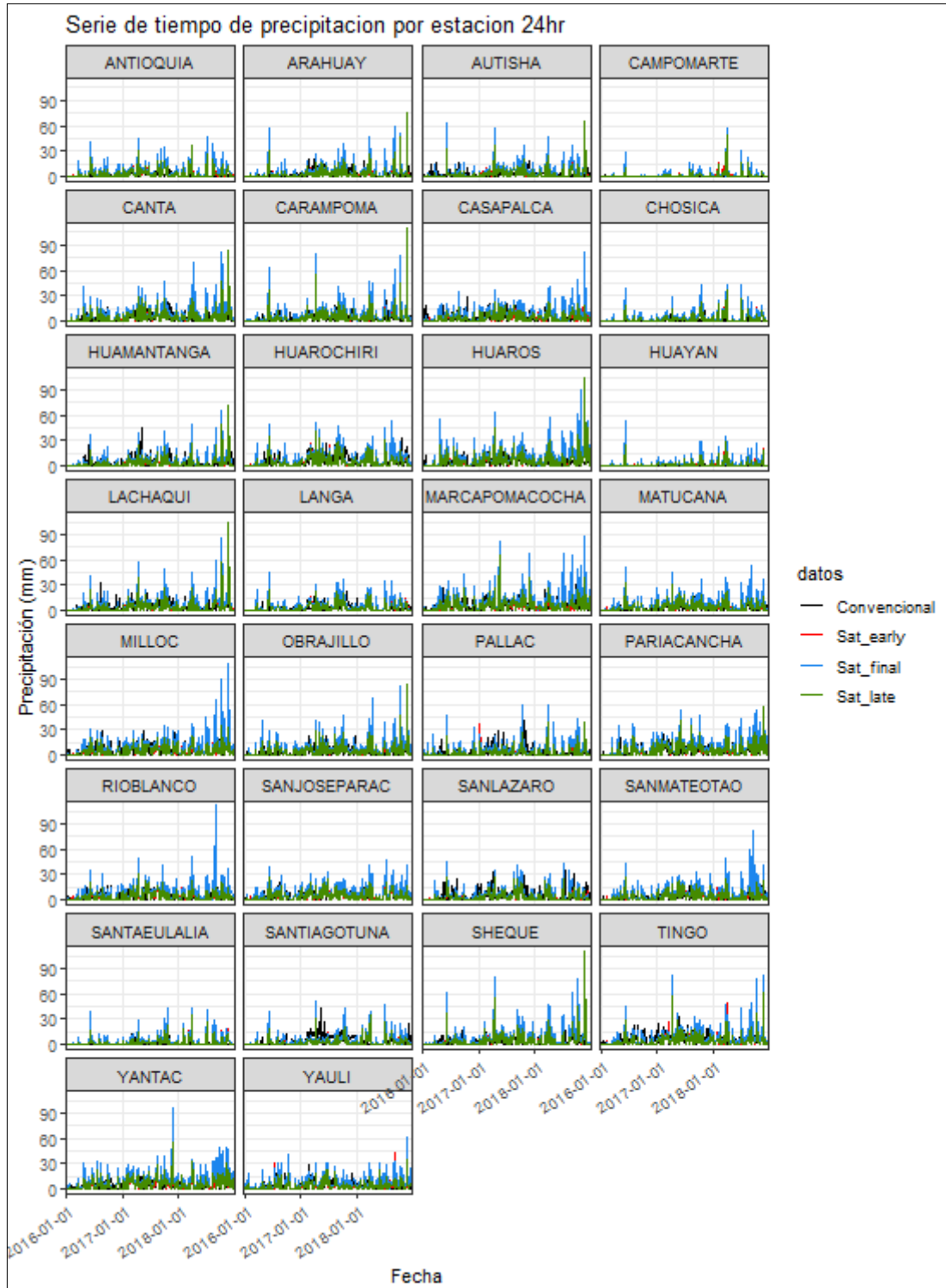
ANEXO 3: Información de estaciones pluviométricas seleccionadas

ID	CODIGO	NOMBRE	LONGITUD(°)	LATITUD(°)	ALTURA (msnm)	TIPO
1	112124	ANTIOQUIA	-76.500	-12.083	1839	Convencional
2	111057	ARAHUAY	-76.700	-11.617	2800	Convencional
3	111077	AUTISHA	-76.611	-11.738	2181	Convencional
4	112171	CAMPO DE MARTE	-77.043	-12.071	123	Convencional
5	111026	CANTA	-76.626	-11.471	2818	Convencional
6	111091	CAMPOMA	-76.515	-11.655	3424	Convencional
7	111114	CASAPALCA	-76.233	-11.638	4233	Convencional
8	111060	CHOSICA	-76.690	-11.930	867	Convencional
9	111085	HUAMANTANGA	-76.750	-11.500	3392	Convencional
10	112055	HUAROCHIRI	-76.234	-12.139	3120	Convencional
11	111089	HUAROS	-76.576	-11.407	3569	Convencional
12	111022	HUAYAN	-77.122	-11.453	346	Convencional
13	111088	LACHAQUI	-76.628	-11.553	3670	Convencional
14	111028	LANGA	-76.421	-12.127	2863	Convencional
15	111028	MARCAPOMACOCHA	-76.325	-11.404	4447	Convencional
16	111027	MATUCA	-76.378	-11.839	2417	Convencional
17	111144	MILLOC	-76.350	-11.571	4384	Convencional
18	111159	OBRAJILLO	-76.622	-11.453	2468	Convencional
19	111083	PALLAC	-76.808	-11.349	2367	Convencional
20	111067	PARIACANCHA	-76.503	-11.394	3842	Convencional
21	111061	RIO BLANCO	-76.259	-11.735	3503	Convencional
22	111093	SAN JOSE DE PARAC	-76.258	-11.801	3829	Convencional
23	112126	SAN LAZARO DE ESCOMARCA	-76.350	-12.183	150	Convencional
24	111175	SAN MATEO DE HUANCHOR	-76.301	-11.760	3506	Convencional
25	111086	SANTA EULALIA	-76.667	-11.918	970	Convencional
26	111092	SANTIAGO DE TUNA	-76.517	-11.983	2924	Convencional
27	111062	SHEQUE	-76.502	-11.661	3188	Convencional
28	111066	TINGO	-76.483	-11.617	3980	Convencional
29	111076	YANTAC	-76.400	-11.333	4617	Convencional
30	111099	YAULI	-76.083	-11.667	4141	Convencional
1	472BF518	AGOCANCHA	-76.159	-11.711	4733	Automática
2	4725F170	HUAROS	-76.567	-11.400	3585	Automática
3	472CE45A	MATUCANA	-76.378	-11.839	241	Automática
4	47E31628	PALLAC EMA	-76.808	-11.349	2367	Automática
5	47E9E7FE	PATAHUAY SAPICANCHA	-76.383	-11.342	4535	Automática
6	47E22148	PUENTE CALLANTAMA	-76.657	-11.223	2740	Automática
7	472C07A8	SAN CRISTOBAL	-76.053	-11.740	4710	Automática
8	472E24B0	SANTA EULALIA	-76.667	-11.920	970	Automática
9	47E214D2	SANTO DOMINGO	-77.028	-11.370	620	Automática
10	472C2144	TOROCOCHA	-76.443	-11.368	4431	Automática
11	472AC278	VON HUMBOLDT	-76.939	-12.082	247	Automática

FUENTE: Elaboración propia

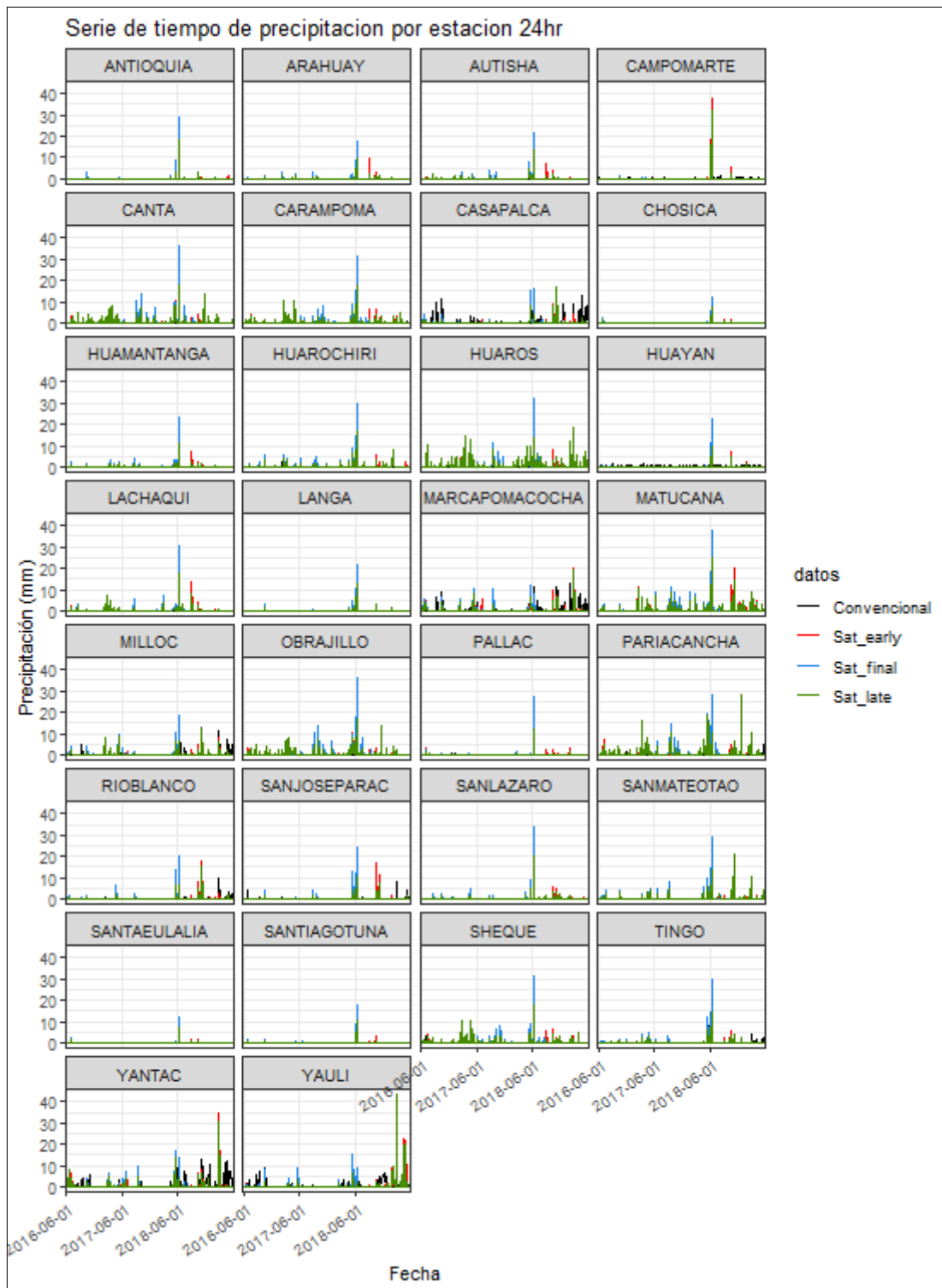
ANEXO 4: Gráficos de series de tiempo de estaciones convencionales y los tres productos satelitales IMERG-Early, IMERG-Late y IMERG-Final

I. Periodo: enero, febrero y marzo (EFM- lluvioso)



FUENTE: Elaboración propia

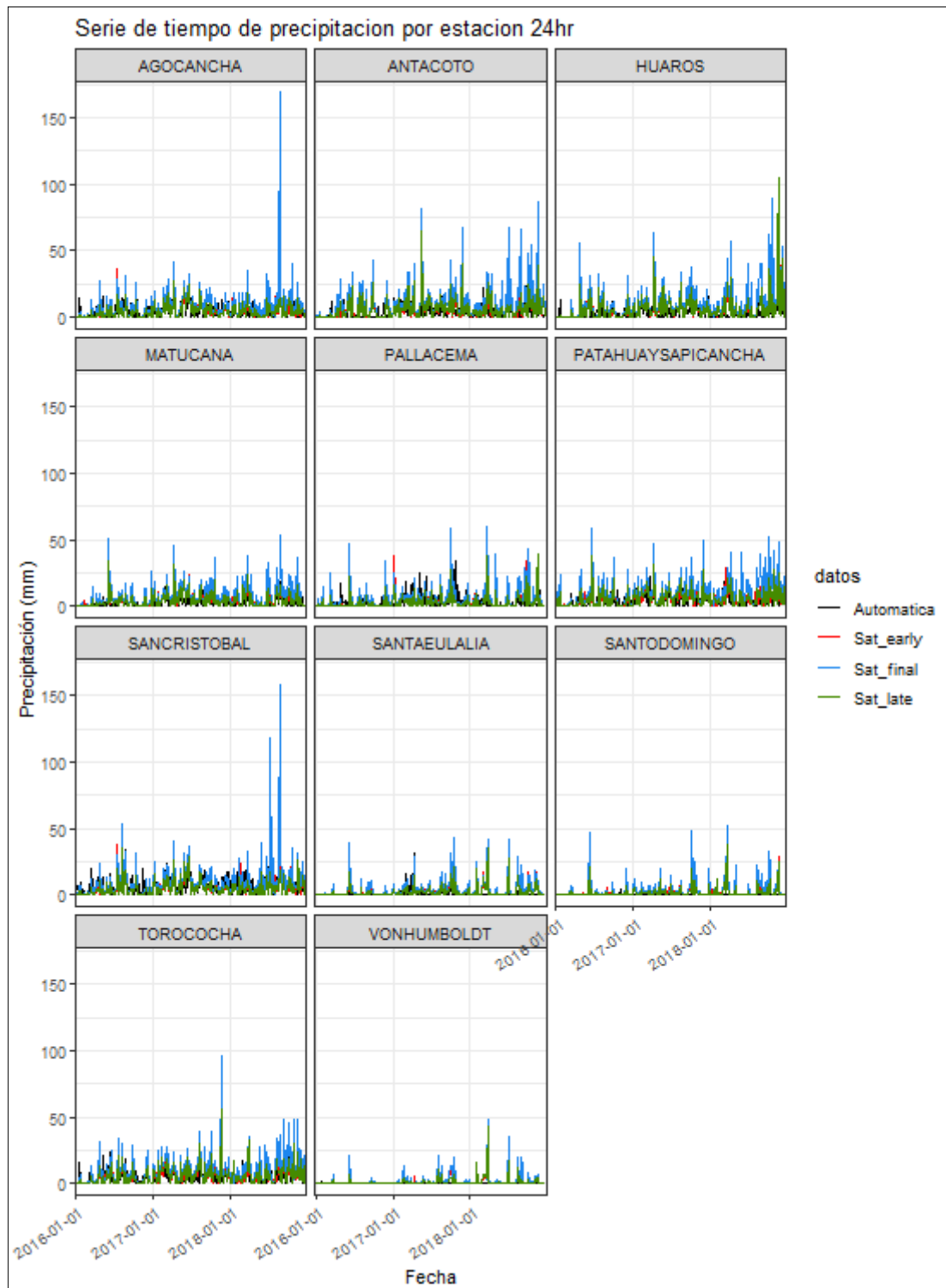
II. Periodo: junio, julio, agosto (JJA- seco)



FUENTE: Elaboración propia

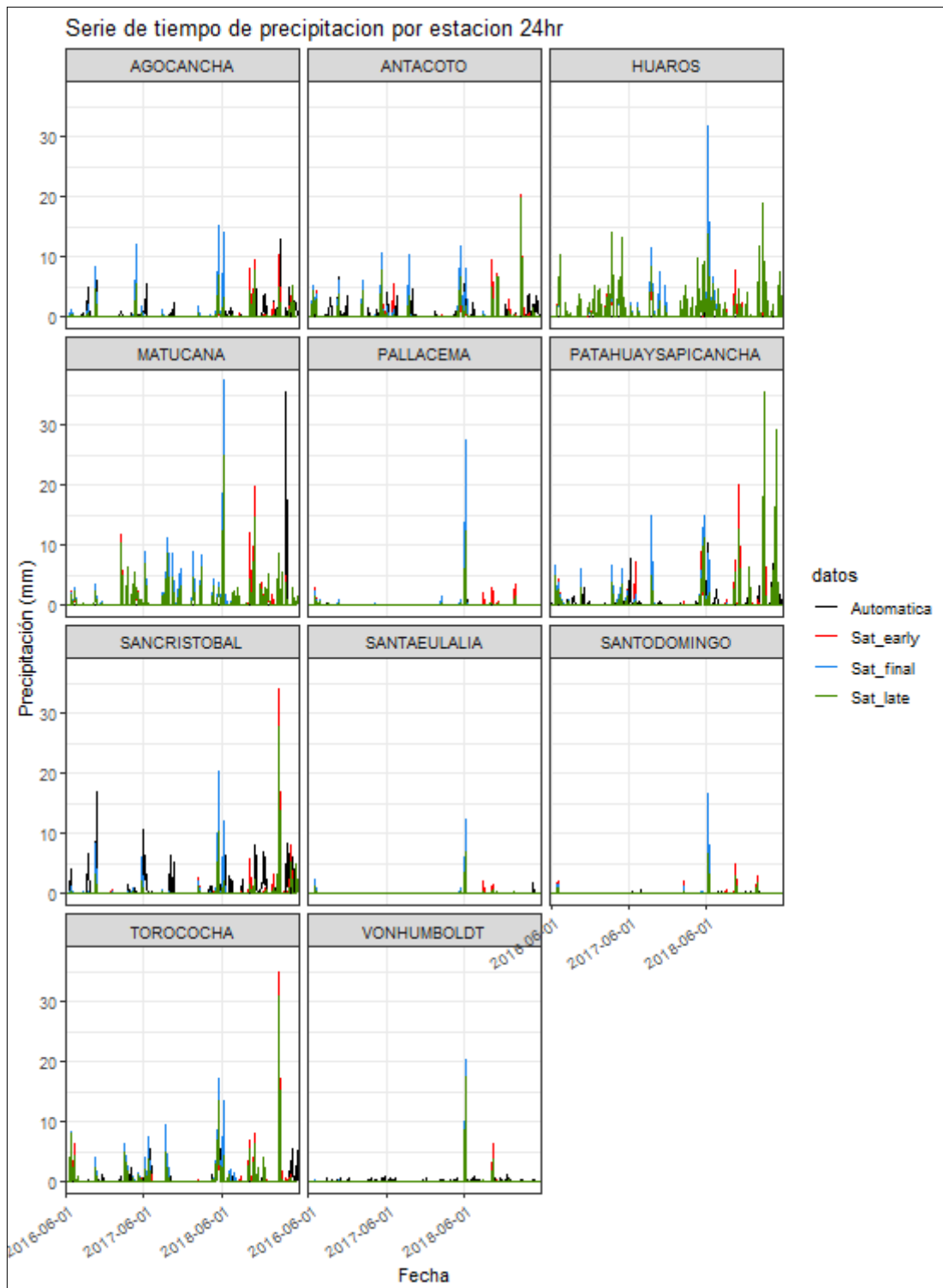
ANEXO 5: Gráficos de series de tiempo de estaciones automáticas y los tres productos satelitales IMERG-Early, IMERG-Late y IMERG-Final

I. Periodo: enero, febrero y marzo (EFM- lluvioso)



FUENTE: Elaboración propia

II. Periodo: junio, julio, agosto (JJA- seco)



FUENTE: Elaboración propia

ANEXO 6: Código en R para el cálculo del método Regression IDW

```
#####REGRESSION INVERSE DISTANCE WEITHING#####
#' @description This function use gstat packages for interpolate spatial point data
(\link[sp]{SpatialPoints})
#' and RasterLayer data (see \link[raster]{raster}).
#' @author Cesar Aybar <aybar1994@gmail.com>
#' @seealso \link[gstat]{idw}
#' @param gauge Is an object of SpatialPointsDataFrame class.
#' @param cov Is An object of RasterLayer.
#' @param idpR Is vector numeric of the power coeficient to evaluate.
#' @param formula that defines the dependent variable as a linear model
#' of independent variables; suppose the dependent variable has
#' name 'z', for Regression Inverse Distance Weigthing (RIDW) use the formula
#' 'z~x+y+....', you do not need define.
#' @param ... parameters that are passed on to \link[gstat]{variogram} variogram when calculating
the sample variogram
#' @details R_IDW use crossvalidation Leave-p-out cross-validation (LpO CV) and force brute
(optimize MSE)
#' for estimate the best idp power coeficient.
#' @return a List that contains: \code{Interpol} is the RIDW result in Raster,
#' \code{params} being \code{bestp} is the best distance weighting power,
#' \code{MSE} is the Residual Mean squared error of the residuals and
#' finally \code{linear_Model} is the adjusted linear Model.
#' @examples
#' library(raster)
#' library(Dorado)
#' data("Dorado")
#' gauge <- mean_doble_Station(gauge = Dorado$gauge,cov = Dorado$TRMM)
#' sat <- Dorado$TRMM
#' norm <- "log1p"
#' x <- RIDW(gauge = gauge,cov = sat,formula = PP_anual~Precipitación_Anual)
#' @importFrom automap autofitVariogram
#' @importFrom dplyr %>% tbl_df mutate_all
#' @importFrom raster extract projection writeRaster stack nlayers
#' @importFrom sp coordinates gridded
#' @importFrom gstat krige.cv idw
#' @importFrom methods as is
#' @importFrom stats lm median na.omit
#' @export

RIDW <- function(gauge, cov, formula, idpR = seq(0.8, 3.5, 0.1),norm) {

  # standardization -----
  if(norm=="log1p"){
    gauge@data[,1] <-log1p(gauge@data[,1])
    cov <-log1p(cov)}

  ext <- raster::extract(cov, gauge, cellnumber = F, sp = T)
  station <- gauge
  linear <- na.omit(ext@data) %>% tbl_df %>% mutate_all(as.character) %>%
  mutate_all(as.numeric)
  llm <- lm(formula,linear)
  station$residuals <- llm$residuals
```

```

# Define Grid -----
point <- rasterToPoints(cov) %>% data.frame
coordinates(point) <- ~x + y
projection(point) <- projection(cov)

# Estimate Best Parameter -----

idpRange <- idpR
mse <- rep(NA, length(idpRange))
for (i in 1:length(idpRange)) {
  mse[i] <- mean(krige.cv(residuals ~ 1, station, nfold = nrow(station), set = list(idp =
idpRange[i]), verbose = F,...)$residual^2)
}
poss <- which(mse %in% min(mse))
bestparam <- idpRange[poss]
residual.best <- krige.cv(residuals ~ 1, station, nfold = nrow(station), set = list(idp =
idpRange[poss]), verbose = F,...)$residual

# Interpolation -----

idwError <- idw(residuals ~ 1, station, point, idp = bestparam,...)
idwError <- idwError["var1.pred"]
gridded(idwError) <- TRUE
mapa <- raster(idwError)
namesF <- unlist(strsplit(as.character(formula), " "))
max_k <- floor(length(namesF)/2) + 1
name_cov = namesF[!namesF %in% c("~", "+", "-", "*", "/")][2:max_k]
cov <- cov[[name_cov]]

OBSp <- sum(stack(mapply(function(i) cov[[i]] * llm$coefficients[i + 1],
1:nlayers(cov)))) + llm$coefficients[1]
Ridw <- OBSp + mapa

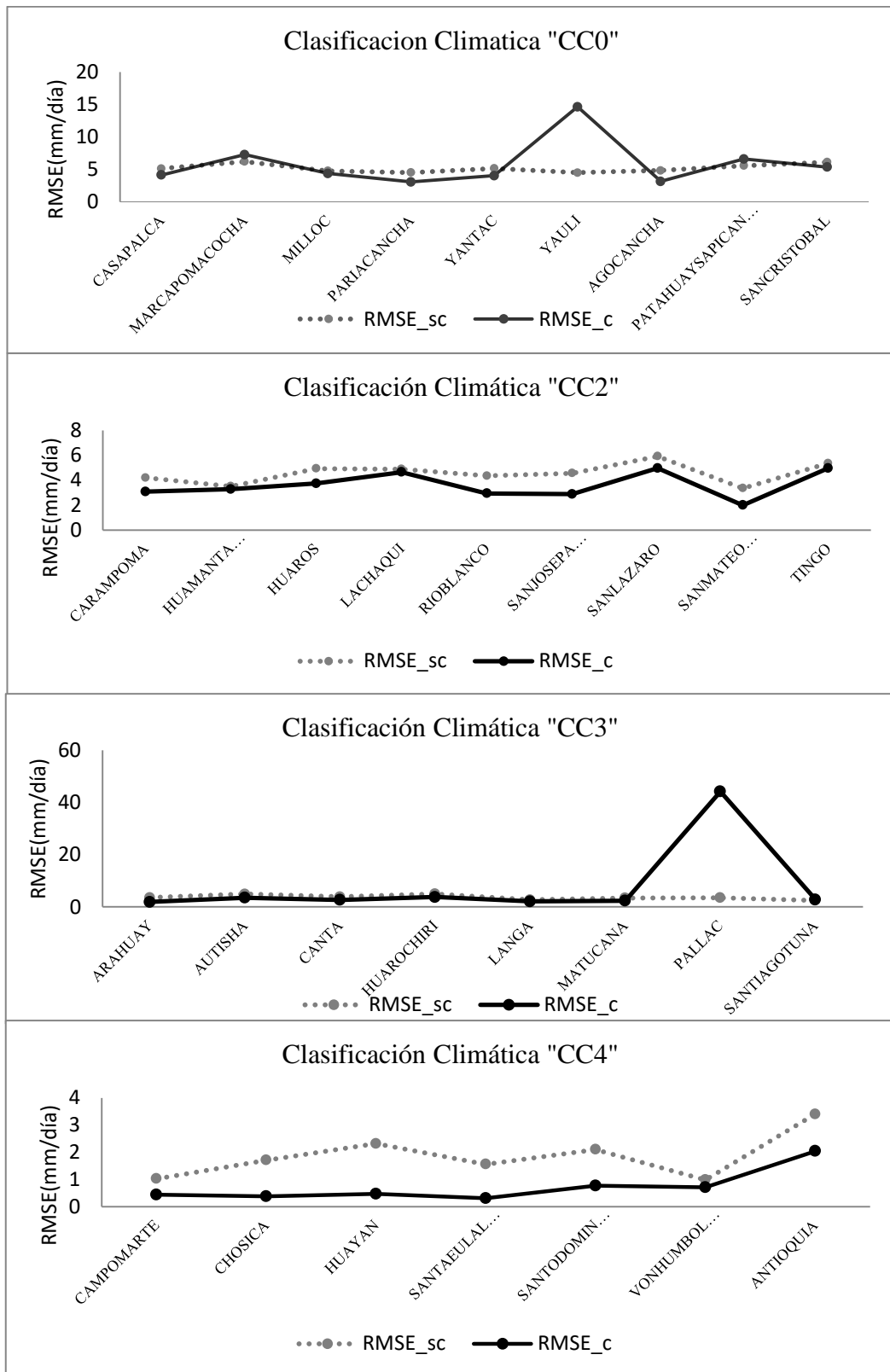
if(norm=="log1p"){ Ridw<-expm1(Ridw)
residual.best<-expm1(residual.best)}
Ridw[Ridw<0]=0

# Save Data -----
list(Interpol = Ridw, params = list(bestp = bestparam, rmse = sqrt(mean(residual.best^2)),
linear_Model = llm))
}

```

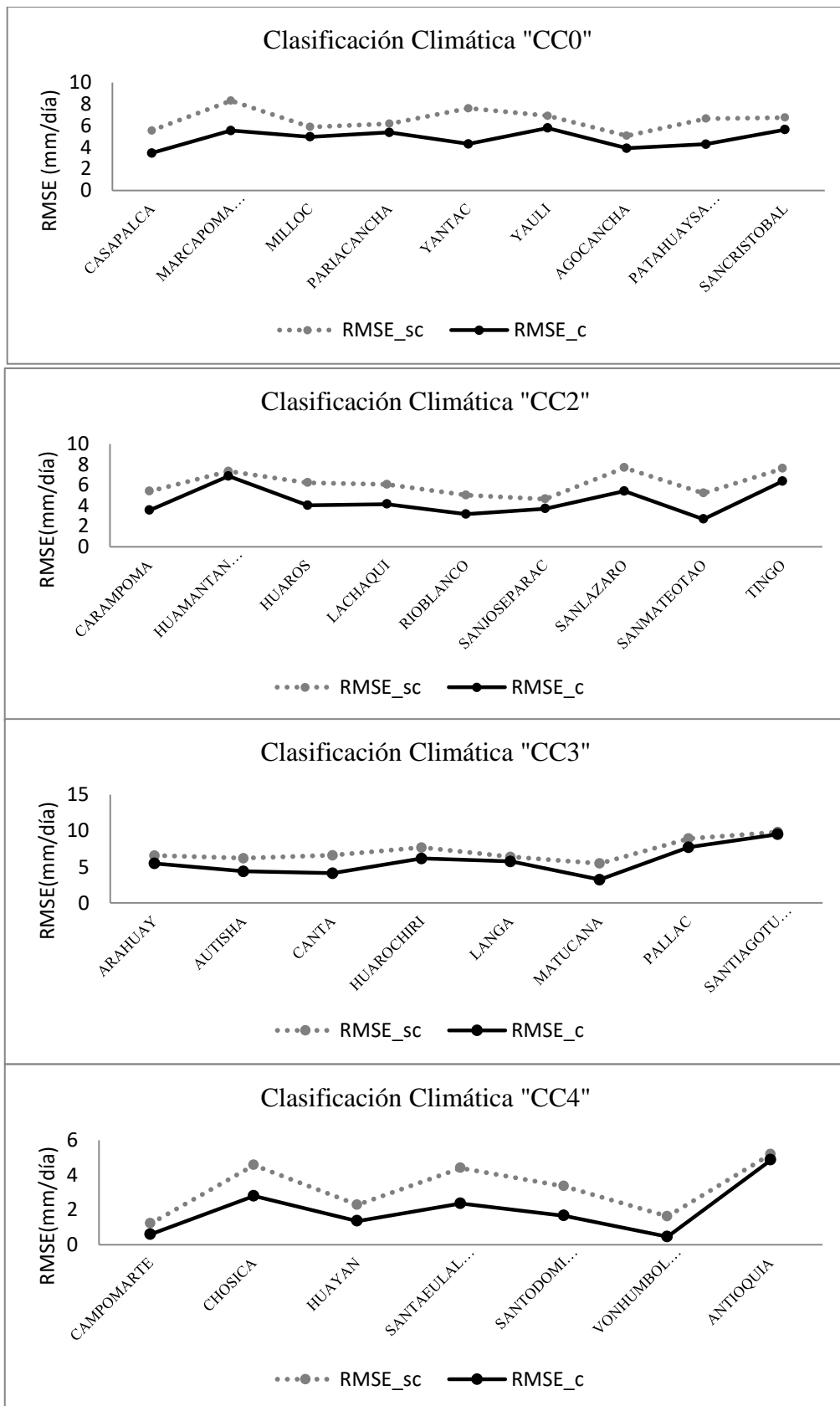
ANEXO 7: RMSE entre los datos observados y el producto satelital IMERG-late sin corregir (RMSE_sc) y la validación cruzada de IMERG-late corregido (RMSE_c).

I. Periodo lluvioso EFM-2016



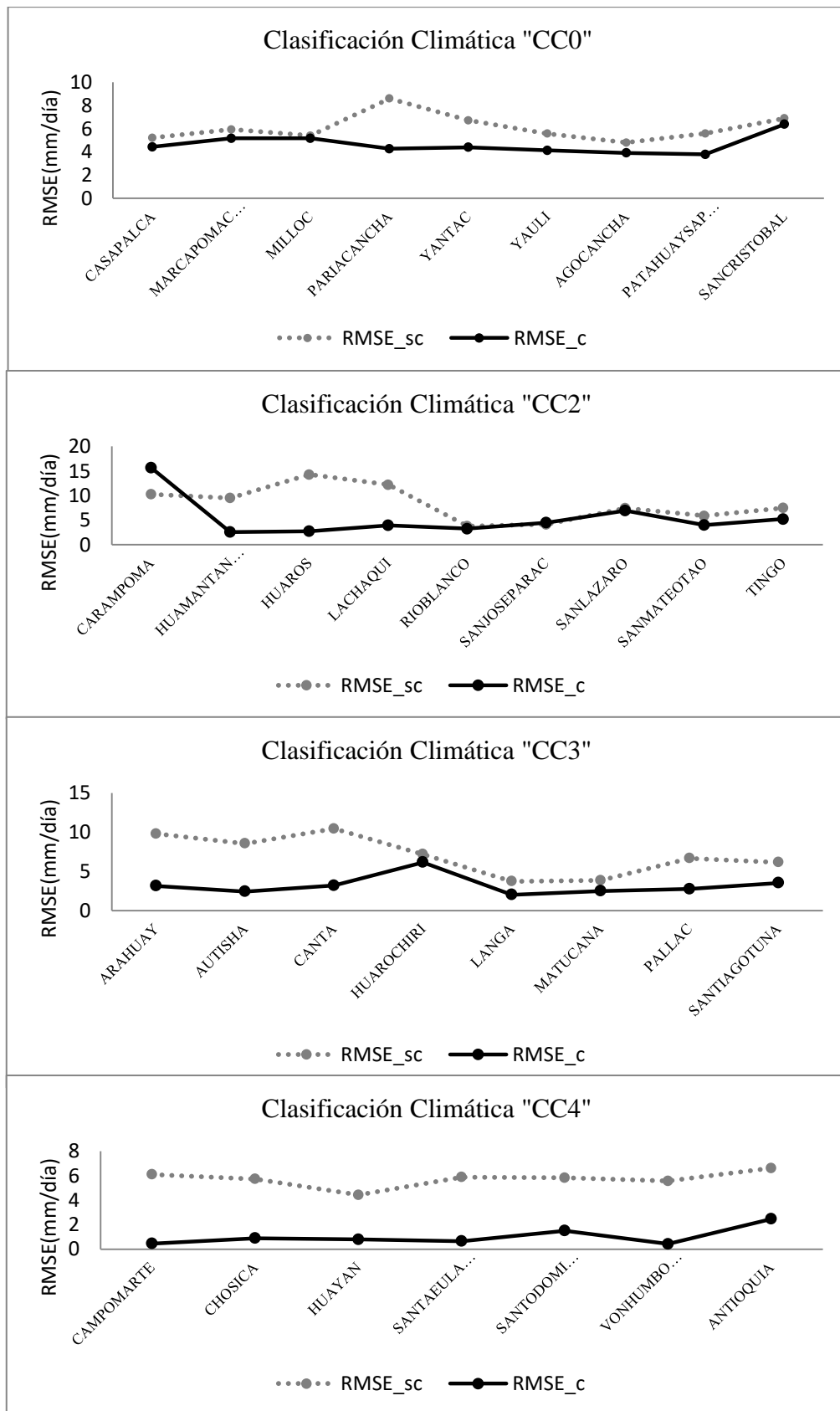
FUENTE: Elaboración propia

II. Periodo lluvioso EFM-2017



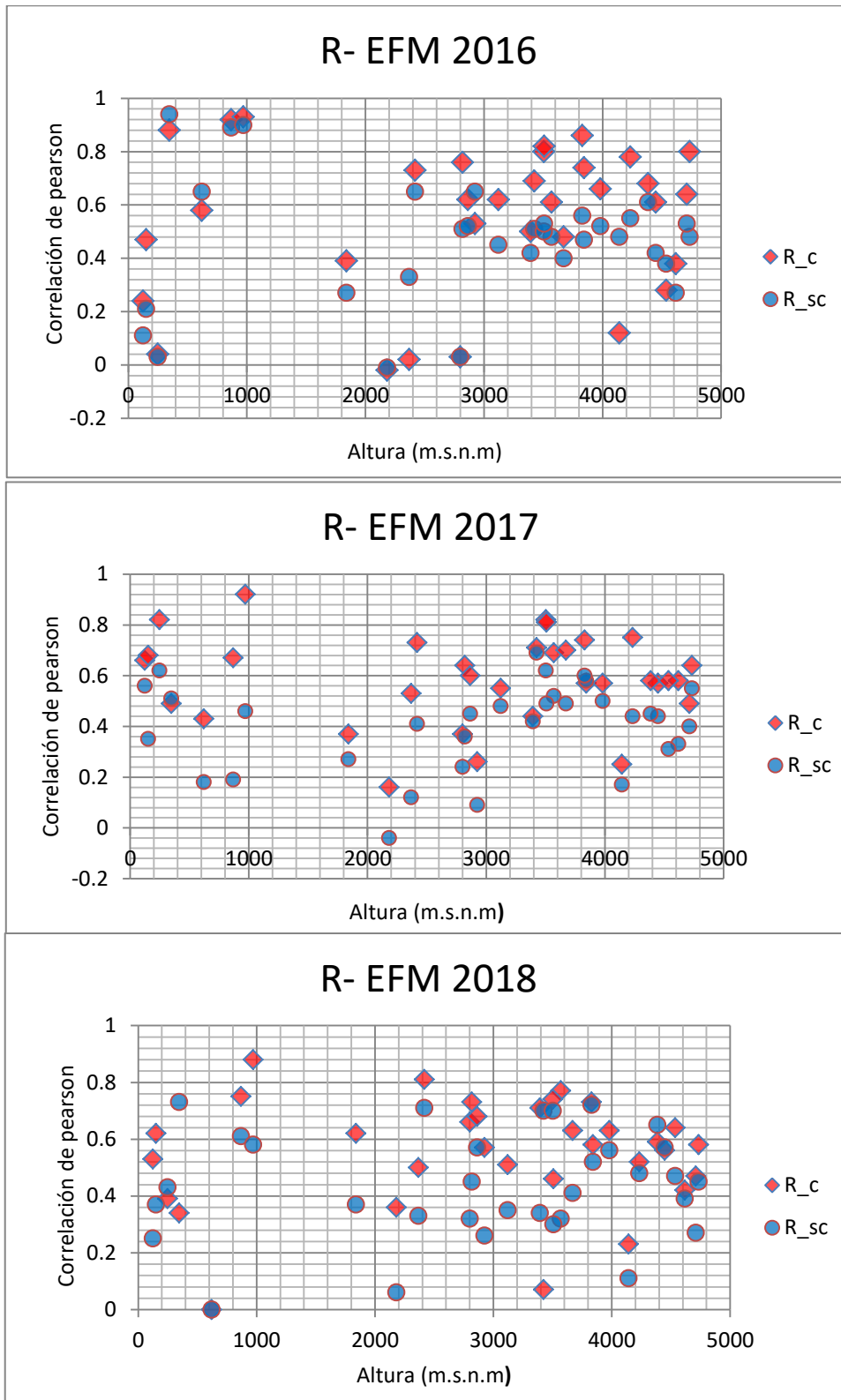
FUENTE: Elaboración propia

III. Periodo lluvioso EFM-2018

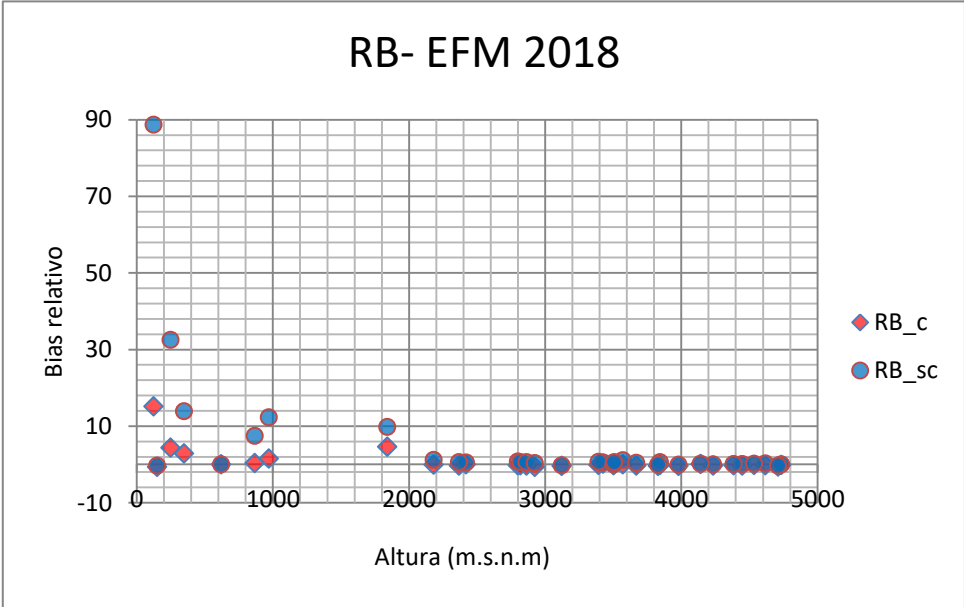
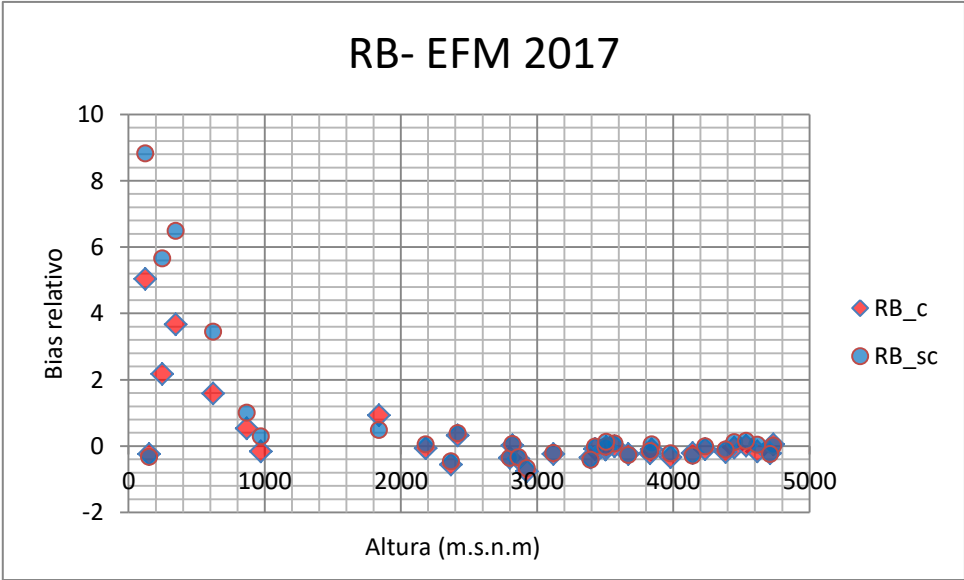
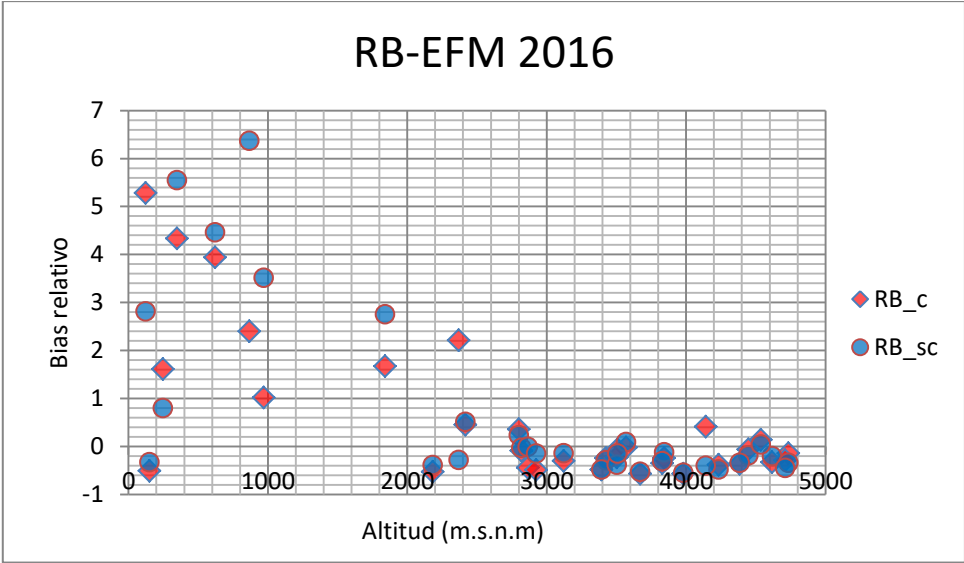


FUENTE: Elaboración propia

ANEXO 8: Distribución altitudinal de los estadísticos de correlación R y sesgo RB entre los datos observados y el producto satelital IMERG-Late sin corregir (R_sc y RB_sc) y la validación cruzada LOOCV de IMERG-late corregido (R_c y RB_c).

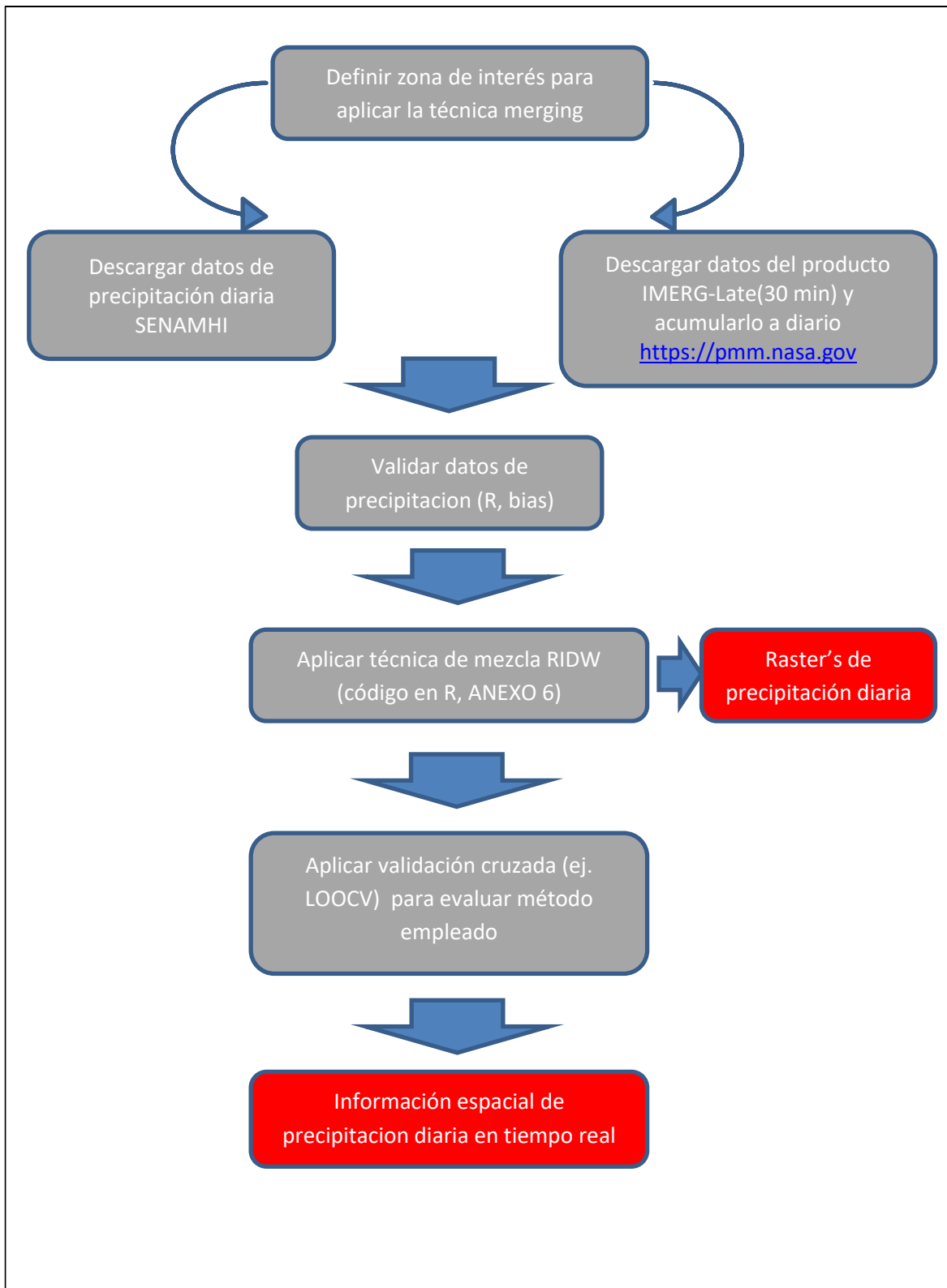


FUENTE: Elaboracion propia



FUENTE: Elaboración propia

ANEXO 9: Esquema resumen para la obtención del producto mezclado mediante la metodología planteada en la tesis



FUENTE: Elaboración propia