



UNIVERSIDAD  
**NACIONAL**  
DE COLOMBIA

# **Metodologías para el uso de la información de satélite como herramienta para rellenar series de datos diarios de lluvia en zonas de montaña colombianas**

**José Jacobo Jiménez Duque**

Universidad Nacional de Colombia  
Facultad de ingeniería y Arquitectura  
Departamento de Ingeniería Civil  
Maestría en Ingeniería – Recursos Hidráulicos  
Manizales, Colombia

2023

---

# **Methodologies for the use of satellite information as a tool for filling daily rainfall data series in Colombian mountainous areas**

**José Jacobo Jiménez Duque**

Tesis de investigación presentada como requisito para optar al título de:  
**Magister en Ingeniería – Recursos Hidráulicos**

Director:

Ph.D, Jorge Julián Vélez Upegui

Línea de Investigación:

Hidrología

Grupo de Investigación:

Grupo de Trabajo Académico en Ingeniería Hidráulica y Ambiental

Universidad Nacional de Colombia

Facultad de ingeniería y Arquitectura

Manizales, Colombia

2023

# Contenido

	Pág.
Contenido	
<b>Resumen</b> .....	<b>10</b>
<b>Abstract</b> .....	<b>11</b>
<b>1. Introducción</b> .....	<b>12</b>
1.1 Identificación del problema .....	13
1.2 Justificación.....	14
1.3 Zona de estudio .....	16
1.3.1 Modelo de elevación digital.....	17
1.3.2 Clima .....	18
1.3.3 Mapa de suelos .....	21
1.3.4 Coberturas del suelo.....	21
<b>2. Objetivos</b> .....	<b>23</b>
2.1 Objetivo General .....	23
2.2 Objetivos específicos .....	23
<b>3. Estado del arte</b> .....	<b>23</b>
3.1 Metodologías de relleno de series.....	25
3.1.1 Metodologías estadísticas y geoestadísticas .....	28
3.1.2 Metodologías estocásticas.....	31
3.1.3 Metodologías de inteligencia artificial.....	34
3.1.4 Metodologías híbridas.....	36
3.2 Productos satelitales de precipitación.....	36
3.2.1 The Climate Hazards group Infrared Precipitation with Stations (CHIRPS 2.0) 38	
3.2.2 Multi-Source Weighted-Ensemble Precipitation (MSWEP).....	40
3.2.3 Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) .....	40
<b>4. Metodología</b> .....	<b>41</b>
4.1 Metodología por actividades.....	41
4.1.1 Revisión de literatura .....	41
• Metodologías para relleno de series .....	41
• Información proveniente de sensores remotos .....	41
4.1.2 Recopilación y análisis de información .....	42
• Información de estaciones en tierra .....	42
• Información de productos satelitales de precipitación .....	42
4.1.3 Selección de las metodologías para relleno de series.....	42
4.1.4 Comparación de la información satelital y las estaciones en tierra .....	42
4.1.5 Relleno de series .....	43
4.1.6 Comparación y análisis de resultados.....	44
4.1.7 Metodologías propuestas.....	44

4	Metodologías para el uso de la información de satélite como herramienta para rellenar series de datos diarios de lluvia en zonas de montaña colombianas	
4.1.8	Preparación de un artículo	44
4.2	Esquema conceptual de la metodología	44
<b>5.</b>	<b>Resultados</b>	<b>46</b>
5.1	Recopilación y análisis de información	46
5.1.1	Series de precipitación de las estaciones en tierra del IDEAM	46
5.1.2	Series de precipitación de los productos satelitales	51
5.1.3	Definición de las zonas homogéneas climáticamente	61
5.2	Comparación de la información satelital y las estaciones en tierra	66
5.2.1	Comparación estadística de la información satelital (magnitud de las precipitaciones)	66
5.2.2	Comparación estadística de la información satelital (detección de los eventos de precipitación)	69
5.3	Técnicas de relleno de datos faltantes de lluvia diaria	73
5.3.1	Selección de las metodologías para el relleno de series	73
5.3.2	Información utilizada en las metodologías de relleno	74
5.3.3	Periodos de calibración y validación	75
5.3.4	Técnicas utilizadas para el relleno de datos faltantes de lluvia diaria	77
5.4	Relleno de series	83
5.4.1	Metodologías estadísticas	84
5.4.2	Metodología estocástica	87
5.4.3	Metodología de inteligencia artificial	91
5.4.4	Resumen metodologías de relleno	92
5.5	Metodologías propuestas	102
<b>6.</b>	<b>Conclusiones</b>	<b>103</b>
<b>7.</b>	<b>Recomendaciones y trabajos futuros</b>	<b>106</b>
<b>8.</b>	<b>Bibliografía</b>	<b>107</b>
<b>9.</b>	<b>Anexos</b>	<b>118</b>
9.1	Series utilizadas	118
9.2	Programas en Python	118
9.3	Gráficas de la comparación de la precipitación media mensual entre los años 1982 y 1986 para las estaciones del IDEAM y los productos satelitales	118
9.4	Curvas de masa doble	118
9.5	Curvas de masa simple	119
9.6	Póster presentando en 2nd LATIN AMERICAN & CARIBBEAN YOUNG WATER PROFESSIONALS CONFERENCE	119
9.7	Certificado ponencia XXIX CONGRESO LATINOAMERICANO DE HIDRÁULICA – MÉXICO 2021	120
9.8	Póster presentando en Spatial Statistics 2023: Climate and the Environment	121

---

## Lista de figuras

Figura 1. Zona de estudio, departamento de Caldas (zona cafetera colombiana). Fuente propia. ....	17
Figura 2. Modelo de elevación digital (MED). Fuente: (ASF, n.d.) .....	18
Figura 3. Precipitación media total anual. Adaptado de (CORPOCALDAS, 2001).....	19
Figura 4. Mapa de zonificación climática. Fuente: (IGAC, 1998) .....	20
Figura 5. Ciclo anual promedio de la precipitación en diversas estaciones localizadas en los Andes de Colombia. Fuente: (Poveda et al., 2002) .....	20
Figura 6. Mapa de suelos. Fuente: (IGAC, 1998) .....	21
Figura 7. Usos del suelo departamento de Caldas. Fuente: (IDEAM, 2014) .....	22
Figura 8. Redes de perceptrones multicapa. Tomada de (Coulibaly & Evora, 2007). .....	35
Figura 9. Metodología punto a pixel usada para comparar los productos satelitales de precipitación con respecto a las estaciones en tierra. Fuente propia. ....	43
Figura 10. Esquema conceptual de la metodología. ....	45
Figura 11. Localización de las estaciones en tierra del IDEAM utilizadas. Fuente propia. ....	47
Figura 12. Completitud de datos de las estaciones en tierra del IDEAM. Fuente propia. ....	53
Figura 13. Estaciones sin datos faltantes entre 1982 y 1986. Fuente propia. ....	55

Figura 14. Precipitación media mensual entre los años 1982 a 1986 en la ubicación de la estación pluviométrica Cenicafé. Fuente propia. ....	56
Figura 15. Precipitación media mensual entre los años 1982 a 1986 en la ubicación de la estación pluviométrica Sub Dorada. Fuente propia. ....	57
Figura 16. Curvas de masa simple. Fuente propia. ....	58
Figura 17. Curva de doble masa de precipitación entre la estación NORCASIA RADIO [23050250] y CHIRPS V2.0, entre los años 1982 y 1986 .....	60
Figura 18. Curva de doble masa de precipitación entre la estación ARAUCA [26150060] y CHIRPS V2.0, entre los años 1982 y 1986.....	61
Figura 19. Relación entre la altitud de las estaciones de la vertiente del Cauca y precipitación media anual entre los años 1982 y 1986 .....	63
Figura 20. Relación entre la altitud de las estaciones de la vertiente del Magdalena y precipitación media anual entre los años 1982 y 1986 .....	64
Figura 21. Resultados por mes del desempeño de los productos satelitales en la estimación de la magnitud de las precipitaciones diarias entre 1982 y 1986. Fuente propia. ....	68
Figura 22. Resultados por mes del desempeño de los productos satelitales en la detección de eventos diarios de precipitación para el periodo entre 1982 y 1986. Fuente propia. ....	72
Figura 23. Índice ONI entre 1990 y 2013. Fuente: (NOAA-CPC, 2022) .....	77
Figura 24. Estaciones utilizadas para aplicar las metodologías de relleno.....	79
Figura 25. Curvas de doble masa de precipitación entre las estaciones y CHIRPS V2.0 entre los años 1982 y 1986 .....	84
Figura 26. Ajustes de las funciones de distribución de probabilidad Gumbel, Generalized Extreme Value (GEV), Log Normal, Weibull y Exponencial para la estación AEROPUERTO LA NUBIA para el periodo 1994 a 2003 .....	88
Figura 27. Desempeño en la estimación de la magnitud de las precipitaciones diarias para el periodo 2004 a 2013.....	93
Figura 28. Desempeño en la detección de eventos diarios de precipitación del relleno de series para el periodo 2004 a 2013 .....	95

---

Figura 29. Mapa del ME (mm) para la metodología MLP con las diferentes fuentes de información, para el periodo 2004 a 2013. La resolución de cuadrícula es de $0.05^\circ$ .....	97
Figura 30. Diagrama de caja y bigotes de las series rellenas con las metodologías estadísticas y MLP, para el periodo 2004 a 2013 .....	98
Figura 31. Diagrama de caja y bigotes de las series rellenas con la metodología Cadenas de Markov de primer orden, para el periodo 2004 a 2013 .....	98
Figura 32. Mapa del promedio de las series rellenas con la metodología MLP, para el periodo 2004 a 2013 .....	100
Figura 33. Mapa de la desviación estándar de las series rellenas con la metodología MLP, para el periodo 2004 a 2013.....	101

## Lista de tablas

Tabla 1. Funciones de distribución de probabilidad.....	33
Tabla 2. Estaciones en tierra del IDEAM utilizadas .....	47
Tabla 3. Resumen características de los productos satelitales descargados de forma gratuita en la red de redes.....	51
Tabla 4. Estaciones sin datos faltantes entre 1982 y 1986 .....	54
Tabla 5. Resultados test de Man Kendall para las 25 estaciones sin datos faltantes entre 1982 y 1986. ....	59
Tabla 6. Estaciones utilizadas para el análisis de la relación altura vs precipitación.....	62
Tabla 7. Estadísticos de comparación por pares. Fuente: basado en (Toté et al., 2015)	66
Tabla 8. Resultados de los productos satelitales en la estimación de la magnitud de las precipitaciones diarias entre 1982 y 1986.....	67
Tabla 9. Estadísticos para evaluar la detección de eventos de precipitación. Fuente: basado en (Toté et al., 2015).....	69
Tabla 10. Resultados del desempeño de los productos satelitales en la detección de eventos diarios de precipitación para el periodo entre 1982 y 1986.....	70
Tabla 11. Estaciones utilizadas aplicar las metodologías de relleno.....	78
Tabla 12. Desempeño promedio de las metodologías NRM, SBE e IDW con respecto a la magnitud de las precipitaciones diarias para el periodo 2004 a 2013 .....	85
Tabla 13. Desempeño promedio de las metodologías NRM, SBE e IDW con respecto a la detección de eventos diarios de precipitación para el periodo 2004 a 2013 .....	86
Tabla 14. Desempeño promedio en la estimación de la magnitud de las precipitaciones diarias utilizando la metodología Cadenas de Markov de primer orden para el periodo 2004 a 2013 .....	89



---

Tabla 15. Desempeño promedio en la detección de eventos diarios de precipitación del relleno de series utilizando la Cadenas de Markov de primer orden para el periodo 2004 a 2013 .....	90
Tabla 16. Desempeño promedio en la estimación de la magnitud de las precipitaciones diarias utilizando la metodología MLP para el periodo 2004 a 2013 .....	91
Tabla 17. Desempeño promedio en la detección de eventos diarios de precipitación del relleno de series utilizando la metodología MLP para el periodo 2004 a 2013.....	91

## Resumen

Los datos faltantes de lluvia diaria afectan la estimación de eventos extremos, los resultados en la gestión del agua y la planificación de los recursos hídricos. Se han implementado diferentes técnicas para abordar este problema, en donde las estrategias más utilizadas implican análisis estadísticos, enfoques estocásticos y técnicas de inteligencia artificial, así como los métodos híbridos que mezclan las estrategias mencionadas anteriormente. La precipitación en la zona andina tropical se ve afectada por este efecto de datos faltantes, ya que apenas se monitorea hasta cierta altitud, y la distribución espacial de las estaciones medidoras de lluvia en tierra es muy dispersa espacialmente y en algunos casos con muchos datos faltantes. Sin embargo, los productos satelitales de precipitación cubren toda el área, dan valores continuos en el tiempo y esta información se puede descargar de forma gratuita desde diferentes bases de datos. Por lo tanto, esta tesis explora el uso de los productos satelitales de precipitación combinados con metodologías tradicionales para mejorar las estrategias de relleno de datos faltantes de las series de lluvia diaria en el departamento de Caldas. Se propone un análisis que compara la eficiencia de los diferentes métodos como las redes neuronales artificiales, los análisis estadísticos clásicos y los métodos estocásticos, con o sin datos de productos satelitales de precipitación y con la combinación de ambas fuentes de información (estaciones en tierra y productos satelitales de precipitación). Los resultados principales revelan que la incorporación de CHIRPS V2.0 junto con la información proveniente de las estaciones en tierra en la zona de estudio no produce cambios significativos en los resultados obtenidos mediante las metodologías aplicadas. Además, se observa que el desempeño de las metodologías con el uso exclusivo la información de CHIRPS V2.0 es inferior en comparación con el uso de las estaciones en tierra. Se destaca como mejor metodología para el relleno de series de lluvia diaria en el departamento de Caldas, las redes de perceptrones multicapa (MLP) cuando se utilizan únicamente las estaciones en tierra como fuente de información.

**Palabras clave:** sensores remotos, relleno de series, precipitación diaria, inteligencia artificial y métodos estocásticos.

## Abstract

The gap-filling rainfall data affects the estimation of extreme events, the results in water management and water resource planning. Different techniques have been implemented to address this problem, where the most used strategies involve statistical analysis, stochastic approaches and artificial intelligence techniques, as well as hybrid methods that mix the strategies mentioned above. Precipitation in the tropical Andean zone is affected by this missing data effect, since it is barely monitored up to a certain altitude, and the spatial distribution of ground-based rain gauge stations is very spatially dispersed and in some cases with many missing data. However, satellite precipitation products cover the entire area, provide continuous values over time, and can be freely downloaded from different databases. Therefore, this thesis explores the use of use of satellite precipitation products combined with traditional methodologies to improve strategies to fill missing data in the daily rainfall series in the Department of Caldas. An analysis is proposed to compare the effectiveness of different methods, such as artificial neural networks, classical statistical analyses, and stochastic methods, both with and without satellite precipitation data, as well as the combination of both data sources (ground-based stations and satellite precipitation products). The main results indicate that the incorporation of CHIRPS V2.0 along with information from ground-based stations in the study area does not yield significant changes in the results obtained from the applied methodologies. Furthermore, it is observed that the performance of methodologies using exclusively CHIRPS V2.0 data is inferior compared to the use of ground-based stations. Notably, the multilayer perceptron (MLP) neural networks with only ground-based stations as the source of information emerge as the superior methodology for filling in daily rainfall series in the Department of Caldas.

**Keywords:** remote sensors, data filling, daily precipitation, artificial intelligence, and stochastic methods.

# 1.Introducción

El ciclo hidrológico representa el mayor flujo de agua que existe en la tierra, además tiene y ha tenido una influencia determinante en la evolución de la tierra. La precipitación y evapotranspiración son las variables más importantes en el ciclo hidrológico global, siendo la precipitación muy variable, lo cual tiene gran influencia en la vegetación, las sequías e inundaciones y un efecto de control sobre la gran escala de la circulación y la atmósfera (Chahine, 1992).

Para el conocimiento de la hidrología en diferentes escalas espaciales y temporales es necesario tener buena cantidad y calidad de registros hidrológicos tanto en su longitud como en su cobertura espacial, en las que se tienen varias aplicaciones fundamentales como lo son el planeamiento para el desarrollo sostenible del territorio y de sus recursos hidráulicos, la prevención de desastres, un mejor conocimiento de la influencia de los fenómenos macro-climáticos para desarrollar mejores predicciones, o para evaluar los posibles impactos del cambio climático global (Hurtado Montoya & Mesa Sánchez, 2014).

Los datos de precipitación global con alta resolución y precisión son el insumo de mayor interés para el estudio del ciclo hidrológico global y de los recursos hídricos en general (Okamoto et al., 2005). Debido a la importancia de las series de precipitación existen múltiples metodologías para estimar los datos faltantes de las series de precipitación, y particularmente en este trabajo el objetivo de crear estrategias para rellenar series de lluvia diaria con el uso de información de satélite en zonas de montaña colombianas (departamento de Caldas, ubicado en la zona cafetera).

## 1.1 Identificación del problema

Generalmente para la medición de la precipitación global se utilizan métodos convencionales como lo son las redes de estaciones pluviométricas y en ocasiones sistemas de radares meteorológicos (Kidd & Levizzani, 2011). A veces no se cuenta con información completa, ya que se presentan situaciones como equipos descalibrados, deterioro en unidades de medición, entre otros. Además, se presentan situaciones que pueden afectar la calidad de los registros de precipitación como lo son el movimiento de instrumentos, transformación del entorno de la estación, o no registrar el dato en el momento adecuado, lo que conlleva a la inconsistencia o ausencia de datos, provocando problemas a la hora de realizar estudios confiables, ya que los estudios hidrológicos y de planificación y gestión del agua requieren series históricas de clima completas (Rivera, 2008). Estos valores faltantes son la principal fuente de problemas e incertidumbres al analizar y modelar la variabilidad espacial y temporal de la variable. Uno de los elementos que hace más difícil la captura confiable de registros hidrológicos en Colombia es su compleja fisiografía caracterizada por los tres ramales de la Cordillera de los Andes (Hurtado Montoya & Mesa Sánchez, 2014), que es donde se encuentra la zona de estudio, ubicada en el departamento de Caldas; zona cafetera colombiana.

El relleno de datos de lluvia permite reducir la incertidumbre en estudios relacionados con análisis de homogeneidad, conglomerados, tendencias, estadísticas multivariadas y pronósticos meteorológicos en regiones con déficit de información, cambio climático, variabilidad climática y las series completadas sirven para la planificación y gestión del recurso agua en el territorio.

En la actualidad las estimaciones cuantitativas de precipitación mediante satélites son ampliamente utilizadas, sin embargo, este tipo de información tiene limitaciones como la deficiencia en la observación, la metodología de evaluación y la selección de ventanas de tiempo cortos para la selección (Bai et al., 2018). La información satelital carece de estudios sobre las incertidumbres asociadas a la confiabilidad de este tipo de fuentes (Aghakouchak et al., 2011), y en general se identifican los siguientes problemas:

14 Metodologías para el uso de la información de satélite como herramienta para rellenar series de datos diarios de lluvia en zonas de montaña colombianas

---

- a. Generalmente los registros hidrológicos son escasos, tanto en su calidad, longitud y cobertura espacial, por lo que en ocasiones no se cuenta con información completa.
- b. Colombia presenta una complejidad geográfica y climática la cual hace difícil la medición de registros hidrológicos. La zona de estudio es compleja orográficamente por tratarse de una zona de montaña en la zona cafetera colombiana.
- c. Los datos de satélite son susceptibles a sesgos sistemáticos, por lo que se debe tener especial cuidado a la hora de utilizar este tipo de información con el fin de hacer estudios de alta confiabilidad.

Por lo tanto, dado que es usual que existan vacíos en las series temporales de lluvia diaria en las diferentes redes de monitoreo de la región, se hace necesario entender las diferentes estrategias disponibles para el relleno de datos de lluvia diaria, y surge la siguiente pregunta de investigación: ¿Cuáles metodologías son las mejores para el relleno de datos de lluvia diaria? ¿Es posible mejorar la calidad de las series temporales de lluvia diaria si su información se completa con el apoyo de la información proveniente de sensores remotos? Este trabajo de investigación pretende dar respuesta a estas inquietudes.

## 1.2 Justificación

La precipitación es el principal componente del ciclo global del agua (Bai et al., 2018). Los modelos ambientales requieren series de tiempo completas de entradas meteorológicas, por lo que la reconstrucción de los datos faltantes es un tema clave en la funcionalidad de dichos modelos. La disponibilidad de registros meteorológicos fiables es fundamental para el desarrollo de estudios ambientales, sin embargo, a veces estos presentan errores, no tienen una extensión suficiente o están incompletos (J. W. Kim & Pachepsky, 2010).

El análisis de la calidad de la información es un paso fundamental en la incertidumbre asociada a la modelación hidrológica, incluye aspectos relacionados con la homogeneidad

y consistencia en los datos, que a su vez incluye el relleno de datos, el análisis de calidad, análisis de tendencias, estacionalidad, entre otros, y todos ellos dependen de la completitud de las series de datos hidroclimatológicos; por lo que una estrategia adecuada para el relleno de la información de lluvia diaria se convierte en una necesidad para la planificación y gestión de agua y del territorio (Hasana & Crokea, 2013).

Las series históricas de clima completas y de períodos largos son un requisito para la elaboración de estudios confiables. Los datos deben ser el resultado de un proceso adecuado, con registros periódicos que den continuidad a las observaciones, para que sea posible obtener resultados ajustados a la realidad del fenómeno que se estudia (Rivera, 2008).

El uso de información satelital es una oferta complementaria a las estaciones pluviométricas, debido a que proporcionan abundante información espaciotemporal de regiones extensas como los océanos, regiones montañosas de gran altitud y otras regiones remotas donde las redes de medición son difíciles de implementar (Shen et al., 2019). El estudio de la precipitación es muy importante para los estudios relacionados con el ambiente, la planificación del territorio y el ciclo hidrológico, además el uso de la información satelital puede ser un medio de información complementaria a las estaciones pluviométricas, ya que es una fuente con abundante información, lo que la vuelve una opción muy atractiva, teniendo en cuenta que se debe ser cuidadoso a la hora de utilizar esta información ya que presentar sesgos sistemáticos o problemas de variabilidad y resolución espaciotemporal.

Las series de precipitación sin datos faltantes son importantes para estudios meteorológicos e hidrológicos, dado esto surge la necesidad de la estimación y desarrollo de metodologías para completar los datos faltantes, las cuales obtengan resultados cercanos al valor real, lo que aporta al desarrollo de estudios hidrológicos, meteorológicos y de manejo y ordenación de los recursos hídricos para que sean consistentes (Cruz-Roa & Barrios, 2018; Suhaila et al., 2008). En general el insumo para las metodologías de relleno de series de estaciones pluviométricas, son las estaciones cercanas (Sattari & Rezazadeh Joudi, 2016). Las fuentes de información de teledetección como satélites o radares son complementarias a las estaciones en tierra, ya que tienen una mayor cobertura

espacial, lo que las hace interesantes para el relleno de las series de las estaciones en tierra (World Meteorological Organization, 2003), además estimaciones de la precipitación global con precisión uniforme no se observan con ningún otro método de medición diferente al de teledetección por satélite (Okamoto et al., 2005). Aunque el uso de los productos satelitales de precipitación es atractivo por las características de la teledetección (cobertura espacial y temporal y precisión uniforme en la estimación), estas fuentes de información deben ser validadas de alguna manera, ya que antes de hacer uso de cualquier base de datos es necesario evaluar el desempeño de esta (Ebert et al., 2007); para lo cual usualmente son utilizadas las estimaciones de los pluviómetros en tierra (Urrea et al., 2016). En vista de lo anterior, la información de la teledetección es complementaria a la de las estaciones en tierra y la información faltante en las series de esta última fuente es muy frecuente, por tanto, se evalúa la posibilidad de rellenar la información faltante de los pluviómetros aplicando las metodologías comúnmente utilizadas y que tienen como insumo la información disponible en las estaciones vecinas utilizando también como recurso adicional la información de algunos productos satelitales de precipitación que se encuentran disponibles de forma gratuita en la red de redes (internet).

### **1.3 Zona de estudio**

Se seleccionó el departamento de Caldas (Figura 1), ubicado en la zona cafetera colombiana, sobre las cordilleras central y occidental de los Andes colombianos. Limita con los departamentos de Tolima y Risaralda (sur), Risaralda (occidente), Antioquia (norte), Boyacá (noroeste) y Cundinamarca (oriente).



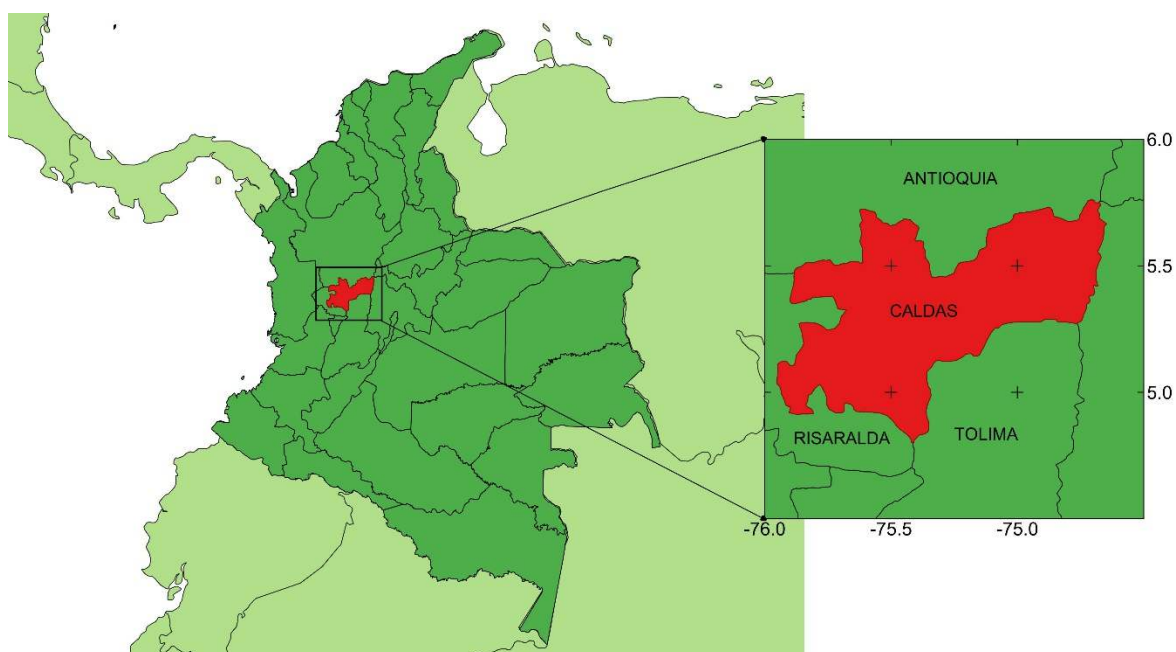


Figura 1. Zona de estudio, departamento de Caldas (zona cafetera colombiana). Fuente propia.

### 1.3.1 Modelo de elevación digital

En proporción, Caldas, es el departamento más montañoso del país, con altitudes entre los 5400 m.s.n.m (Nevado del Ruiz) y 170 m.s.n.m (municipio de La Dorada) (Gobierno de Caldas - Geografía de Caldas, n.d.), por lo tanto, según la clasificación climática de Caldas presenta todos los pisos térmicos (cálido, templado, frío, muy frío, extremadamente frío y nival). En la Figura 2 se presenta el modelo de elevación digital (MED) de tamaño 12.5 m (ASF, n.d.).

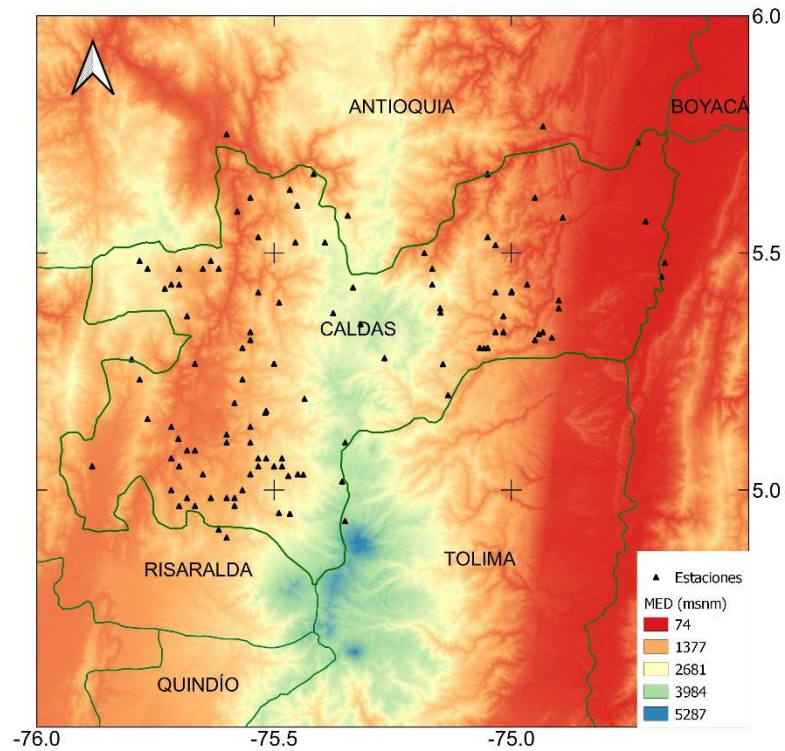


Figura 2. Modelo de elevación digital (MED). Fuente: (ASF, n.d.)

### 1.3.2 Clima

En la Figura 3 se muestra la lluvia media anual en el departamento de Caldas, en donde se destaca la alta variabilidad de la lluvia en el departamento de Caldas, lo cual indica que incluir variables adicionales con la lluvia, como la proveniente de sensores remotos puede ser una buena aproximación para el relleno de datos faltantes.

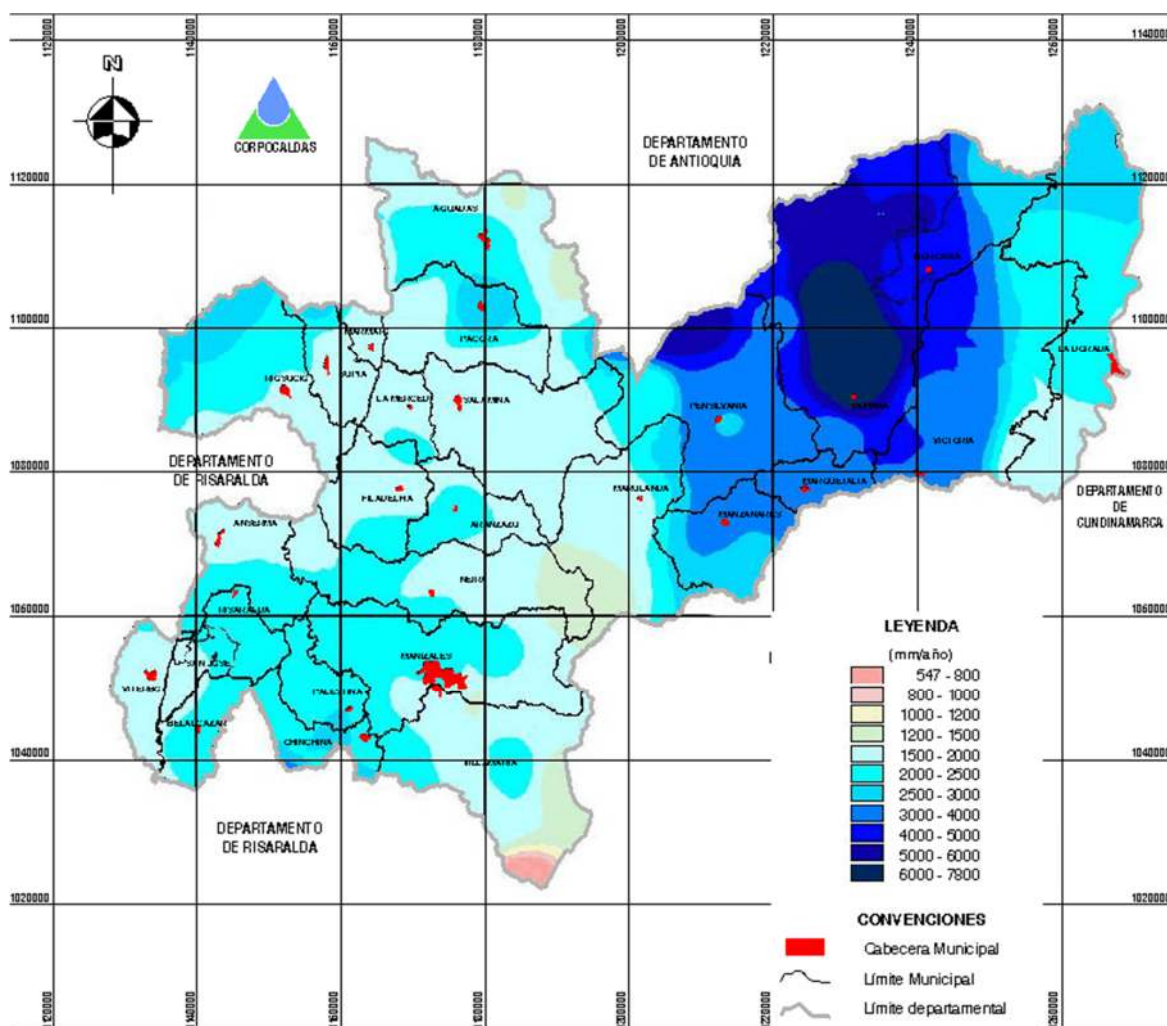


Figura 3. Precipitación media total anual. Adaptado de (CORPOCALDAS, 2001).

Como se mencionó en 1.3.1 debido a que el departamento de Caldas cuenta con un gran rango de variación altitudinal, se tienen climas cálidos, templados, fríos, muy fríos, extremadamente fríos y nivales, según la clasificación de Caldas (IDEAM, 2013).

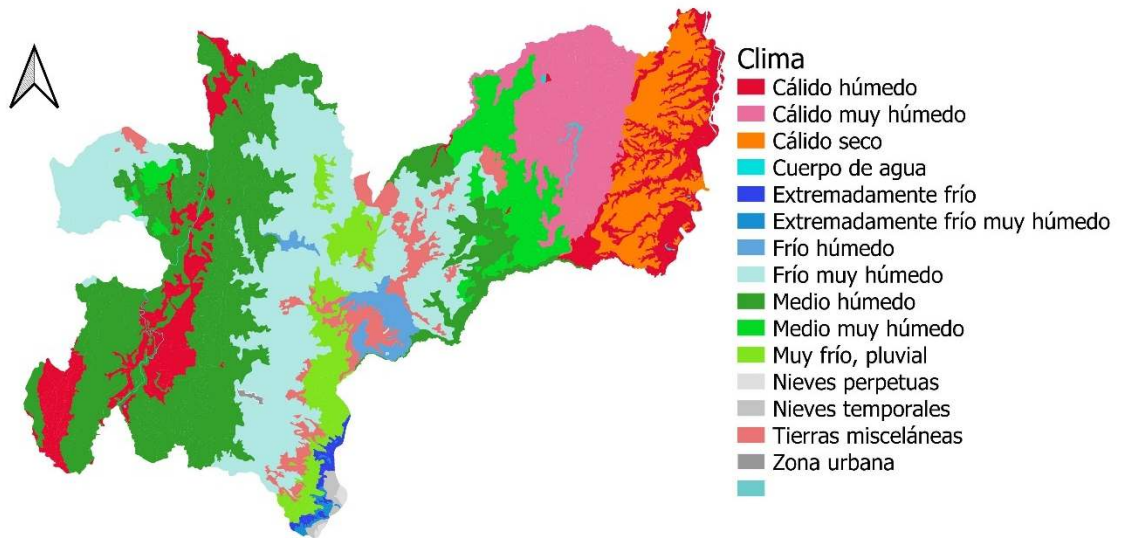


Figura 4. Mapa de zonificación climática. Fuente: (IGAC, 1998)

La región Andina, presenta una distribución bimodal de las precipitaciones mensuales, lo cual es característico de la Zona de Confluencia Intertropical (ZCIT), con dos periodos lluviosos (abril-mayo y octubre-noviembre) y dos periodos secos (diciembre-febrero y junio-agosto) (Poveda, 2004; Jaramillo & Chaves, 2000). En la Figura 5 se presenta un ejemplo del ciclo anual promedio de la precipitación en la región Andina colombiana.

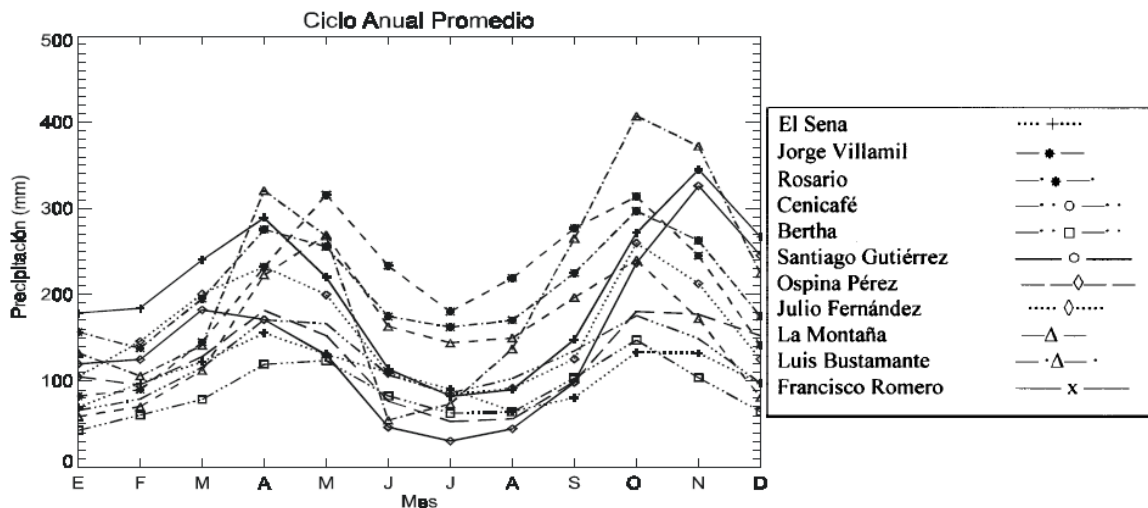


Figura 5. Ciclo anual promedio de la precipitación en diversas estaciones localizadas en los Andes de Colombia. Fuente: (Poveda et al., 2002)

### 1.3.3 Mapa de suelos

En la Figura 6 se presenta el mapa de suelos para el departamento de Caldas, tomado de (IGAC, 1998). Se destaca el gran porcentaje de cenizas volcánicas y rocas metamórficas en la zona central de Caldas, de roca metamórfica, sedimentaria, ceniza volcánica y suelo de origen aluvial y coluvial en la parte oriental y la presencia de rocas ígneas y metamórficas en el occidente.

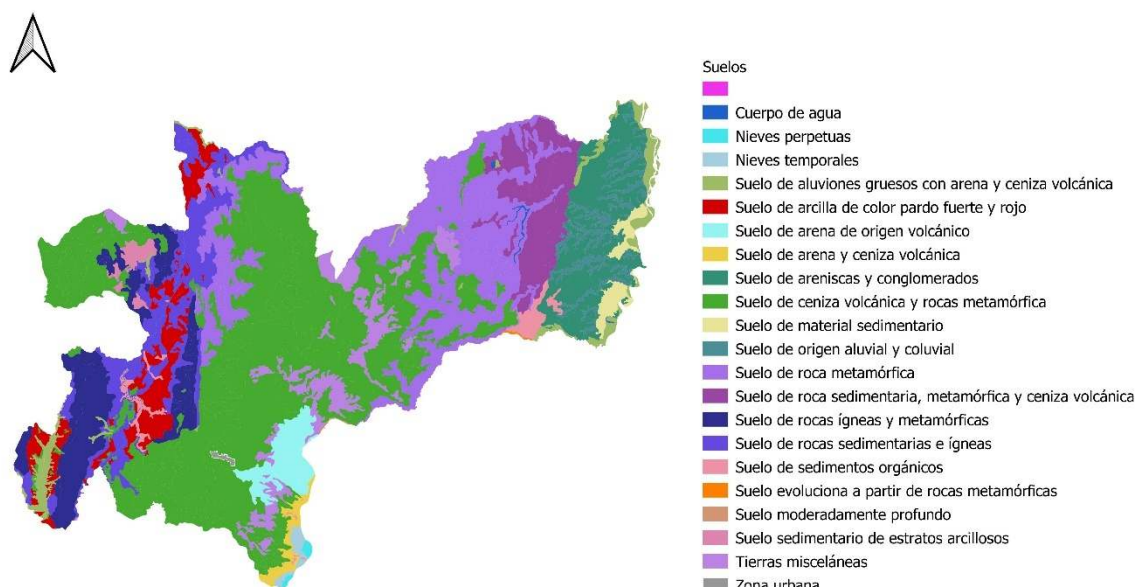


Figura 6. Mapa de suelos. Fuente: (IGAC, 1998)

### 1.3.4 Coberturas del suelo

En la Figura 7 se presenta el mapa de usos y coberturas del suelo para el departamento de Caldas, tomado de (IDEAM, 2014). Se observa la gran variedad de coberturas, donde se destaca la presencia de gran porcentaje de bosques y plantaciones forestales, mosaicos de pastos, cultivos y espacios naturales. Además, se puede observar en la parte alta las zonas glaciares y nivales correspondiente a el Parque Nacional Natural de Los Nevados.

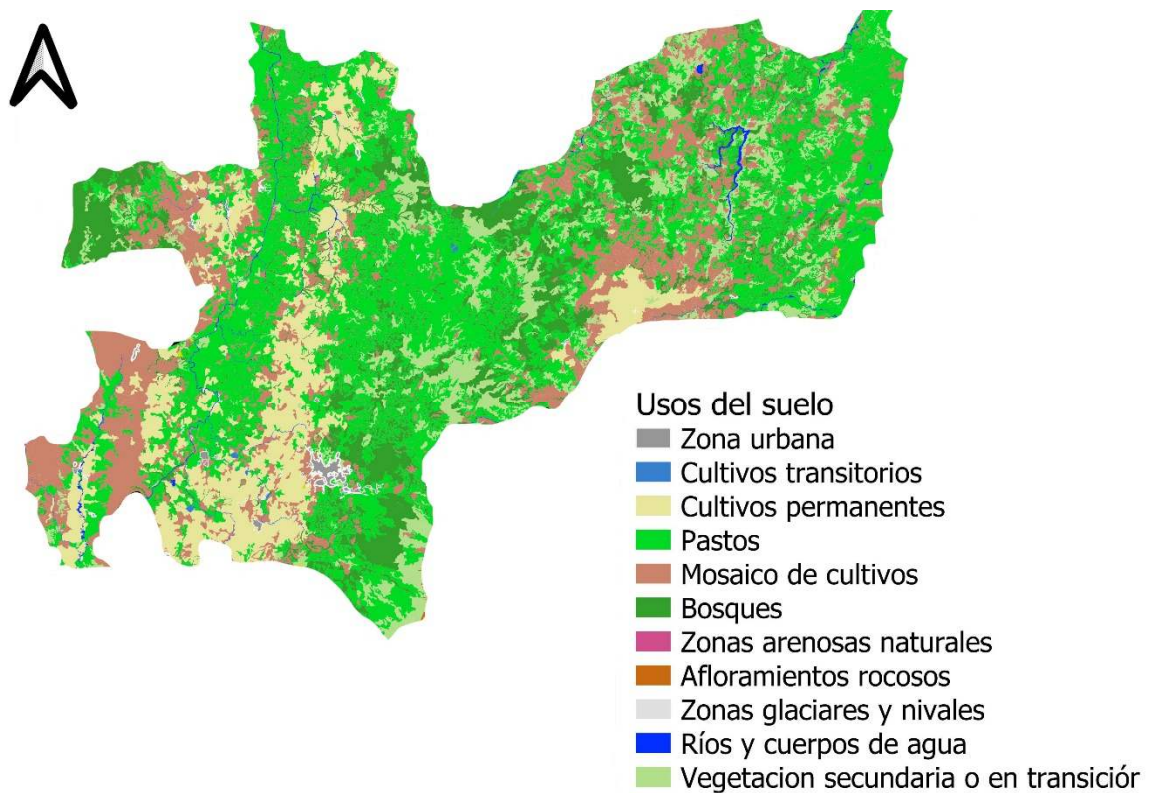


Figura 7. Usos del suelo departamento de Caldas. Fuente: (IDEAM, 2014)

Según lo anterior, en el departamento de Caldas, se presenta una notable variabilidad climática, evidenciada por la presencia de diversos tipos de suelo y una amplia gama de coberturas vegetales en toda la región. Sin embargo, esta diversidad también plantea un desafío importante, ya que la falta de patrones claros y consistentes entre estas covariables, dificulta considerablemente el uso de modelos que usen como insumo estas variables para estimar las precipitaciones en la zona de estudio.

## **2. Objetivos**

### **2.1 Objetivo General**

Desarrollar estrategias para el uso de información de satélite con el fin de rellenar series de lluvia diaria en zonas de montaña colombianas (departamento de Caldas, ubicado en la zona cafetera colombiana).

### **2.2 Objetivos específicos**

Los objetivos específicos para el logro del objetivo general son:

1. Descargar y analizar la información satelital de lluvia diaria para la zona de estudio de al menos dos fuentes.
2. Estudiar metodologías estadísticas, estocásticas y de inteligencia artificial para rellenar series de lluvia diaria.
3. Seleccionar y aplicar por lo menos una metodología por enfoque (estadístico, estocástico y de inteligencia artificial) en las estaciones en tierra del Instituto de Hidrología, Meteorología y estudios Ambientales de Colombia (IDEAM) ubicadas en la zona de estudio.
4. Evaluar el desempeño de las metodologías aplicadas para rellenar series de precipitación diaria con el uso de información de satélite y sin esta en las series de las estaciones en tierra de la zona de estudio.

## **3. Estado del arte**

Durante décadas se han desarrollado técnicas para estimar las precipitaciones mediante sensores remotos, como fuentes de información complementarias o alternas de las

estaciones en tierra. En 1987 (Arkin & Meisner) clasificaron estas técnicas como directas (mediciones de precipitación realizadas por medio de la radiación de microondas sobre hidrometeoros) e indirectas (medición de variables que son productos de la precipitación, como las nubes, a partir de las cuales se estiman las precipitaciones). El enfoque indirecto más usado en la época consistía en usar la información de imágenes visibles y/o del espectro infrarrojo (VIS/IR) de la distribución temporal y espacial de la temperatura superior de las nubes y su brillo, mediante las cuales se calculaban las distribuciones e intensidades de las precipitaciones (Arkin & Meisner, 1987).

Las técnicas VIS/IR se comenzaron a usar desde mediados de la década de 1960 y las microondas desde mediados de la década de 1970 (Arkin, 1990). Las microondas debido al desarrollo tecnológico de la época solo pudieron ser aplicadas hasta 1987 donde el primer instrumento con esta capacidad entró en operación, el SSM/I (Special Sensor Microwave/Imager) (Arkin, 1990). También existen enfoques en los cuales se combinan varios tipos de información con el fin de aumentar la exactitud en la estimación de las precipitaciones. El primer producto que combinó la información de microondas de satélites de órbita baja (tomadas del SSM/I), datos infrarrojos de satélites de órbita geosíncrona (basados en Defense Meteorological Satellite Program, DMSP) y de estaciones (del Deutscher Wetterdienst) fue el Global Precipitation Climatology Project (GPCP), el cual fue creado en 1986 por el World Climate Research Program (WCRP) (Huffman et al., 1997). Actualmente los sensores satelitales se basan en la medición de variables indirectas con las cuales se estima la precipitación (Toté et al., 2015). Estas variables se clasifican en (1) información espacial de estaciones, (2) reanálisis de modelos físicos y dinámicos y (3) microondas pasivas (PMW) y datos del espectro infrarrojo (IR); la mayoría de satelitales se basan en la combinación de estas para obtener la mayor exactitud posible (Satgé et al., 2020). El desarrollo tecnológico abre una gran cantidad de posibilidades para cuantificar las precipitaciones como lo son los productos satelitales de precipitación, que resultan ser fuentes adicionales de información a las estaciones en tierra. Sin embargo, las incertidumbres en la información satelital, por ser estimaciones calculadas a partir de las variables indirectas de medición, obligan a realizar la validación de este tipo de información para su uso.



Las imprecisiones de las precipitaciones satelitales varían con la topografía, región y clima. A lo largo de los años se han realizado estudios en los cuales se comprueba esta dependencia. El ajuste de los satélites en bajas latitudes es mejor con respecto a las altas latitudes (Lei et al., 2022), las incertidumbres son altas en las zonas con topografías complejas (Amjad et al., 2020), bajas en los océanos y se incrementan en los trópicos (Tian & Peters-Lidard, 2010). Una adecuada precisión de la información es dependiente de la aplicación que se le quiera dar (Tapiador et al., 2012). Según lo anterior, el estudio detallado de la precipitación satelital en una zona específica de interés es importante para evaluar el desempeño real de estos. Además, debido a estas incertidumbres generadas por los satélites, así como la escala espacial y temporal de su información, hacen que las estaciones en tierra sigan siendo la fuente de información más confiable para analizar una zona específica. En consecuencia, estudios como este buscan analizar si la información satelital puede ser un buen complemento para estimar los datos faltantes de las series de precipitación diaria en las estaciones de tierra.

### **3.1 Metodologías de relleno de series**

Los métodos para estimar datos faltantes se pueden clasificar en métodos determinísticos, estocásticos, de inteligencia artificial y geoestadísticos (Armanuos et al., 2020). En los métodos estadísticos los datos faltantes se calculan a partir de una suma ponderada de datos de entrada (Xia et al., 1999). Los métodos estocásticos proporcionan estimaciones probabilísticas de los resultados, los cuales son un poco más demandantes computacionalmente con respecto a los métodos determinísticos (Campozano et al., 2014). Los métodos de inteligencia artificial, son altamente demandantes computacionalmente, ya que requieren cálculos matemáticos de alta complejidad, por lo que es más difícil su aplicación (Campozano et al., 2014).

Una de las principales aplicaciones de la estadística en la climatología es la estimación de datos faltantes en las series temporales, en donde es fundamental tener en cuenta que los

valores calculados deben ser cercanos a los valores reales y coherentes con los fundamentos físicos (OMM, 2011; OMM, 2018).

El método de la relación normal (por sus siglas en inglés, NRM) y el método de distancia inversa ponderada (por sus siglas en inglés, IDW) son los dos métodos de ponderación tradicionales que son utilizados más comúnmente en la estimación de datos climáticos faltantes (Suhaila et al., 2008). La Sociedad Americana de Ingenieros Civiles (American Society of Civil Engineers., 1996) recomiendan el uso de los métodos NRM e IDW para la estimación de datos faltantes. El NRM fue usado en múltiples estudios tradicionales (J. L. H. Paulhus & Kohler, 1952; Young, 1992) y sigue siendo usado en estudios actuales como en el propuesto por Morales et al., (2019). Asimismo ocurre con el IDW en donde hay múltiples estudios tradicionales en los que se utiliza (R. S. V. Teegavarapu & Chandramouli, 2005; Chang et al., 2006) y en estudios más recientes como en (Moeletsi et al., 2016; Morales et al., 2019) , es decir, que la aplicación de estas dos metodologías ha trascendido sobre el paso de los años, aun con la aparición de metodologías más complejas, por lo que se hace interesante aplicarlas en este estudio.

Métodos estocásticos como las cadenas de Markov dan buenos resultados para la generación de series sintéticas cuando se estiman grandes periodos de tiempo (Mehan et al., 2017). Las cadenas de Markov de primer orden con dos estados se han usado en múltiples estudios con buenos resultados (Guevara Ochoa et al., 2017; Mehan et al., 2017; Telesca et al., 2017; Wilks, 1999). Esta metodología usa la información de la serie histórica de la misma estación para calibrar sus parámetros y estimar los datos faltantes. Según lo anterior, se hace interesante el uso de esta metodología para el caso de estudio debido a que presenta un buen desempeño para el relleno de largos periodos de tiempo, situación que se adapta a las series de la zona de estudio con algunas estaciones que presentan grandes periodos de datos faltantes. Además, dentro de las metodologías propuestas, es la única que no requiere información de las estaciones vecinas como insumo para su implementación.

Las metodologías para completar los datos faltantes, se han desarrollado y mejorado con los años (J. Kim & Ryu, 2016). Las metodologías determinísticas y estocásticas son las más usadas típicamente para completar datos faltantes de precipitación, pero estos

enfoques no siempre son lo más exactos por no caracterizar la variabilidad temporal y espacial de las zonas de estudio (R. Teegavarapu, 2009). Metodologías nuevas que implementan otros enfoques como por ejemplo el uso de la inteligencia artificial (IA) se han desarrollado, y aunque su implementación lleva muchos años, su aplicación actualmente es más accesible debido a los grandes avances computacionales. Dado esto, recientemente el uso de las redes neuronales se ha empleado para el relleno de series de precipitación, predicción de precipitación, caudal y nivel de agua (Guevara Ochoa et al., 2017). Las redes usadas comúnmente son función de base radial y perceptrón multicapa (Guevara Ochoa et al., 2017). Coulibaly & Evora (2007) utilizaron seis tipos de redes neuronales artificiales con el fin de evaluar su desempeño para rellenar series de lluvia diaria en 15 estaciones de una cuenca de Canadá; encontraron que el tipo de red más efectiva para rellenar las series de lluvia diaria es la red de perceptrones multicapa (MLP).

En este estudio se quiere ver la aplicabilidad de la información de los productos satelitales para completar datos faltantes de las series de precipitación de las estaciones en tierra, por lo que se aplicaron diferentes tipos de metodologías (estadísticas, estocásticas y de inteligencia artificial), con el fin de comparar el desempeño utilizando como insumo algunos productos satelitales con diferentes enfoques. La coincidencia entre las observaciones estimadas por las metodologías y las utilizadas para validar no garantiza el rendimiento de un modelo en escenarios de climas futuros (Tapiador et al., 2012).

A continuación, se presentan diferentes metodologías estadísticas, estocásticas y de inteligencia artificial que son comúnmente usadas en la estimación de datos faltantes de series de precipitación, y que serán aplicadas en las series de la zona de estudio para evaluar su desempeño. Cabe resaltar que la finalidad del estudio es conocer la aplicabilidad de la información de los productos satelitales para la estimación de datos faltantes de las series de lluvia diaria en la zona de estudio, más no encontrar cual es la metodología o el enfoque (estadístico, estocástico o de inteligencia artificial) que presenta el mejor desempeño a la hora de rellenar datos.

### 3.1.1 Metodologías estadísticas y geoestadísticas

Se trata de métodos clásicos, algunos de ellos muy simples y que son ampliamente utilizados por su sencillez.

- **Curvas de doble masa**

Las curvas de doble masa de precipitación son utilizadas para medir la consistencia de la información entre estaciones cercanas y muestran la relación entre las lluvias acumuladas de dos estaciones, de esta forma, si se aprecia una relación lineal entre ellas, sin cambios bruscos en la pendiente, se dice que ambas estaciones están sometidas a lluvias de un mismo origen y son consistentes. Además, es posible utilizar esta relación para rellenar el valor faltante en una de las estaciones con base en la otra, simplemente ajustando la pendiente entre las estaciones para totalizar la lluvia faltante en función de la estación con datos. Sin embargo, si no se observa relación entre ellas es un indicador de que la lluvia en ambas estaciones es de origen diferente y no se considera consistencia entre las estaciones y por lo tanto no se puede utilizar para el relleno. Para el relleno de datos se sugiere el uso de la estación cercana que tenga una curva de doble masa sin quiebres en la pendiente.

- **Distancia inversa ponderada (IDW)**

El método IDW es probablemente el método más usado para estimar datos faltantes en el campo de la hidrología (Di Piazza et al., 2011; Teegavarapu & Chandramouli, 2005; Sattari et al., 2016; R. Teegavarapu, 2009), se clasifica como un método determinístico y fue desarrollado por la U.S. National Weather Service en 1972 (F.-W. Chen & Liu, 2012).

Este método se basa en el supuesto de que los valores de la lluvia en la estación objetivo podrían verse más influenciados por las estaciones cercanas y menos por las más lejanas (Suhaila et al., 2008). Un problema significativo de usar este método es la selección de estaciones vecinas arbitrarias para la estimación de datos faltantes de la estación objetivo (Di Piazza et al., 2011). Esta metodología se define por la siguiente ecuación (Campozano et al., 2014):

$$V_{est} = \frac{\sum_{i=1}^n V_i d_i^{-k}}{\sum_{i=1}^n d_i^{-k}}$$

Donde  $V_{est}$  es el valor estimado;  $n$  el número de estaciones;  $V_i$  el valor observado en la estación  $i$ , la distancia entre la estación objetivo y la  $i$ -ésima estación, y  $k$  la potencia de la distancia. A mayor distancia la influencia de la estación es menor, por lo que es usual definir una distancia o rango óptimo para el método, del mismo modo es usual utilizar para la estimación sólo entre tres o cuatro estaciones cercanas, para evitar el fenómeno de apantallamiento entre estaciones más lejanas.

- **Método de la relación normal modificado, basado en la correlación (NRM)**

En 1952 (Paulhus & Kohler) propusieron el NRM ya que los métodos utilizados hasta el momento, no siempre tenían buen desempeño en las zonas montañosas, por lo que proponen este método el cual utiliza estaciones circundantes las cuales tienen relación con respecto a la estación objetivo; este método fue posteriormente modificado por Young (1992) en el cual se determinan los coeficientes de correlación entre cada estación y la estación objetivo.

Las tres estaciones con el coeficiente de correlación más alto positivo se seleccionan. La estimación de datos faltantes se realiza a partir de las siguientes ecuación (Kanda et al., 2018):

$$V_0 = \frac{\sum_{i=1}^n W_i V_i}{\sum_{i=1}^n W_i} \quad W_i = r_i^2 \left( \frac{n_i - 2}{1 - r_i^2} \right)$$

Donde  $W_i$  es el peso de la  $i$ -ésima estación;  $V_i$  es el valor de la variable de interés en la  $i$ -ésima estación;  $r_i$  es el coeficiente de correlación entre la estación objetivo y la  $i$ -ésima estación circundante y  $n_i$  es el número de datos utilizados para calcular el índice de correlación.

- **Mejor estimador único (SBE)**

Este es uno de los métodos más simples para estimar datos faltantes, en donde es necesario tener disponibilidad de información histórica (R. S. V. Teegavarapu, 2012). Es análogo a usar la información de la estación vecina más cercana para estimar los datos de la estación objetivo. La estación objetivo se estima a partir de tomar el valor real observado de la estación vecina con la correlación más alta positiva con respecto a la estación objetivo (Eischeid et al., 2000).

- **Mapeo de cuantiles (QM)**

Para una estación objetivo dada, los valores de lluvia que faltan se completan utilizando el método de mapeo de cuantiles (Panofsky & Brier, 1958), de la siguiente forma: (i) las estaciones predictoras se clasifican de mayor a menor en función de su correlación con una estación objetivo, (ii) las funciones de distribución acumulativa (CDF) se calculan para algún período de datos superpuestos definido por el usuario en las estaciones objetivo y predictoras, (iii) la probabilidad acumulativa de una observación en una estación predictora se calcula a partir de la CDF de la estación predictora, y la probabilidad acumulativa se usa en la CDF de la estación de destino para determinar el valor de llenado en la estación de destino. Longman et al. (2020) lo comparan con otros métodos de relleno de datos a varias escalas temporales usando diferentes indicadores de desempeño con buenos resultados.

- **Kriging**

Esta metodología es muy utilizada para el relleno de datos satelitales y para rellenar series de lluvia, y es el mejor estimador lineal insesgado, en geoestadística se refiere a una familia de métodos de regresión de mínimos cuadrados generalizados (Webster & Oliver, 2007). Esta metodología ayuda a estimar los valores de variables desconocidas en ubicaciones no observadas en función de los valores conocidos observados en ubicaciones circundantes.

Kajornrit et al. (2011) comparan varios métodos de kriging utilizados para estimar los datos faltantes de precipitaciones mensuales y muestran que las características de los conjuntos de datos tienen un efecto significativo en el rendimiento de la estimación, por lo que recomiendan utilizar el valor de curtosis del histograma de las observaciones y el efecto

pepita de los modelos de semivariograma ajustados como guía para seleccionar entre los métodos kriging ordinario y kriging universal. Además, mencionan que, si existe una relación entre la precipitación y la altitud, el método ordinario de co-kriging puede ser utilizado con la altitud como una característica complementaria para mejorar el rendimiento de la estimación.

- **Otros métodos**

Toro et al. (2017) utilizaron para rellenar datos faltantes de lluvia diaria los métodos del Servicio Nacional de Clima de los Estados Unidos (NWS) y la regresión lineal múltiple (RLM), donde la variable meteorológica dependiente es el valor del dato faltante en la estación a completar y la independiente es el valor de la variable de las estaciones vecinas el mismo día que disponen de datos. También es común el uso las regresiones lineales basadas en transformaciones logarítmicas que convierten los datos en variables estándar normalizadas (Mora et al., 2014). Finalmente, están los métodos de regresión multivariada que incluyen variables adicionales, cuando se encuentran disponibles, como pueden ser la temperatura, la radicación solar, la humedad relativa, entre otras variables.

### **3.1.2 Metodologías estocásticas**

Son métodos que de alguna manera incluyen la medida de la incertidumbre en la estimación de los valores faltantes, por esta razón se consideran métodos más robustos; es importante mencionar que el método de kriging no es un método estocástico, pero está en capacidad de suministrar el error.

- **Cadenas de Markov de primer orden con dos estados**

Las cadenas de Markov de primer orden con dos estados son usadas para generar estocásticamente días lluviosos y días secos (Telesca et al., 2017). Telesca et al. (2017) describieron este modelo de la siguiente manera: la generación de la precipitación está basada en dos suposiciones: La primera es que la lluvia de un día  $i$  está relacionada con la condición de lluvia de un día  $i-1$  y la segunda es que la cantidad de lluvia de los días lluviosos se describen mediante una función de distribución adecuada.

La primera suposición describe un modelo llamado cadena de Markov, definiendo  $p(w/w)$  como la probabilidad de tener un día húmedo dado que el día  $i-1$  fue un día húmedo y  $p(w/d)$  como la probabilidad de tener un día húmedo dado que el día  $i-1$  fue un día seco. A continuación, se muestran las probabilidades de transición complementarias:

$$p\left(\frac{d}{w}\right) = 1 - p\left(\frac{w}{w}\right)$$

$$p\left(\frac{d}{d}\right) = 1 - p\left(\frac{w}{d}\right)$$

Donde  $p(d/w)$  es la probabilidad de un día seco dado que hubo un día húmedo en el día  $i-1$  y  $p(d/d)$  es la probabilidad de un día seco, dado que el día  $i-1$  fue un día seco.

Los valores diarios de estas probabilidades se interpolan mediante funciones *spline*. Este generador de tiempo determina si el día es seco o húmedo restando a las probabilidades  $p(w/w)$  o  $p(w/d)$  un número aleatorio entre 0 y 1; si el resultado es mayor a 0, se asume que no hay lluvia ese día. Si el resultado es menor a 0 se determina que ese día hay lluvia y se determina la cantidad de lluvia utilizando una función de distribución de probabilidad (FDP) para las cantidades de lluvia en los días húmedos. Para calcular la cantidad de lluvia utilizando las FDP se estima un número aleatorio entre 0 y 1 que representa la probabilidad (este número es diferente al utilizado para estimar si es un día seco o lluvioso).

Se hicieron ajustes a varias FDP con el fin de identificar cual se ajusta mejor para cada estación y estimar de mejor manera la magnitud de las precipitaciones; para hacer los ajustes se utilizaron las series históricas con lluvias mayores o iguales a 0.2 mm. En la Tabla 1 se presentan las FDP utilizadas.



Tabla 1. Funciones de distribución de probabilidad.

FDP	Ecuación	Parámetros
Gumbel	$F(x) = \exp(-\lambda \exp(-\theta x))$ $f(x) = \lambda \theta \exp(-\theta x) \exp(-\lambda \exp(-\theta x))$	Donde $\Theta$ es el parámetro de escala y $\lambda$ es el parámetro de forma.
GEV	$F(x) = \exp \left[ - \left( 1 - \frac{\beta}{\alpha} (x - x_0) \right)^{\frac{1}{\beta}} \right]$ $f(x) = F(x) \left[ \frac{1}{\alpha} \left( 1 - \frac{\beta}{\alpha} (x - x_0) \right)^{\frac{1}{\beta} - 1} \right]$	Donde $\alpha$ es el parámetro de escala, $\beta$ es el parámetro de forma y $x_0$ es el parámetro de localización. Si $\beta > 0$ entonces $x < \left( x_0 - \frac{\alpha}{\beta} \right)$ Si $\beta < 0$ entonces $x < \left( x_0 + \frac{\alpha}{\beta} \right)$
Log Normal	$F(x) = \Phi \left( \frac{\ln(x) - \mu_y}{\sigma_y} \right)$ $f(x) = \frac{1}{x\sqrt{2\pi\sigma_y^2}} \exp \left[ -\frac{1}{2} \left( \frac{\ln(x) - \mu_y}{\sigma_y} \right)^2 \right]$	Donde $\Phi$ es la función de distribución acumulada normal estándar.
Exponencial	$F(x) = \lambda \exp(-\lambda x)$ $f(x) = \frac{\lambda}{2} \exp(-\lambda x) \exp \left( \left( \frac{\lambda}{2} \right) (\mu + \sigma^2 - 2x) \right) \operatorname{erfc} \left( \frac{(\mu + \sigma^2 - x)}{\sigma\sqrt{2}} \right)$	Donde $\lambda$ es el parámetro de escala de la distribución exponencial, $\mu$ es el parámetro de ubicación de la distribución normal, $\sigma$ es el parámetro de escala de la distribución normal, $\exp()$ es la función exponencial y $\operatorname{erfc}()$ es la función de complemento del error.

## • ARIMA

Twumasi et al. (2021) describieron ARIMA mediante dos módulos; Autoregresión (AR) y Promedio móvil (MA). En algunos casos, se agrega un módulo adicional de integración cuando los datos no son estacionarios y requieren diferenciación para lograr estacionariedad. El módulo de Autoregresión (AR) asume que el valor más reciente es igual al valor inmediatamente anterior más un término de ruido blanco aleatorio. Por otro lado, el módulo de Promedio móvil (MA) asume que el valor actual es igual al valor anterior más un término de ruido blanco adicional. La combinación de los módulos AR y MA da como resultado el modelo denominado Autoregressive Moving Average (ARMA).

El modelo ARIMA se representa en la forma ARIMA (p, d, q) (P, D, Q), donde p y P indican el orden de rezago para los componentes AR y AR estacionales respectivamente. El

parámetro  $d$  y  $D$  representan el grado de diferenciación (es decir, el número de veces que los datos se diferenciaron) para los componentes no estacionales y estacionales. Finalmente, los parámetros  $q$  y  $Q$  indican el orden de rezago para los componentes MA y MA estacionales respectivamente.

### 3.1.3 Metodologías de inteligencia artificial

Estos métodos hacen uso de la misma información para entrenar o proponer estrategias de aprendizaje que permitan obtener relaciones regionales que permitan completar los datos faltantes en series temporales hidroclimatológicas.

- **Redes de perceptrones multicapa**

Este tipo de red tienen una capacidad cartográfica de entrada-salida que parece adecuada para la modelación de la precipitación dado que las series de precipitación diaria no suelen estar correlacionadas (Coulibaly & Evora, 2007). Coulibaly & Evora (2007) describieron este tipo de red como se menciona a continuación: las redes de perceptrones multicapa constan de tres componentes: una capa de entrada, una capa oculta y una capa de salida.

En las redes de perceptrones multicapa convencionales, la información de entrada se transmite como se muestra en la Figura 8. La salida de la red suponiendo una neurona de salida lineal  $j$ , una capa oculta con  $h$  nodos ocultos sigmoides y una entrada variable  $x(y)$  viene dada por:

$$Y_k = F \left( \sum_{j=1}^h W_j G(S_j) + b_k \right)$$

Donde  $F(\cdot)$  es la función de activación lineal de la salida neurona  $k$  y  $b_k$  es su sesgo (o umbral),  $w_j$  representa los pesos de conexión entre las unidades ocultas y de salida,  $G(\cdot)$  es el sigmoide tangente hiperbólico utilizado como función de activación para los nodos ocultos, y se puede expresar como sigue:

$$G(S_i) = \frac{e^{S_i} - e^{-S_i}}{e^{S_i} + e^{-S_i}}$$

Donde  $x_i$  son las entradas a la red,  $w_i$  son los pesos de la conexión entre los nodos de las capas de entrada y ocultas.

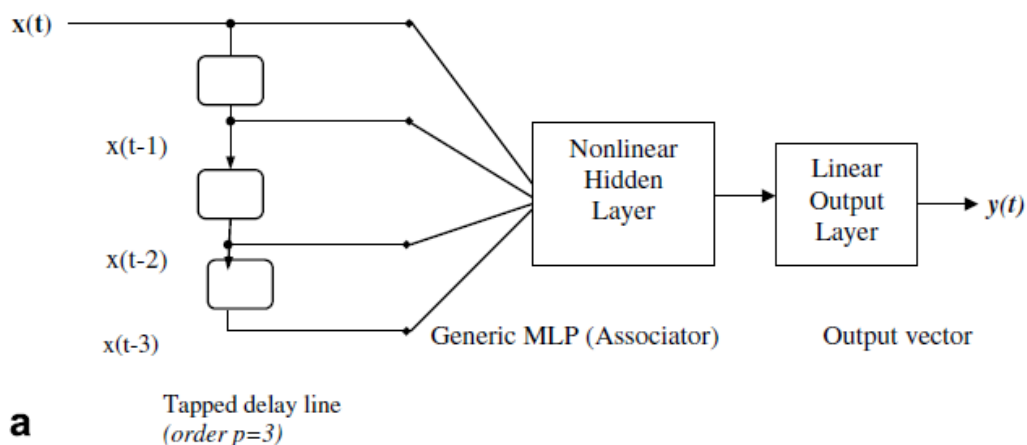


Figura 8. Redes de perceptrones multicapa. Tomada de (Coulibaly & Evora, 2007).

Es importante tener en cuenta que todos los métodos de relleno de datos requieren un análisis previo que consiste en separar las estaciones en regiones climáticamente homogéneas, para lo cual también es posible utilizar las técnicas estadísticas o de IA.

### • Otros métodos

Los métodos de inteligencia artificial y minería de datos también son muy utilizados para identificar zonas con patrones de lluvia similar, entre ellos se destacan el análisis de conglomerados (Clustering) que sirve para clasificar las estaciones o los puntos de la cuadrícula que muestran diferentes regímenes de lluvia y el análisis de conglomerados jerárquicos (por su siglas en inglés, HCA) cuyo análisis de los datos de lluvia revela el patrón de comportamiento del régimen de lluvias en una región específica mediante la identificación de conglomerados homogéneos (Basri et al., 2021).

Friedman et al. (2000) introdujeron el método del aumento de gradiente (Gradient Boosting, GB) cuyo enfoque trata sobre la construcción de un conjunto de árboles de regresión con una salida que depende de uno o más predictores. Por ejemplo, un modelo de árbol único

hace una predicción de valor único, pero tiene un rendimiento bajo. La ventaja del enfoque es que utiliza no solo uno, sino muchos de estos árboles. Los resultados presentados indican que hay varias ventajas en comparación con otros métodos como la regresión lineal múltiple o las redes neuronales.

Kalteh & Hjorth (2009) emplean técnicas de relleno de valores faltantes por medio de la IA utilizando el mapa autoorganizado (SOM), el perceptrón multicapa (MLP), el vecino más cercano multivariado (MNN), el algoritmo de maximización de expectativas regularizadas (REGEM) y la imputación múltiple (MI) con el fin de construir una base de datos completa para la predicción de la escorrentía. Concluyen que SOM y MNN tienden a dar resultados similares y robustos.

### **3.1.4 Metodologías híbridas**

Se trata de métodos que combinan las estrategias anteriores para beneficiarse de las ventajas de cada una de ellas, dado que son más complejas pueden requerir un mayor esfuerzo computacional. También se pueden considerar métodos híbridos aquellos realizan una separación por regiones homogéneas antes de realizar el relleno de datos con información de una misma región.

En la comparación de 10 métodos presentados por Miró et al. (2017), los autores realizan el pre-proceso para identificar regiones comunes utilizando el mapeo de cuantiles y destacan que el mejor desempeño es para los métodos híbridos y no lineales, resaltando que el método PCA no lineal (NLPCA) supera considerablemente al método de mapas autoorganizados (SOM) dentro de los enfoques no lineales. En los métodos lineales, el método Regularized Expectation Maximization (RegEM) fue el mejor, pero lejos de NLPCA. La aplicación del filtrado EOF como post procesamiento de NLPCA (enfoque híbrido) arrojó los mejores resultados.

## **3.2 Productos satelitales de precipitación**

Urrea et al. (2016) realizaron la validación a escalas diarias, mensuales y anuales de CHIRPS v2.0 en Colombia, para esto utilizaron 904 series de las estaciones en tierra

---

disponibles de precipitación diaria entre 1981 y 2014. Concluyeron que CHIRPS es un producto valioso para el estudio y la gestión de agua en Colombia, ya que conserva características importantes de la precipitación como la precipitación media, la precipitación acumulada total y la estacionalidad para las escalas diaria, mensual y anual. Adicionalmente resaltan características importantes de este producto satelital, las cuales se adaptan a las necesidades de los estudios en Colombia, 1) la resolución espacial es la más alta entre los productos satelitales disponibles para Colombia, 2) la resolución temporal es alta y suficiente para muchas aplicaciones, 3) representa bien los valores medios multianuales, la estacionalidad y la precipitación acumulada total. Estas características y resultados hacen interesante la aplicación de CHIRPS al caso de estudio, donde se quiere evaluar el desempeño de utilizar la información espacial (productos satelitales) para rellenar los datos de las series de la información puntual (estaciones en tierra); dado que CHIRPS es el producto satelital con mejor resolución en Colombia se hace interesante para evaluarlo en esta aplicación. En el capítulo 3.2.1 se presenta un resumen de las características de CHIRPS 2.0.

H E Beck et al. (2017) compararon 22 productos globales de precipitación para el periodo 2000-2016, evaluándolos con respecto a series diarias de precipitación de 76,086 estaciones en tierra; de los 22 productos analizados, MSWEP V2 y V1 obtuvieron las mejores correlaciones con respecto a las estaciones, es decir, que MSWEP V2 y V1 son los mejores productos satelitales de precipitación si se hace una comparación con estaciones a nivel global. Lo anterior es interesante para este estudio ya que la evaluación fue hecha con base en estaciones en tierra, lo cual también se realizó en este estudio (relleno de series de estaciones).

Beck et al. (2017) mencionan las características que en conjunto hacen único al MSWEP V2:

- Cobertura global de toda la tierra la cual incluye los océanos.
- Alta resolución espacial ( $0.1^\circ$ ) y temporal (3 horas).
- Combinación de diferentes fuentes de datos (reanálisis, satélites y pluviómetros).
- Corrección por sesgos distributivos, para eliminar lluvias falsas y restaurar picos atenuados.

- Corrección de los sesgos sistemáticos de precipitación terrestre debido a la medición insuficiente de las estaciones, utilizando las mediciones de caudal en 13.762 cuencas en todo el mundo.
- Correcciones utilizando observaciones diarias de 76.747 estaciones en todo el mundo; dentro de las cuales hay estaciones del IDEAM.
- Un sistema de corrección de las estaciones que tiene en cuenta los tiempos de medición de las estaciones.

Según lo anterior el uso del MSWEP V2 es interesante para la estimación de datos faltantes de las series de precipitación diaria de las estaciones de la zona de estudio, ya que su alta resolución temporal y espacial lo hacen resaltar con respecto a otros productos satelitales (este producto hace estimaciones más locales de la precipitación). Adicionalmente para su creación se usaron estaciones en tierra dentro de las cuales hay estaciones del IDEAM (base de datos de las estaciones usadas para este estudio).

A continuación, en los numerales 3.2.1 y 3.2.2 se presenta la descripción de los dos productos satelitales usados en este estudio. Estos satélites se escogieron por los argumentos presentados en esta sección, que hacen a estos productos atractivos para el caso de estudio.

### **3.2.1 The Climate Hazards group Infrared Precipitation with Stations (CHIRPS 2.0)**

Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) es un producto global de precipitación con datos desde 1981 hasta el presente, en el que se incorporan diferentes fuentes de información; datos del Climate Hazards Center, CHPclim, imágenes satelitales de 0.05° de resolución y datos de estaciones en campo, para crear series temporales de precipitación distribuidas en una rejilla regular cuadrículada (CHIRPS: Rainfall Estimates from Rain Gauge and Satellite Observations | Climate Hazards Center - UC Santa Barbara, n.d.). Los tres principales componentes del desarrollo de CHIRPS son

the Climate Hazards group Precipitation climatology (CHPclim), the satellite-only Climate Hazards group Infrared precipitation (CHIRP) y finalmente la combinación con las observaciones de estaciones en campo que conforman CHIRPS (Chris Funk et al., 2015).

Climate Hazards Group's Precipitation Climatology versión 1 (CHPclim v.1.0) es un producto de precipitación mensual global con una resolución de  $0.05^\circ$ , el cual se desarrolló a partir de la combinación de información satelital, indicadores fisiográficos de rejilla regular cuadrículada y normales climatológicas en campo, las cuales se combinan a partir de un modelado geoespacial basado regresiones de ventana móvil e interpolación de ponderación inversa (C. Funk et al., 2015). En la validación de este producto se obtuvieron resultados buenos comparándolo con productos climatológicos globales similares, especialmente en zonas topográficas complejas y de baja densidad de estaciones en tierra; adicionalmente se obtuvieron conjuntos de datos climatológicos de alta calidad para Colombia ( $R^2=0.84$ , con 194 estaciones) (C. Funk et al., 2015).

Los campos de CHIRP son variaciones de la media del CHPclim (Chris Funk et al., 2015). El enfoque utiliza la calibración local de las estimaciones de la precipitación satelital con estadísticas del infrarrojo térmico (TIR) y CCD (Chris Funk et al., 2015). El CHIRP se basa en dos archivos TIR geosincrónicos globales, el 1981–2008 Globally Gridded Satellite (GriSat) producido por la NOAA y el 2000-presente conjunto de datos del Centro de predicción climática de la NOAA (CPC TIR) (Chris Funk et al., 2015).

Para la creación de CHIRPS agencias meteorológicas de México, América Central, América del Sur y África Subsahariana otorgaron observaciones de estaciones para el proceso de desarrollo de CHIRPS (Chris Funk et al., 2015). En el caso específico de Colombia se realizó una validación mensual de CHIRPS (en la principal época de lluvias de cada año; comprendida entre septiembre a noviembre) utilizando 338 estaciones del IDEAM, donde se obtuvieron buenos resultados ( $R=0.97$  y MAE 21mm); las estaciones utilizadas en la validación fueron incluidas en la versión final de CHIRPS (Chris Funk et al., 2015).

### **3.2.2 Multi-Source Weighted-Ensemble Precipitation (MSWEP)**

Multi-Source Weighted-Ensemble Precipitation (MSWEP V2) es un producto global de precipitación con resolución temporal de 3 horas con datos desde 1979 hasta el presente, basado en datos de reanálisis, satélites y pluviómetros (Hylke E. Beck et al., 2019). La idea para la creación de MSWEP fue fusionar diferentes fuentes de datos de precipitación de alta calidad así; i) La media a largo plazo de MSWEP está basada en CHPClim, pero se reemplaza con datos más precisos regionales (dependiendo su disponibilidad), ii) una corrección de la precipitación a partir de las observaciones de caudal de 13762 estaciones de todo el mundo con las cuales se calcula la precipitación promedio de esas cuencas, iii) la variabilidad temporal se determinó a partir de los conjuntos de datos (CPC Unificado y GPCC), tres en teledetección satelital (CMORPH, GSMaP-MVK, y TMPA 3B42RT), y dos sobre el reanálisis del modelo atmosférico (ERA-Interim y JRA-55) (Hylke E. Beck et al., 2017).

### **3.2.3 Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN)**

PERSIANN, un sistema automatizado para la estimación de la precipitación a partir de información obtenida por teledetección utilizando redes neuronales artificiales, para la estimación de la precipitación a partir de imágenes infrarrojas de onda larga satelitales geosincrónicas (GOES-IR) con una resolución de  $0,25^\circ \times 0,25^\circ$  cada media hora. La precisión del producto de lluvia se mejora ajustando de forma adaptativa los parámetros de la red utilizando las estimaciones instantáneas de la tasa de lluvia del generador de imágenes de microondas de la Misión de medición de lluvia tropical (TRMM) (producto TMI 2A12), y los errores aleatorios se reducen aún más mediante la acumulación a una resolución de  $1^\circ \times 1^\circ$  diario. El producto basado en TMI GOES-IR-TRMM actual de los autores, denominado PERSIANN-GT, se evaluó en la región  $30^\circ\text{S}$ – $30^\circ\text{N}$ ,  $90^\circ\text{E}$ – $30^\circ\text{W}$ , que incluye el Océano Pacífico tropical y partes de Asia, Australia y las Américas.



## **4. Metodología**

Se presenta una metodología por actividades y la descripción de la zona de estudio. Con la metodología por actividades se pretende que con el avance secuencial en cada actividad se de cumplimiento a los objetivos específicos, para finalmente lograr el objetivo de la investigación.

### **4.1 Metodología por actividades**

Las actividades a realizar se describen brevemente a continuación.

#### **4.1.1 Revisión de literatura**

- **Metodologías para relleno de series**

Se realiza una revisión de literatura de metodologías estadísticas, estocásticas y de inteligencia artificial que son comúnmente usadas en la estimación de datos faltantes de series de precipitación.

- **Información proveniente de sensores remotos**

Se realiza una revisión de literatura de las fuentes de sensores remotos, especialmente de los productos satelitales disponibles de forma gratuita en internet, los cuales han sido utilizados en el relleno de datos y en general en el análisis de la calidad de información de la lluvia diaria.

### **4.1.2 Recopilación y análisis de información**

Se acude a la información espacio temporal disponible de forma gratuita en internet y en las diferentes plataformas.

- **Información de estaciones en tierra**

Se revisó la calidad y disponibilidad de diferentes bases de datos con información de precipitación medida a partir de estaciones en tierra en la zona de estudio; se examinaron las bases de datos del Centro de Datos e Indicadores Ambientales de Caldas (CDIAC), Centro Nacional de Investigaciones de Café (Cenicafé) y del Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM).

- **Información de productos satelitales de precipitación**

Se recopila la información diaria de los productos satelitales CHIRPS 2.0, MSWEP 1.0 y la información de cada 3 horas del MSWEP 2.8 (posteriormente se agregó a escala diaria) desde el año 1981 hasta el 2014 (incluidos). La justificación del uso de estas tres fuentes se presenta en el capítulo 3.2.

### **4.1.3 Selección de las metodologías para relleno de series**

De las diferentes metodologías disponibles en la literatura, se realiza una selección de aquellas que de acuerdo con la información disponible y con los buenos resultados obtenidos en otros estudios puedan ser aplicadas en la zona de estudio.

### **4.1.4 Comparación de la información satelital y las estaciones en tierra**

Para realizar las comparaciones entre la información satelital y las estaciones en tierra se usa la metodología “punto a pixel” en la cual cada una de las estaciones en tierra se compara con las celdas de los productos satelitales que correspondan con sus localizaciones (en la Figura 9 se ilustra esta metodología). En la Figura 9, la celda 2 (donde hay dos estaciones en tierra, puntos azules), cada estación se compara con el valor de la

celda 2 de manera individual, es decir, que no se realizó un promedio u otra operación matemática con las estaciones que se encuentren en una misma celda, ya que se quería evaluar el desempeño de la información espacial (productos satelitales) para complementar la información puntual (relleno de las estaciones en tierra).

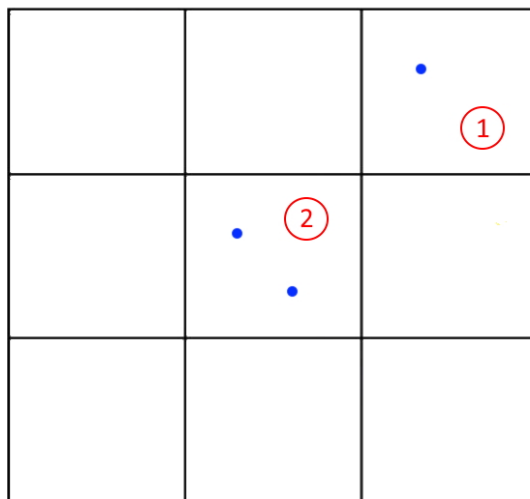


Figura 9. Metodología punto a píxel usada para comparar los productos satelitales de precipitación con respecto a las estaciones en tierra. Fuente propia.

#### 4.1.5 Relleno de series

Se aplican las metodologías seleccionadas para rellenar las series utilizando 1) solamente la información de las estaciones en tierra del IDEAM, 2) solamente la información satelital y 3) utilizando la información de las estaciones junto con la de los productos satelitales; lo anterior con el fin de comparar los resultados obtenidos utilizando las diferentes fuentes de información y evaluar la aplicabilidad de utilizar la información de los productos satelitales en el relleno de las series de precipitación diaria.

La diferencia en la escala espacial entre los productos satelitales de precipitación y las estaciones terrestres introduce un sesgo en las mediciones. Sin embargo, la hipótesis del estudio sostiene que la información proveniente de los productos satelitales de precipitación enriquece el conjunto de datos de las estaciones en tierra al proporcionar una fuente adicional de información.

#### **4.1.6 Comparación y análisis de resultados**

Con los principales resultados obtenidos de la comparación entre el relleno de datos de lluvia diaria empleando solo información de estaciones de tierra, sólo información satelital y la combinación de ambas, es posible plantear una discusión sobre el uso de datos satelitales en el relleno de datos de precipitación diaria.

#### **4.1.7 Metodologías propuestas**

Con base en los resultados obtenidos al utilizar diversas metodologías y fuentes de información para el relleno de datos, se plantea la alternativa que demuestre ofrecer los mejores resultados para completar series de lluvia diaria en el departamento de Caldas, Colombia.

#### **4.1.8 Preparación de un artículo**

Con los principales resultados de la investigación se prepara un artículo para ser sometido en una revista indexada por el Ministerio de Ciencia y Tecnología, Minciencias.

### **4.2 Esquema conceptual de la metodología**

Un esquema conceptual de la metodología propuesta se presenta en la Figura 10.

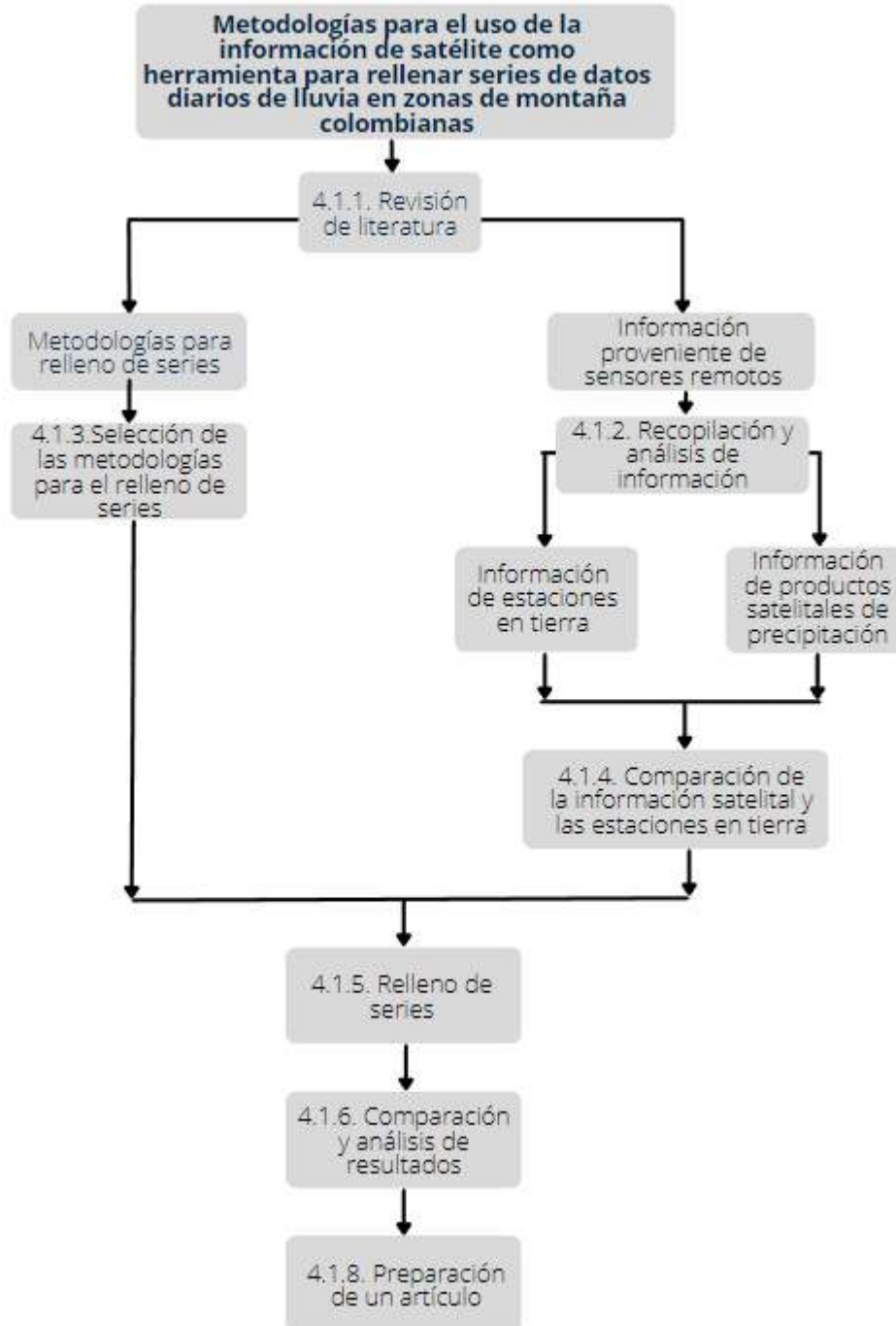


Figura 10. Esquema conceptual de la metodología.

## **5. Resultados**

### **5.1 Recopilación y análisis de información**

Se recopiló la información de las bases de datos con información de precipitación diaria medida a partir de estaciones en tierra (CDIAC, Cenicafé e IDEAM) entre los años 1981 y 2014, pero finalmente solamente se usó la información de las estaciones del IDEAM, debido a que i) las series de las estaciones del CDIAC tienen gran cantidad de datos faltantes lo cual no garantiza la continuidad de datos necesaria para hacer los análisis requeridos y ii) la información de Cenicafé solamente se encuentra disponible para el público desde el año 2006 hasta el 2020, por lo cual, se tienen solamente nueve años de concurrencia con el periodo de interés. Por lo anterior se analizaron solamente los datos del IDEAM debido a que es la fuente de datos con mayor cantidad y continuidad en las series de precipitación diaria medidas en tierra en la zona de estudio.

#### **5.1.1 Series de precipitación de las estaciones en tierra del IDEAM**

Se recopiló información de 120 estaciones pluviométricas, pluviográficas y climatológicas del IDEAM ubicadas en el departamento de Caldas (ver Figura 11), las cuales tienen datos entre 1981 y 2014 (este periodo fue escogido, debido a que es el periodo de concurrencia donde los productos satelitales utilizados tienen datos).

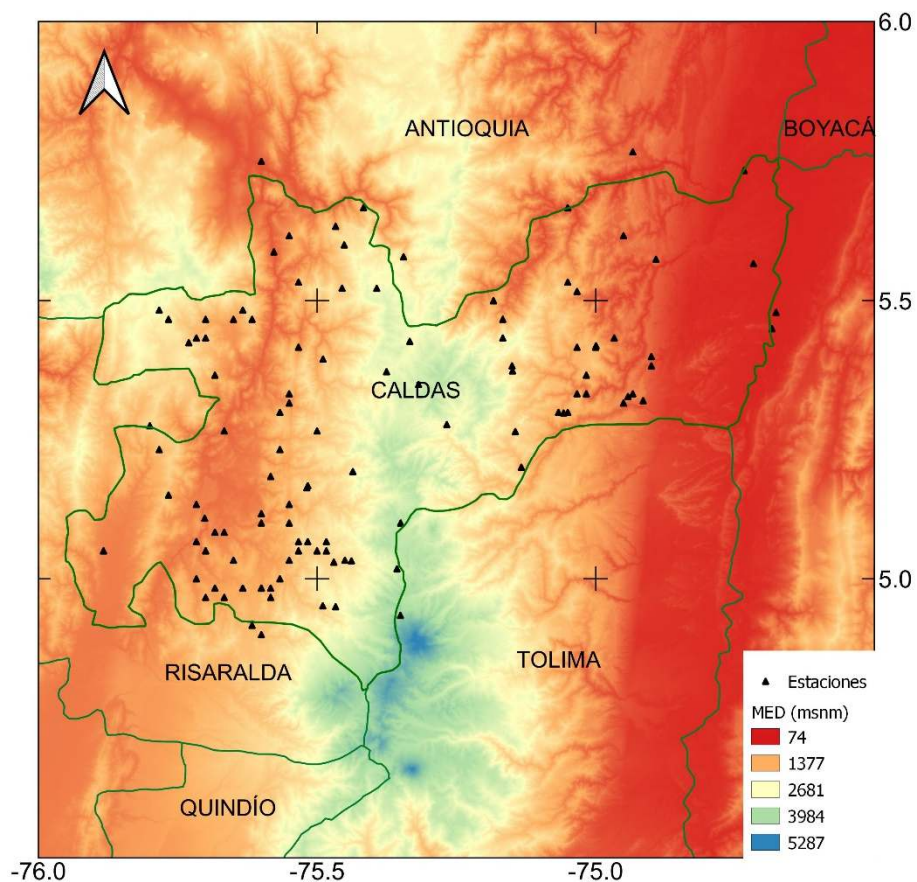


Figura 11. Localización de las estaciones en tierra del IDEAM utilizadas. Fuente propia.

Tabla 2. Estaciones en tierra del IDEAM utilizadas

Código	Estación	Latitud	Longitud	Altitud (m)
26155110	AEROPUERTO LA NUBIA	5.030	-75.470	2104
26155120	AEROPUERTO SANTAGUEDA	5.083	-75.683	933
26155050	AGRONOMIA	5.050	-75.483	2150
26180190	AGUADAS	5.600	-75.451	2400
26150060	ARAUCA	5.109	-75.702	917
23050340	ARBOLEDA	5.500	-75.183	1550
26170320	ARGENTINA LA	5.467	-75.700	1420
23050270	BELEN	5.517	-75.033	1400
26145020	BELLAVISTA	5.276	-75.800	217
23050280	BOLIVIA	5.333	-75.033	1840
26160060	BOTIJA SUB IRRA	5.267	-75.667	800

23050110	BUTANTAN	5.417	-75.000	1475
23055020	BUTANTAN	5.733	-74.733	1475
26130050	CAMPOALEGRE	4.917	-75.617	1475
23040070	CANAVERAL	5.328	-74.942	189
26130140	CAUCAYA	4.900	-75.600	1500
26155020	CENICAFE	4.983	-75.583	1310
26170240	CHANCOS LOS	5.467	-75.767	2500
23040110	CORNETA LA	5.383	-74.900	800
26160120	CRISTALINA LA	5.192	-75.436	2411
26180040	CUATRO ESQUINAS	5.667	-75.417	1900
23040090	CUBA	5.333	-74.933	1050
26170120	DESCANSO EL	5.483	-75.633	1650
26150380	DIVISA LA	4.983	-75.600	1590
26130300	ESMERALDA LA	5.067	-75.717	1040
26170350	FLECHA LA	5.467	-75.617	650
23055080	FLORENCIA	5.533	-75.050	1575
26130370	FRANCIA LA	4.983	-75.683	1360
23050200	GUACAS	5.467	-75.167	1700
26180050	GUAIMARAL	5.633	-75.467	1600
23050290	HIGUERON	5.367	-75.017	1610
23050360	HOGAR JUVENIL	5.500	-75.183	1560
23040030	IDEMA-DORADA	5.479	-74.676	192
26175020	INSTITUCION LOS FUNDADORE	5.433	-75.717	1820
26150350	JAVA	5.033	-75.550	1780
26160260	JULIA LA	5.300	-75.567	1650
26150160	LA ESPERANZA	5.018	-75.357	3341
26155150	LAS BRISAS - AUT	4.935	-75.350	4133
23020130	LEONERA LA	5.100	-75.350	3580
23050310	LIBANO	5.433	-75.167	2270
26160270	LINDA LA	5.533	-75.533	1850
23025010	LLANADAS	5.200	-75.133	1420
26160030	LLANADAS SALAMINA	5.417	-75.533	1500
26155140	LUKER GJA	5.083	-75.683	1020
26150370	MANGO EL	5.133	-75.550	1350
26170130	MANUELITA LA	5.367	-75.683	1320
23020080	MANZANARES	5.266	-75.144	1974
26170270	MARGARITA LA	5.133	-75.717	980
26160160	MARIA LA	5.588	-75.578	1480



23050080	MARQUETALIA	5.299	-75.057	1550
23020090	MARULANDA	5.278	-75.267	1700
26180260	MIRANDA LA	5.427	-75.334	2175
26150240	MONTENEGRO	4.950	-75.467	2420
26150080	MONTEVIDEO	5.000	-75.567	1370
26135050	NARANJAL	4.967	-75.700	1400
26160100	NEIRA	5.163	-75.518	198
23050250	NORCASIA RADIO	5.574	-74.892	737
23050070	NORCASIA	5.567	-74.717	800
26160090	PACORA PLAZA FERIA	5.522	-75.455	712
26140160	PALESTINA	5.483	-75.783	2560
23050170	PALMA LA	5.433	-74.967	1300
23050300	PALMA LA	5.417	-75.033	2150
26130440	PALMA LA	5.000	-75.717	1300
26160180	PAN DE AZUCAR	5.167	-75.517	1780
26150150	PAPAYAL	4.952	-75.490	2304
23020020	PASTORITA LA	5.317	-74.950	1100
26150340	PAVAS LAS	5.100	-75.600	1420
26180240	PAVAS LAS	5.522	-75.393	3210
26185010	PELADA LA	5.579	-75.345	2180
23055050	PENSILVANIA	5.383	-75.150	1975
26160170	PERNAMBUCO	5.333	-75.550	1400
26180130	PINTADA LA	5.750	-75.600	650
26150330	PLTA INSULA	5.050	-75.700	1100
26150270	POMOS LOS	4.967	-75.583	1500
26130490	PROVIDENCIA	5.050	-75.700	1250
23050180	PTE LINDA	5.667	-75.050	650
23055090	PTE LINDA	5.667	-75.050	650
26175070	RAFAEL ESCOBAR	5.467	-75.650	1320
26150280	RECAUDACION CHEC	4.983	-75.600	1340
26150090	RECREO EL	5.033	-75.650	1450
26160250	REFUGIO	5.183	-75.583	1320
26170290	RIOSUCIO	5.425	-75.730	1946
26175010	RIOSUCIO	5.433	-75.700	1810
26130530	ROMELIA LA	4.967	-75.667	1400
26160080	SALAMINA CONC BARC	5.395	-75.489	1890
26170370	SAMAN EL	5.150	-75.767	1550
23055040	SAMANA	5.419	-74.999	1532
26145050	SAMARIA LA	5.050	-75.883	120
26150250	SAN CANCIO PLANTA	5.050	-75.500	1850

23050320	SAN DANIEL	5.333	-75.017	1660
26185040	SAN FELIX	5.373	-75.375	2690
26160240	SAN ISIDRO	5.233	-75.567	1600
23050230	SAN JOSE PENSILVAN	5.374	-75.149	216
26155090	SANTAGUEDA	5.083	-75.667	1010
26150200	SELVA LA	5.117	-75.600	1350
26130470	SIERRA LA	4.983	-75.633	1450
26160140	SILVANIA	5.317	-75.550	1200
26130160	STA ANA	4.983	-75.633	1250
23025030	STA HELENA	5.300	-75.067	1450
26160130	STA HELENA	5.617	-75.550	1400
26150140	STA TERESITA	5.117	-75.600	1350
26150030	SUB ALTA SUIZA	5.067	-75.483	2055
26140170	SUB ANSERMA	5.233	-75.783	1660
26160110	SUB ARANZAZU	5.267	-75.500	1870
26150220	SUB CHINCHINA	4.983	-75.600	1340
26150050	SUB CHIPRE	5.067	-75.533	2170
23040100	SUB DORADA	5.450	-74.683	200
26150070	SUB LA URIBE	5.050	-75.533	1850
26150040	SUB MARMATO	5.067	-75.517	2072
23050330	SUB MARQUETALIA	5.300	-75.050	1550
26160150	SUB NEIRA	5.167	-75.517	2020
26140130	SUB VITERBO	5.050	-75.883	1000
23045010	TAMBOR EL	5.400	-74.900	850
23055030	TENDIDOS	5.767	-74.933	1100
26150180	TESORITO	5.033	-75.450	2200
26155170	TESORITO FINCA - AUT	5.032	-75.438	2325
26150360	TRILLADORA LA	5.100	-75.550	1650
26180250	VALLE ALTO - AUT	5.350	-75.318	2800
23055060	VERGEL EL	5.617	-74.950	650
23020100	VICTORIA LA	5.321	-74.915	760

El proceso de la evaluación de la calidad de la información temporal es importante debido a que permite encontrar datos atípicos, inconsistentes e incoherentes (IDEAM, 2019). Aunque este proceso es realizado por las autoridades ambientales para su publicación en el Sistema de Información sobre Calidad del Aire (SISAIRE), el IDEAM realiza la evaluación crítica de la calidad de los datos (IDEAM, 2019). En (IDEAM, 2019) se describe el proceso

de evaluación hecho por el IDEAM que comprende la identificación de valores atípicos (valores mínimos y máximos extremos) e inconsistencia de los datos (revisión de las relaciones existentes entre variables, consistencia de las unidades, revisión temporal de las series, comparación según tendencias interanuales, revisión de datos negativos que físicamente no pueden ser negativos, verificación de metadatos como la fecha y la hora y comparación de los reportes de información hechos por las autoridades ambientales).

Adicionalmente se evaluó la consistencia de los datos a partir de la realización de curvas de doble masa. Las curvas de doble masa permiten identificar la lluvia acumulada de una estación con respecto al promedio de la precipitación acumulada de las estaciones cercanas (en este caso se utilizó la más cercana), con el fin de comparar los comportamientos de las precipitaciones en una misma zona.

### 5.1.2 Series de precipitación de los productos satelitales

Se seleccionaron para el estudio los datos de los productos satelitales CHIRPS V2.0, MSWEP V2.8 y MSWEP V1.0 entre los años 1981 y 2014 (periodo de concurrencia en el que los cuatro productos satelitales tienen datos). Las características de estos productos se presentan en la Tabla 3.

Tabla 3. Resumen características de los productos satelitales descargados de forma gratuita en la red de redes.

Producto satelital	Nombre completo	Resolución espacial	Cobertura temporal	Resolución temporal	Fuente
CHIRPS V2.0	Climate Hazards group Infrared Precipitation with Stations Version 2.0	0.05°	1981-presente	Diaria	<a href="https://data.chc.ucsb.edu/products/CHIRPS-2.0/">https://data.chc.ucsb.edu/products/CHIRPS-2.0/</a>
MSWEP V2.8	Multi-Source Weighted-Ensemble Precipitation Version 2.2	0.1°	1979-presente	3-horas	<a href="http://www.gloh2o.org/mswep/">http://www.gloh2o.org/mswep/</a>
MSWEP V1.0	Multi-Source Weighted-Ensemble Precipitation Version 1.0	0.25°	1979-2014	Diaria	<a href="https://wci.earth2observe.eu/portal/">https://wci.earth2observe.eu/portal/</a>

La información diaria de los productos satelitales corresponde a el acumulado de precipitación desde las 00:00 a las 24:00 en hora UTC (Universal Time Coordinated); de manera equivalente para los productos satelitales con mediciones cada 3 horas. Se realizó un ajuste de la información debido a que la zona horaria colombiana es igual a la hora UTC menos 5 horas, teniendo en cuenta que las series de las estaciones del IDEAM utilizadas se clasifican como “Día pluviométrico”, es decir, que las mediciones están hechas desde las 7:00 de un día hasta las 7:00 del otro día (IDEAM, 2021a; IDEAM, 2021b; IDEAM, 2019). Por lo anterior las series de los productos satelitales se ajustaron a el horario de los días pluviométricos de las series del IDEAM (de 7:00 a 7:00 en hora colombiana), por lo cual según la hora UTC que está adelantada 5 horas con respecto a la hora colombiana, hay una diferencia de dos horas con respecto a el horario del día pluviométrico. Para los productos satelitales CHIRPS V2.0 y MSWEP V1.0 (los cuales fueron descargados con una agregación diaria) el ajuste fue realizado como se muestra en la  $PPT_{día_i} = \frac{22h}{24h} \times$

$$PPT_{día_i} + \frac{2h}{24h} \times PPT_{día_{i+1}}$$

Ecuación 1:

$$PPT_{día_i} = \frac{22h}{24h} \times PPT_{día_i} + \frac{2h}{24h} \times PPT_{día_{i+1}}$$

Ecuación 1. Ajuste a la zona horaria colombiana para los productos CHIRPS V2.0 y MSWEP V1.0. Donde  $i$  corresponde al  $i$ ésimo día y  $PPT_{día}$ , precipitación diaria

Para el MSWEP V2.8 la información satelital se descargó con una resolución temporal de cada tres horas y se agregó a información diaria sumando la precipitación desde las 6:00 del presente día, hasta las 6:00 del siguiente día (por lo que hay una diferencia de una hora entre la información del MSWEP V2.8 agregada de esta manera y la de las estaciones del IDEAM con su horario de día pluviométrico). Posteriormente se aplicó la Ecuación 2 para hacer la corrección a la zona horaria colombiana:

$$PPT_{día_i} = \frac{23h}{24h} \times PPT_{día_i} + \frac{1h}{24h} \times PPT_{día_{i+1}}$$

Ecuación 2. Ajuste a la zona horaria colombiana para la información del producto MSWEP V2.0. Donde  $i$  corresponde al  $i$ ésimo día y  $PPT_{día}$ , precipitación diaria

Para analizar la información de los tres productos satelitales se buscó el periodo de concurrencia de 5 años en donde hubiera más estaciones sin datos faltantes; este periodo se encuentra entre 1982 y 1986, en donde hay 25 estaciones sin datos faltantes (ver Figura 12 y Figura 13).

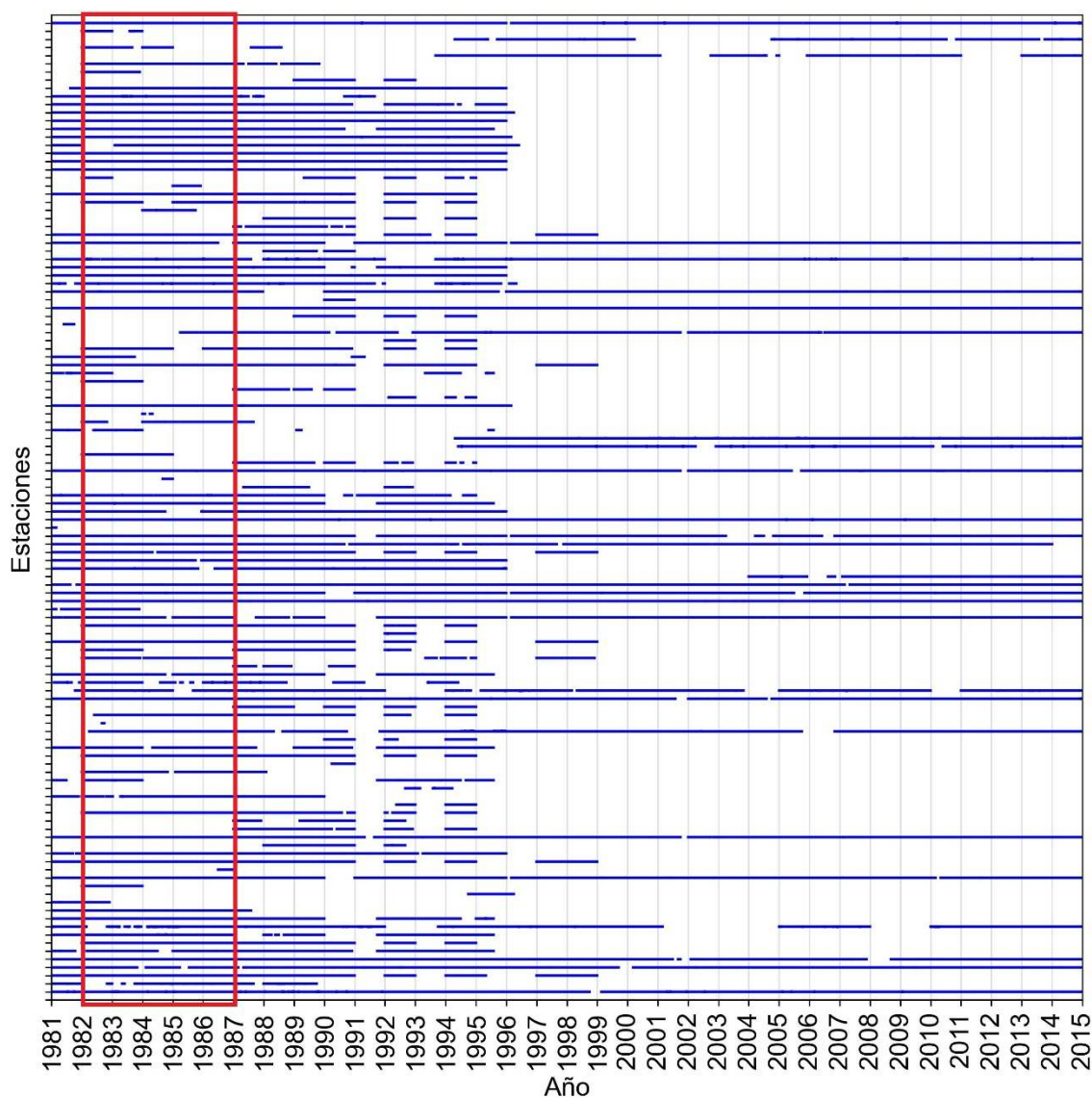


Figura 12. Completitud de datos de las estaciones en tierra del IDEAM. Fuente propia.

En la Figura 12, se destaca el periodo resaltado en rojo, que indica el rango de tiempo de 5 años (desde el año 1982 a 1986) en cual se aprecia que existen mayor número de estaciones sin datos faltantes, cuya ubicación se observa en la Figura 13. Se buscan estaciones con series de datos completas para quitar algunos datos, luego aplicar las técnicas de relleno de datos para poder comparar el dato simulado con el dato original y de esta forma estimar la precisión de cada técnica de relleno de datos faltantes. En la Tabla 4 se presentan las estaciones sin datos faltantes entre 1982 y 1986.

Tabla 4. Estaciones sin datos faltantes entre 1982 y 1986

<b>Código</b>	<b>Estación</b>	<b>Municipio</b>	<b>Latitud</b>	<b>Longitud</b>	<b>Altitud (m)</b>
26155050	AGRONOMIA	Manizales	5.050	-75.483	2150
26150060	ARAUCA	Palestina	5.109	-75.702	917
26170320	ARGENTINA LA	Palestina	5.467	-75.700	1420
26155020	CENICAFE	Chinchiná	4.983	-75.583	1310
26170240	CHANCOS LOS	Riosucio	5.467	-75.767	2500
26160120	CRISTALINA LA	Neira	5.192	-75.436	2411
26170120	DESCANSO EL	Marmato	5.483	-75.633	1650
26180050	GUAIMARAL	Aguadas	5.633	-75.467	1600
26150160	LA ESPERANZA	Manizales	5.018	-75.357	3341
26155140	LUKER GJA	Palestina	5.083	-75.683	1020
26170130	MANUELITA LA	Riosucio	5.367	-75.683	1320
26160160	MARIA LA	Aguadas	5.588	-75.578	1480
23050250	NORCASIA RADIO	Norcasia	5.574	-74.892	737
26160090	PACORA PLAZA FERIA	Pácora	5.522	-75.455	712
26150330	PLTA INSULA	Chinchiná	5.050	-75.700	1100
26175070	RAFAEL ESCOBAR	Supía	5.467	-75.650	1320
26160080	SALAMINA CONC BARC	Salamina	5.395	-75.489	1890
26155090	SANTAGUEDA	Palestina	5.083	-75.667	1010
23025030	STA HELENA	Marquetalia	5.300	-75.067	1450
26160110	SUB ARANZAZU	Aránzazu	5.267	-75.500	1870
23040100	SUB DORADA	Dorada	5.450	-74.683	200
26150070	SUB LA URIBE	Manizales	5.050	-75.533	1850
23050330	SUB MARQUETALIA	Marquetalia	5.300	-75.050	1550
26150180	TESORITO	Manizales	5.033	-75.450	2200
23020100	VICTORIA LA	Victoria	5.321	-74.915	760

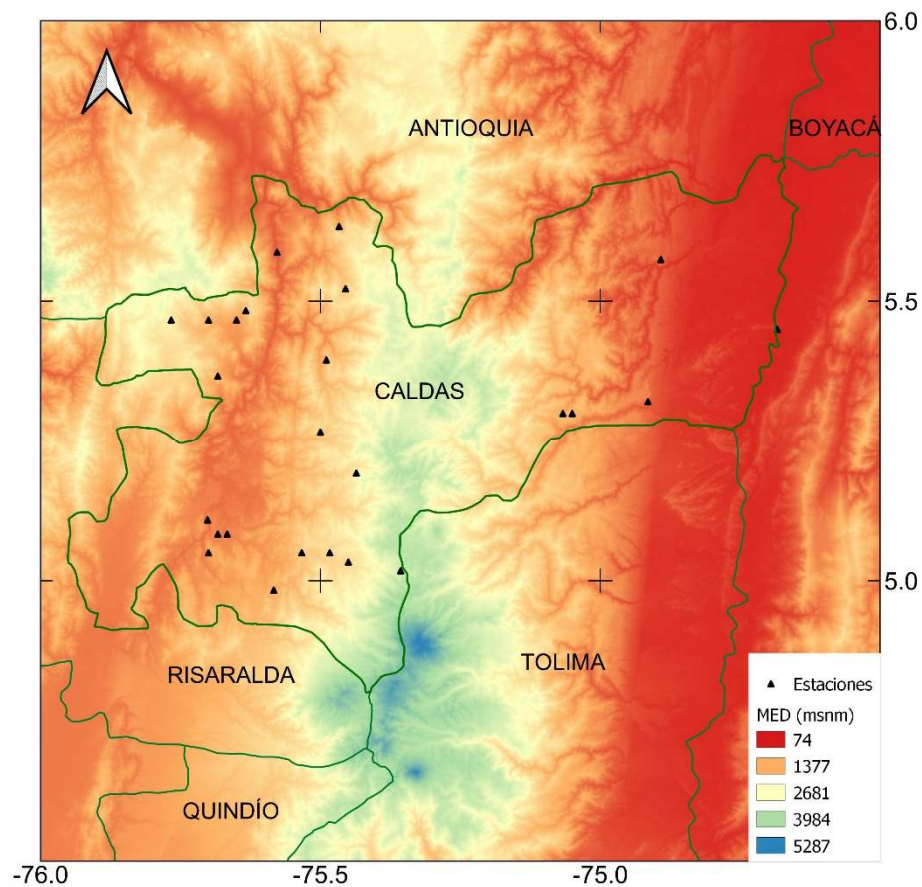


Figura 13. Estaciones sin datos faltantes entre 1982 y 1986. Fuente propia.

Para evaluar la consistencia de los datos de los productos satelitales, se realizaron curvas de masa simple con los datos de las 25 estaciones (estaciones sin datos faltantes entre 1982 y 1986), la estación más cercana y los datos de las celdas correspondientes a la ubicación de las estaciones de los tres productos satelitales.

Se realizaron las gráficas de la precipitación media mensual entre 1982 y 1986 comparando las 25 estaciones observadas en la Figura 13 y los tres productos satelitales, con lo cual se busca verificar que la información de los productos satelitales siga el comportamiento bimodal de las precipitaciones mensuales característico de la zona de estudio. En la Figura 14 y Figura 15 se presentan dos de las estaciones para las cuales se

hicieron las gráficas, donde se observa que los productos satelitales utilizados siguen el comportamiento bimodal característico de la zona de estudio.

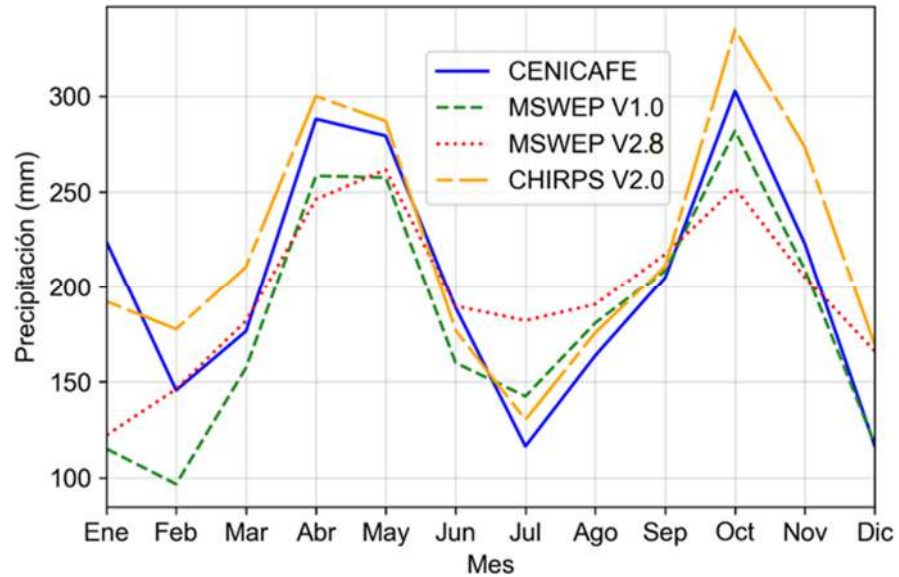


Figura 14. Precipitación media mensual entre los años 1982 a 1986 en la ubicación de la estación pluviométrica Cenicafé. Fuente propia.



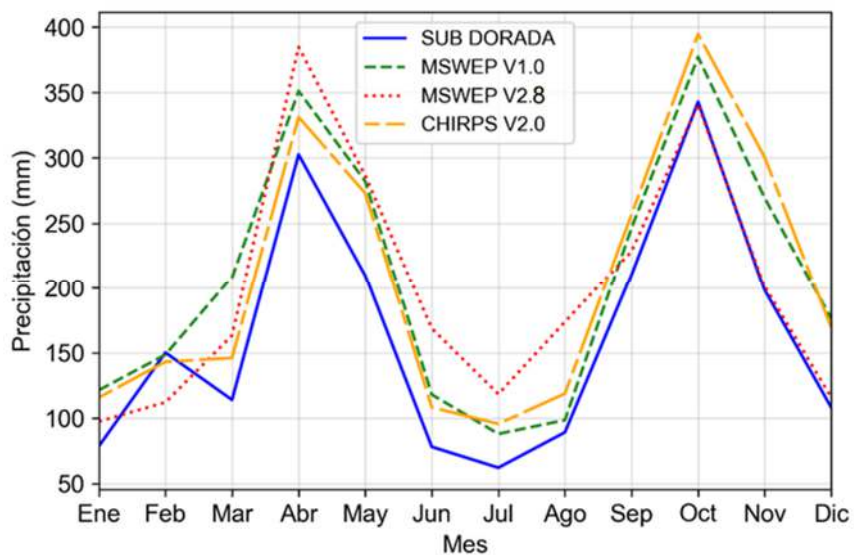


Figura 15. Precipitación media mensual entre los años 1982 a 1986 en la ubicación de la estación pluviométrica Sub Dorada. Fuente propia.

Aunque los tres productos satelitales siguen el comportamiento bimodal, se observan diferencias en la estimación de la magnitud de las precipitaciones, lo cual es entendible por la diferencia entre las escalas espaciales de los dos tipos de fuentes de información, donde las estaciones en tierra miden precipitaciones puntuales y los productos satelitales miden precipitaciones en una región (escala espacial de las mallas de datos).

Adicionalmente se evaluó la consistencia de la información satelital y de las estaciones en tierra, mediante la realización de curvas de masa simple. En la Figura 16 se presentan cuatro ejemplos para las estaciones La María (norte del departamento), Cenicafé (sur del departamento), Norcasia Radio (oriente del departamento) y Los Chancos (occidente del departamento), en donde se grafican las curvas de masa simple entre la estación objetivo, estación más cercana y los productos satelitales para el periodo 1982 a 1986.

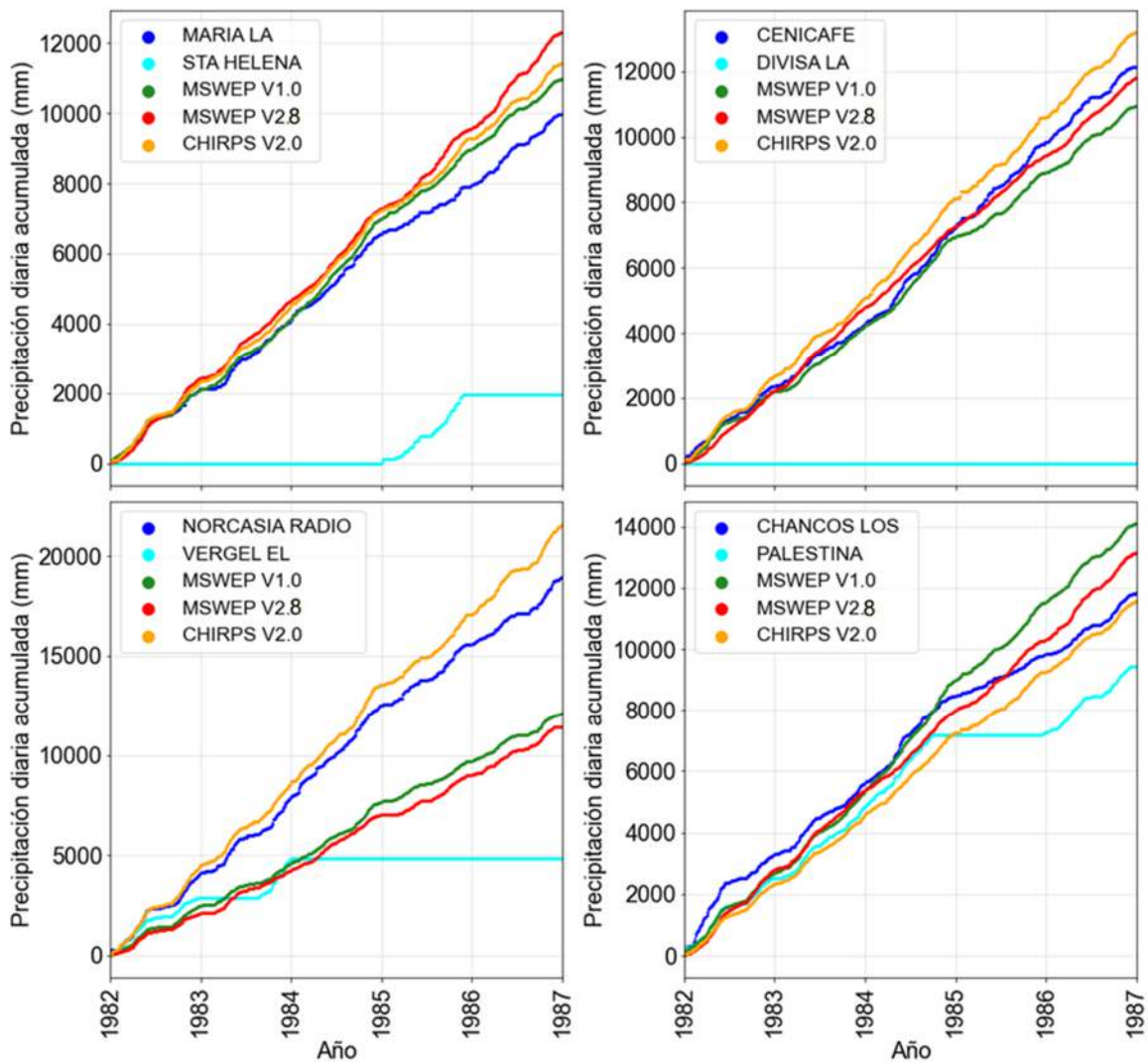


Figura 16. Curvas de masa simple. Fuente propia.

En la Figura 16 se observa que CHIRPS V2.0 es el producto que más se ajusta a las estaciones en tierra. Los productos MSWEP V1.0 y MSWEP 2.8 al compararlos con la mayoría de estaciones son consistentes y se ajustan bien. Además, se observa la carencia de información de las estaciones más cercanas, donde existen grandes periodos donde no se tienen datos.

Todo este análisis realizado a las estaciones en tierra y a la información satelital sirve como un análisis de la calidad de información, ya que se ha evaluado la homogeneidad de los datos y la consistencia de la información recopilada. Adicionalmente, se lleva a cabo un análisis de tendencias a cada una de las series temporales para observar si las series siguen un patrón. Se pretende con todo este análisis, tener un conocimiento más amplio sobre las series utilizadas en el estudio.

Se aplicó el test de Mann Kendall con el fin de evaluar si existe una tendencia significativa en el periodo entre los años 1982 y 1986 para las 25 estaciones sin datos faltantes en este periodo. Para realizar esta prueba se tomó un nivel de significancia de 0.05. En la Tabla 5 se presentan los resultados del test; donde Z es el estadístico de prueba y P-Value es la probabilidad que tiene de ocurrir el estadístico calculado a partir de los datos.

Tabla 5. Resultados test de Man Kendall para las 25 estaciones sin datos faltantes entre 1982 y 1986.

Estaciones	Z	P-Value	Evidencia de tendencia
AGRONOMIA [26155050]	-1.29	0.20	No hay tendencia
ARAUCA [26150060]	2.67	0.01	Hay tendencia
ARGENTINA LA [26170320]	0.90	0.37	No hay tendencia
CENICAFE [26155020]	-0.64	0.53	No hay tendencia
CHANCOS LOS [26170240]	-4.70	0.00	Hay tendencia
CRISTALINA LA [26160120]	-0.22	0.83	No hay tendencia
DESCANSO EL [26170120]	-0.94	0.35	No hay tendencia
GUAIMARAL [26180050]	-1.95	0.05	No hay tendencia
LA ESPERANZA [26150160]	-7.87	0.00	Hay tendencia
LUKER GJA [26155140]	-0.62	0.54	No hay tendencia
MANUELITA LA [26170130]	0.58	0.56	No hay tendencia
MARIA LA [26160160]	-1.36	0.17	No hay tendencia
NORCASIA RADIO [23050250]	-2.44	0.01	Hay tendencia
PACORA PLAZA FERIA [26160090]	-1.85	0.06	No hay tendencia
PLTA INSULA [26150330]	0.82	0.41	No hay tendencia
RAFAEL ESCOBAR [26175070]	0.88	0.38	No hay tendencia
SALAMINA CONC BARC [26160080]	-0.34	0.73	No hay tendencia
SANTAGUEDA [26155090]	-1.18	0.24	No hay tendencia
STA HELENA [23025030]	0.28	0.78	No hay tendencia
SUB ARANZAZU [26160110]	-2.65	0.01	Hay tendencia

SUB DORADA [23040100]	0.64	0.52	No hay tendencia
SUB LA URIBE [26150070]	-3.27	0.00	Hay tendencia
SUB MARQUETALIA [23050330]	1.22	0.22	No hay tendencia
TESORITO [26150180]	-1.36	0.17	No hay tendencia
VICTORIA LA [23020100]	-0.38	0.70	No hay tendencia

Se observa que de las 25 estaciones sin datos faltantes durante el periodo comprendido entre 1982 y 1986, hay seis de ellas con evidencia de una tendencia. Esto no afecta las metodologías de relleno empleadas en el estudio, ya que estas no predicen escenarios futuros, si no que se basan en la información de los días a completar de las estaciones vecinas o de los productos satelitales.

En la Figura 17 y Figura 18 se presentan dos ejemplos de curvas de doble masa para las estaciones NORCASIA RADIO [23050250] y ARAUCA [26150060]. Donde se observa que la información entre las estaciones y CHIRPS V2.0 es consistente, ya que no se aprecian cambios significativos en la pendiente de las curvas.

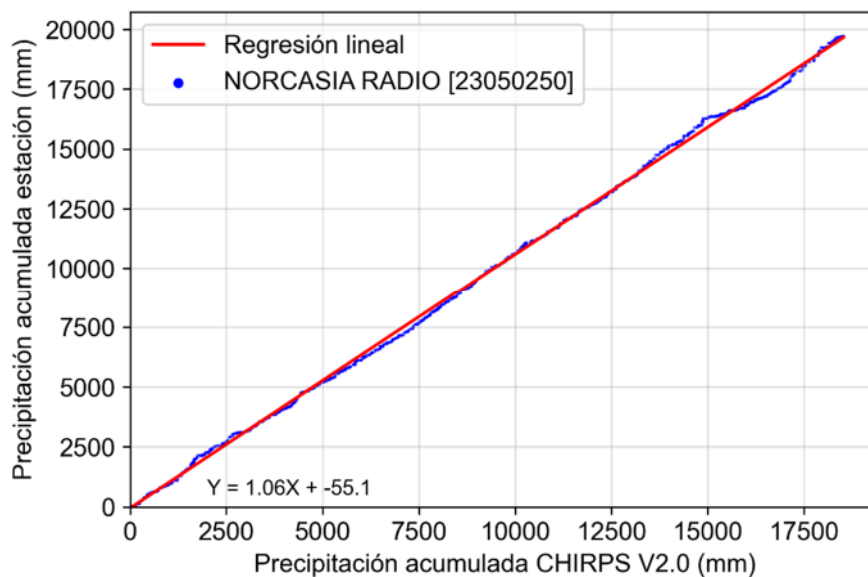


Figura 17. Curva de doble masa de precipitación entre la estación NORCASIA RADIO [23050250] y CHIRPS V2.0, entre los años 1982 y 1986

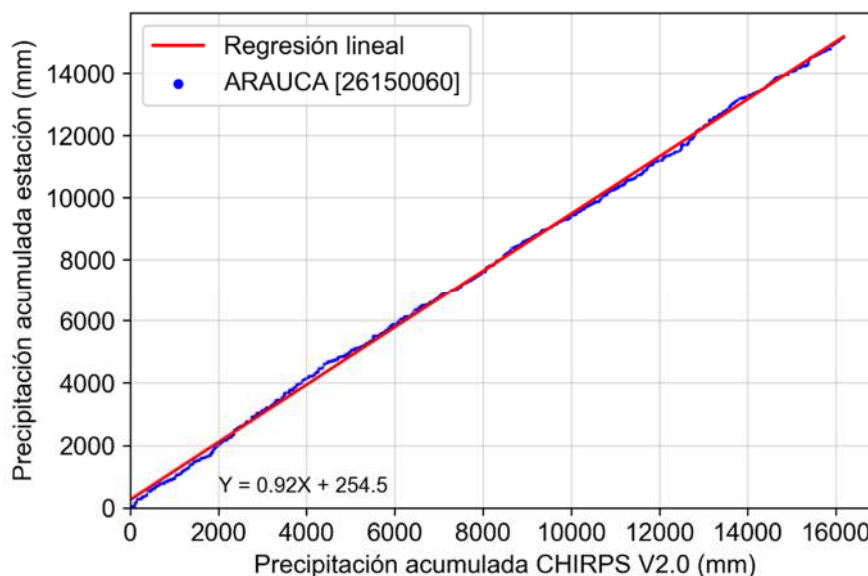


Figura 18. Curva de doble masa de precipitación entre la estación ARAUCA [26150060] y CHIRPS V2.0, entre los años 1982 y 1986

### 5.1.3 Definición de las zonas homogéneas climáticamente

En este apartado se exploran las relaciones que permitan encontrar zonas homogéneas climáticamente, en las cuales sea posible utilizar de forma confiable las técnicas de relleno de datos, por lo que si se analiza la información de la geología del departamento de Caldas, las series de lluvia diarias, la distribución de la temperatura máxima, media y mínima, que es función de la altitud y con el soporte de la clasificación climática del IDEAM (ver capítulo 1.3), se propone inicialmente para este estudio dos zonas climáticas, las cuales corresponden a las vertientes del oriente (valle del río Magdalena) y del occidente (valle del río Cauca) en el departamento de Caldas. Además, se observan las relaciones de la lluvia con la altitud, de tal forma que sea posible incluir una variable adicional para mejorar la calidad en el relleno de datos.

La relación entre la lluvia y la elevación es un apoyo para la definición áreas homogéneas, ya que, si se logra identificar una relación con la altitud por regiones, es posible utilizar este criterio para definir zonas homogéneas. Sin embargo, un análisis realizado con las distribuciones de la lluvia, tanto para el satélite como para las estaciones de tierra, y un método para desagregar espacialmente la información satelital con base en la altura; no muestra un buen desempeño en la desagregación, ni una relación clara entre la altitud y

precipitación anual de las estaciones. Por lo anterior, aunado a que el interés de la investigación se centra en lluvias diarias, se toma la decisión de utilizar como única región para el análisis y evaluación de la información al departamento de Caldas.

El análisis realizado consistió en seleccionar la mayor cantidad de estaciones que presentaran cinco años consecutivos sin datos faltantes (ver Tabla 6); estas fueron divididas en dos zonas climáticas (según su ubicación en la vertiente del Cauca y la vertiente del Magdalena), ya que estas vertientes representan la principal red hidrográfica del departamento. A partir de las gráficas de altitud contra precipitación media anual, se ajustaron funciones que describen la relación entre ambas variables (ver Figura 19 y Figura 20).

Tabla 6. Estaciones utilizadas para el análisis de la relación altura vs precipitación

Código	Estación	Municipio	Lat.	Long.	Alt. (m)	Precip anual prom (mm)	Vertiente
26180050	GUAIMARAL	Aguadas	5.633	-75.467	1600	1682.0	Cauca
26160160	MARIA LA	Aguadas	5.588	-75.578	1480	1983.2	Cauca
26160090	PACORA PLAZA FERIA	Pácora	5.522	-75.455	712	2686.2	Cauca
26160080	SALAMINA CONC BARC	Salamina	5.395	-75.489	1890	1949.2	Cauca
26160110	SUB ARANZAZU	Aranzazu	5.267	-75.500	1870	1958.9	Cauca
26170320	ARGENTINA LA	Palestina	5.467	-75.700	1420	2523.4	Cauca
26170240	CHANCOS LOS	Riosucio	5.467	-75.767	2500	2180.8	Cauca
26170120	DESCANSO EL	Marmato	5.483	-75.633	1650	1824.4	Cauca
26170130	MANUELITA LA	Riosucio	5.367	-75.683	1320	2095.0	Cauca
26175070	RAFAEL ESCOBAR	Supía	5.467	-75.650	1320	1951.6	Cauca
26155050	AGRONOMIA	Manizales	5.050	-75.483	2150	1889.4	Cauca
26150060	ARAUCA	Palestina	5.109	-75.702	917	2224.0	Cauca
26155020	CENICAFE	Chinchiná	4.983	-75.583	1310	2353.3	Cauca
26160120	CRISTALINA LA	Neira	5.192	-75.436	2411	1970.0	Cauca
26150160	LA ESPERANZA	Manizales	5.018	-75.357	3341	1431.4	Cauca

26155140	LUKER GJA	Palestina	5.083	-75.683	1020	2052.6	Cauca
26150330	PLTA INSULA	Chinchiná	5.050	-75.700	1100	2453.2	Cauca
26155090	SANTAGUEDA	Palestina	5.083	-75.667	1010	2030.7	Cauca
26150070	SUB LA URIBE	Manizales	5.050	-75.533	1850	2142.8	Cauca
26150180	TESORITO	Manizales	5.033	-75.450	2200	2037.0	Cauca
23040030	IDEMA-DORADA	Dorada	5.479	-74.676	192	1928.8	Magdalena
23050250	NORCASIA RADIO	Norcasia	5.574	-74.892	737	3714.3	Magdalena
23025030	STA HELENA	Marquetalia	5.300	-75.067	1450	3597.2	Magdalena
23040100	SUB DORADA	Dorada	5.450	-74.683	200	1896.7	Magdalena
23050330	SUB MARQUETALIA	Marquetalia	5.300	-75.050	1550	2650.1	Magdalena
23020100	VICTORIA LA	Victoria	5.321	-74.915	760	3679.1	Magdalena

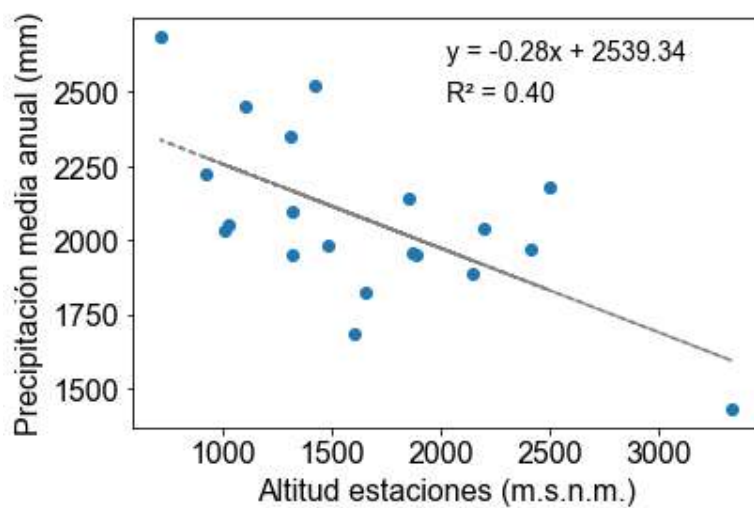


Figura 19. Relación entre la altitud de las estaciones de la vertiente del Cauca y precipitación media anual entre los años 1982 y 1986

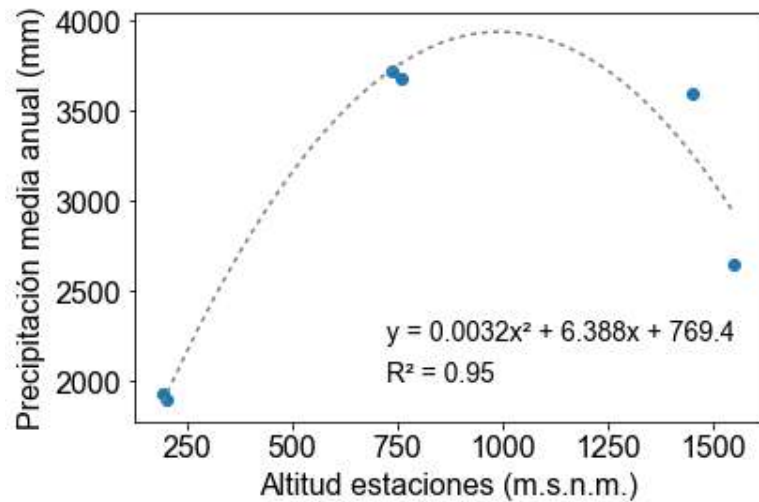


Figura 20. Relación entre la altitud de las estaciones de la vertiente del Magdalena y precipitación media anual entre los años 1982 y 1986

Se observó una relación baja entre la altitud y la precipitación anual para las estaciones ubicadas en la vertiente del Cauca (Figura 19). Sin embargo, debido a la limitada variabilidad altitudinal de las estaciones consideradas en el análisis, la fiabilidad de esta relación es limitada, dado que el departamento presenta altitudes por debajo de los 100 m.s.n.m y superiores a los 5000 m.s.n.m. Mientras que, para las estaciones de la vertiente del Magdalena, las cuales se presentan en la Figura 20 si se observa una mejor relación, sin embargo, esta relación fue calculada con base en solo seis estaciones, por lo que no se considera confiable.

El clima entendido como el conjunto de condiciones atmosféricas promedio y variaciones a largo plazo en una región determinada, es también otro factor importante para poder utilizar estaciones vecinas en el relleno de datos, ya que, si las condiciones geomorfológicas de la zona y el clima son similares, se esperan patrones de lluvia similar, por lo tanto, sería una variable factible para utilizar. Pero dada la alta variabilidad de climas en el departamento, no es posible utilizar esta variable (ver Figura 4). Además, como se observa en la Figura 4, Figura 6 y Figura 7, no se observa un patrón evidente entre el clima, los suelos y las coberturas de suelo en el departamento de Caldas.



Finalmente, dado que no se aprecia una relación de la lluvia con la elevación de forma contundente, hay alta variabilidad de climas en el departamento y la geología regional se caracteriza por rocas volcánica, con presencia de cenizas volcánicas y zonas aluviales en la zona adyacente al río Magdalena, se decide considerar a todo el departamento como una sola región homogénea en la cual sea posible utilizar las diferentes técnicas de relleno de datos. Lo anterior es una simplificación que toca adoptar por no encontrar relaciones claras.

Las técnicas de desagregación espacial permiten mejorar la precisión de las mediciones de los productos satelitales de precipitación con datos con una mejor resolución espacial. En general, las técnicas de desagregación espacial se basan en el uso de variables ambientales relacionadas con la lluvia, como la vegetación, la topografía o la ubicación espacial. En estudios como (Vergara et al., 2018; F. Chen et al., 2014; Xu et al., 2015), se utilizan diversas fuentes de información, como el Índice de Vegetación de Diferencias Normalizadas (NDVI), el Modelo de Elevación Digital (DEM) o las coordenadas geográficas (latitud y longitud), como variables adicionales para alimentar las técnicas de desagregación espacial. Estas variables demuestran claramente su relación con la lluvia en escalas temporales mayores a la diaria, como mensuales o anuales. Sin embargo, encontrar una relación clara para desagregar la lluvia a escala diaria (como en el caso de estudio) resulta difícil, ya que se considera que la lluvia diaria es aleatoria. Teniendo en cuenta lo expuesto anteriormente, se ha decidido no evaluar técnicas de desagregación espacial en este estudio; esto se debe a la falta de una relación contundente entre la precipitación anual y la altitud, como se puede observar en la Figura 19 y Figura 20. Además, es importante destacar que la escala temporal de la lluvia en este estudio es diaria, lo que agrega mayor complejidad a la tarea de desagregarla.

## 5.2 Comparación de la información satelital y las estaciones en tierra

### 5.2.1 Comparación estadística de la información satelital (magnitud de las precipitaciones)

Se evaluó el desempeño por pares de la cantidad de precipitación estimada por los satélites con respecto a las estaciones en tierra del IDEAM. Se realizaron evaluaciones punto a pixel, es decir, que cada estación en tierra del IDEAM se comparó con su respectiva celda asociada a los productos satelitales según su ubicación (CHIRPS V2.0 y MSWEP V2.8 y MSWEP V1.0). En la Tabla 7 se presentan los estadísticos usados para realizar esta evaluación. Estas evaluaciones fueron hechas con las resoluciones originales de cada producto satelital; con conocimiento de la gran diferencia entre escalas espaciales de la información de las estaciones con respecto a la de los satélites.

Tabla 7. Estadísticos de comparación por pares. Fuente: basado en (Toté et al., 2015)

Estadístico	Fórmula	Descripción
Mean Error	$ME = \frac{1}{N} \sum (S - G)$	Mide el error promedio de la estimación del satélite con respecto a la estación (en milímetros). El valor perfecto es de 0.
Relative mean absolute error	$RMAE = \frac{1}{N} \sum \frac{ S - G }{\bar{G}}$	Mide el error promedio de la estimación del satélite con respecto a la estación (en fracción). El valor perfecto es de 0.
Nash-Sutcliffe Efficiency coefficient	$Eff = 1 - \frac{(\sum(S - G)^2)}{(\sum(G - \bar{G})^2)}$	Mide el ajuste de las estimaciones con respecto a la media de los valores observados (estaciones). Si el valor es negativo, significa que el valor medio observado es mejor predictor que las estimaciones satelitales, si es igual a 0, quiere decir que el valor medio observado es tan bueno como las estimaciones satelitales y si es 1 quiere decir que las estimaciones satelitales tienen una coincidencia perfecta con los valores observados por las estaciones.
Bias	$Bias = \frac{\sum S}{\sum G}$	Mide la correspondencia entre el valor medio de las estimaciones satelitales y de las estaciones. El valor perfecto es de 1.

Donde  $G$  es la medición de la estación,  $\bar{G}$  el promedio de las mediciones de la estación,  $S$  la estimación del satélite,  $\bar{S}$  el promedio de las estimaciones del satélite y  $N$  el número de pares de datos.

En la Tabla 8 se presentan los resultados de los productos satelitales en la estimación de la magnitud de las precipitaciones diarias con respecto a las 25 estaciones sin datos faltantes entre 1982 y 1986 (ver Figura 13).

Tabla 8. Resultados de los productos satelitales en la estimación de la magnitud de las precipitaciones diarias entre 1982 y 1986

Producto satelital	ME (mm)	RMAE	Nash-Sutcliffe	Bias
<b>CHIRPS V2.0</b>	0.23	1.32	-0.40	1.05
<b>MSWEP V2.8</b>	0.06	1.29	-0.17	1.06
<b>MSWEP V1.0</b>	-0.08	1.08	0.09	1.03

Se observa que en promedio el producto que tiene mayores errores para estimar la precipitación diaria es CHIRPS V2.0, lo cual se ve reflejado en los valores más altos de ME y RMAE, y el menor valor del coeficiente de Nash. También se observa que el producto con mejor desempeño es el MSWEP V1.0 el cual obtuvo el mejor coeficiente de Nash (siendo cercano a 0, es decir, bajo ajuste) y el menor RMAE. Aunque se esperaba que CHIRPS V2.0 fuera el producto con el mejor desempeño para estimar la precipitación diaria, debido a que es el producto con la mejor resolución espacial entre los tres utilizados (resolución de  $0.05^\circ$ ), el que obtuvo el mejor desempeño fue el MSWEP V1.0 (producto que tiene la menor resolución;  $0.25^\circ$ ).

Los tres productos presentan el indicador de Bias muy similar y los tres son cercanos a 1, lo cual refleja que los productos son buenos para estimar las precipitaciones acumuladas; así haya diferencia entre la escala espacial de las estaciones y los productos. Lo anterior también se encontró en la Figura 16, donde se presentaron las curvas de masa simple y se observó que CHIRPS V2.0 presenta el mejor desempeño (MSWEP V2.0 y V2.8 también obtuvieron un buen comportamiento en este análisis de precipitación acumulada). Que los valores sean mayores a 1 demuestra que los tres productos sobrestiman las precipitaciones acumuladas; así sea en una baja magnitud.

En la Figura 21 se presenta el análisis mensual multianual de los mismos estadísticos en el mismo periodo (1982 y 1986), para observar el desempeño mensual de la estimación de precipitaciones diarias

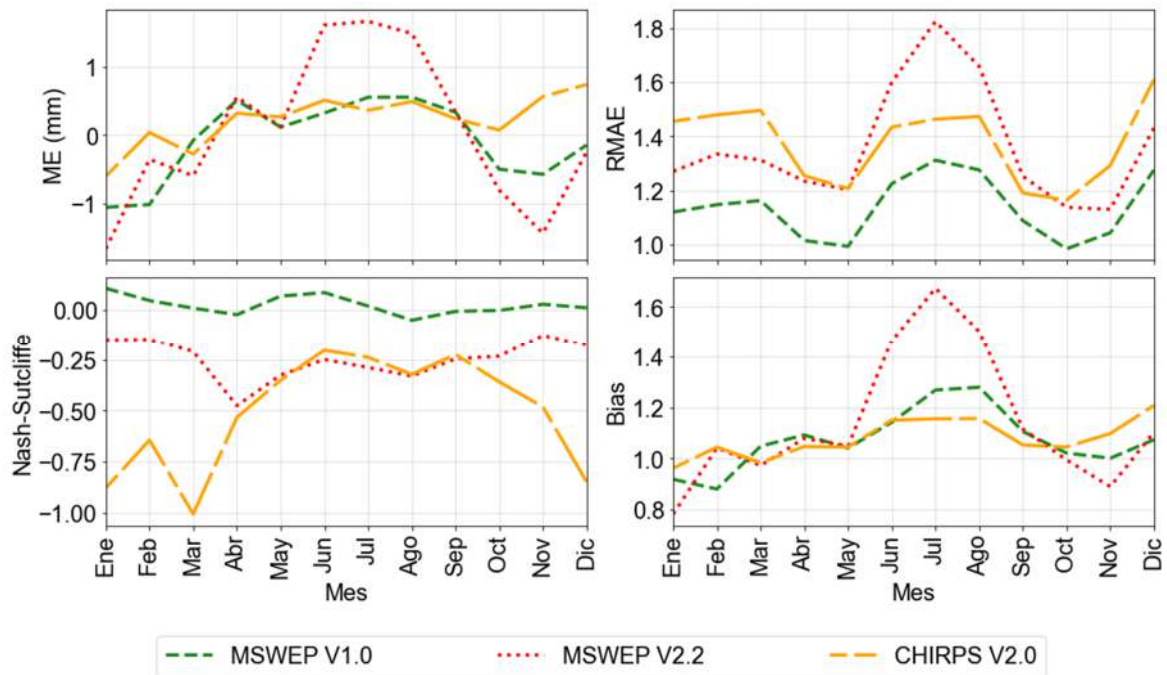


Figura 21. Resultados por mes del desempeño de los productos satelitales en la estimación de la magnitud de las precipitaciones diarias entre 1982 y 1986. Fuente propia.

El MSWEP V1.0 es el producto que presenta menor RMAE durante todos los meses del año, lo cual finalmente se ve reflejado en menor RMAE a nivel general, esto también se observa en el coeficiente de Nash donde obtuvo los mejores resultados durante todo el año. Durante la mayoría de meses, el MSWEP V2.8 es el satélite que sobreestima más las precipitaciones.

Aunque para los tres productos el ME es muy bajo (lo cual indica un alto desempeño en la estimación de la precipitación diaria), el coeficiente de Nash para los tres es muy bajo, es decir, que los errores positivos y negativos se compensan al sumarlos para calcular el ME. Lo anterior corresponde con el buen desempeño que se obtiene al estimar la precipitación acumulada (los errores positivos y negativos se compensan al sumar las precipitaciones

diarias, haciendo que la estimación de la precipitación acumulada sea buena). En conclusión, los tres productos obtienen un bajo desempeño para calcular las precipitaciones diarias, lo cual se puede explicar debido a la alta diferencia entre la escala espacial de los productos y las estaciones en tierra (información por área e información puntual respectivamente). Asimismo, se evidencia la importancia de calcular diferentes estadísticos que se complementen para evaluar las series de precipitación diaria.

### 5.2.2 Comparación estadística de la información satelital (detección de los eventos de precipitación)

También se evaluó el desempeño de los productos satelitales para detectar los eventos de lluvia diaria. Para esto se utilizaron los estadísticos presentados en la Tabla 9.

Para clasificar los días en que hubo lluvia y en los que no hubo se utilizó un límite de 0.2 mm, esto debido a que las lecturas diarias del IDEAM se redondean a los 0.2 mm más cercanos de acuerdo con la precisión de los aparatos de medida. Sin embargo, la OMM (2013) sugiere valores mayores a 1 mm para definir el umbral de lluvia, pero este umbral puede ser adecuado para balances hidrológicos a largo plazo y cambios en este umbral no se exploran en este estudio.

Tabla 9. Estadísticos para evaluar la detección de eventos de precipitación. Fuente: basado en (Toté et al., 2015)

Estadístico	Fórmula	Descripción
Probability of detection	$POD = \frac{A}{(A + C)}$	Proporción de eventos medidos por la estación que fueron estimados por el satélite. El valor perfecto es de 1.
False alarm ratio	$FAR = \frac{B}{(A + B)}$	Proporción de eventos estimados por el satélite, que no fueron medidos por la estación. El valor perfecto es de 0.
Equitable threat score	$ETS = \frac{A - Ar}{(A + B + C - Ar)}$ $Ar = \frac{((A + C)(A + B))}{N}$	Proporción de eventos medidos por la estación y que fueron estimados por el satélite correctamente (ajustados por aciertos relacionados con probabilidad aleatoria). El valor perfecto es de 1.

Hansen and Kuipers discriminate	$HK = \frac{A}{(A + C)} - \frac{B}{(B + D)}$	Mide que tan bien el satélite estima los eventos y no eventos de precipitación. El valor perfecto es de 1.
Heidke Skill Score	$HSS = \frac{(2(AD - BC))}{((A + C)(C + D) + (A + B)(B + D))}$	Mide la proporción de aciertos del satélite (después de eliminar los aciertos asociados a la probabilidad aleatoria). El valor perfecto es de 1.
Frequency Bias	$FB = \frac{A + B}{(A + C)}$	Relación entre la frecuencia de eventos estimados por el satélite y los observados por la estación. El valor perfecto es de 1.

Donde  $A$  representa el número de eventos estimados por el satélite que realmente ocurrieron (medidos por la estación),  $B$  los eventos detectados por el satélite que no fueron observados por la estación (eventos que no ocurrieron),  $C$  eventos que no fueron detectados por el satélite pero que si fueron observados por la estación y  $D$  el número de eventos que no fueron detectados por el satélite y tampoco observados por la estación.

En la Tabla 10 se presentan los resultados del desempeño de los productos satelitales para detectar los eventos diarios de lluvia entre los años 1982 y 1986. Estos resultados fueron obtenidos al comparar los productos satelitales con las 25 estaciones de la Figura 13.

Tabla 10. Resultados del desempeño de los productos satelitales en la detección de eventos diarios de precipitación para el periodo entre 1982 y 1986

Producto satelital	POD	FAR	ETS	HK	HSS	FB
<b>CHIRPS V2.0</b>	0.76	0.44	0.13	0.23	0.22	1.42
<b>MSWEP V2.8</b>	0.95	0.51	0.04	0.08	0.08	1.99
<b>MSWEP V1.0</b>	0.99	0.49	0.08	0.15	0.14	2.02

MSWEP V2.8 y V1.0 presentaron los mayores valores de POD, lo que significa que aciertan la mayor cantidad de días donde hay lluvia, pero también tienen los valores más altos de FAR, es decir, que estiman muchos días con lluvia en los que realmente no se presentaron precipitaciones. Asimismo, el producto con mayor cantidad de aciertos (días

de lluvia y de no lluvia) es el CHIRPS V2.0; obteniendo los mejores indicadores de ETS, HK y HSS.

Los tres productos satelitales sobrestiman la cantidad de días con lluvia (FB mayores a 1), siendo el MSWEP V1.0 el producto que más sobrestima la cantidad de días con precipitación y el CHIRPS V2.0 el que menos los sobrestima (FB más cercano a 1).

Según lo anterior el producto con el mejor desempeño en la estimación de los eventos diarios de precipitación es el de la mayor resolución espacial (CHIRPS V2.0), ya que es el que tiene la mayor cantidad de aciertos para detectar días con lluvia, así como los días sin lluvia.

En la Figura 22 se observa que en promedio durante 11 meses del año CHIRPS V2.0 presenta la mayor cantidad de aciertos de días con precipitaciones y sin precipitaciones (menos enero) ya que obtuvo mayores valores de ETS, HK y HSS. Además, durante todos los meses del año sobrestima en menor medida los eventos con lluvia (menor FB), estima menos falsas alarmas de eventos con lluvia (menor FAR) y detecta en menor proporción los eventos con lluvia (mejor desempeño de MSWEP V2.8 y V1.0). Lo anterior, demuestra que, si se analiza de manera mensual, se mantiene el mismo comportamiento que se había presentado al hacer el análisis anual, en donde CHIRPS V2.0 presenta el mejor desempeño.

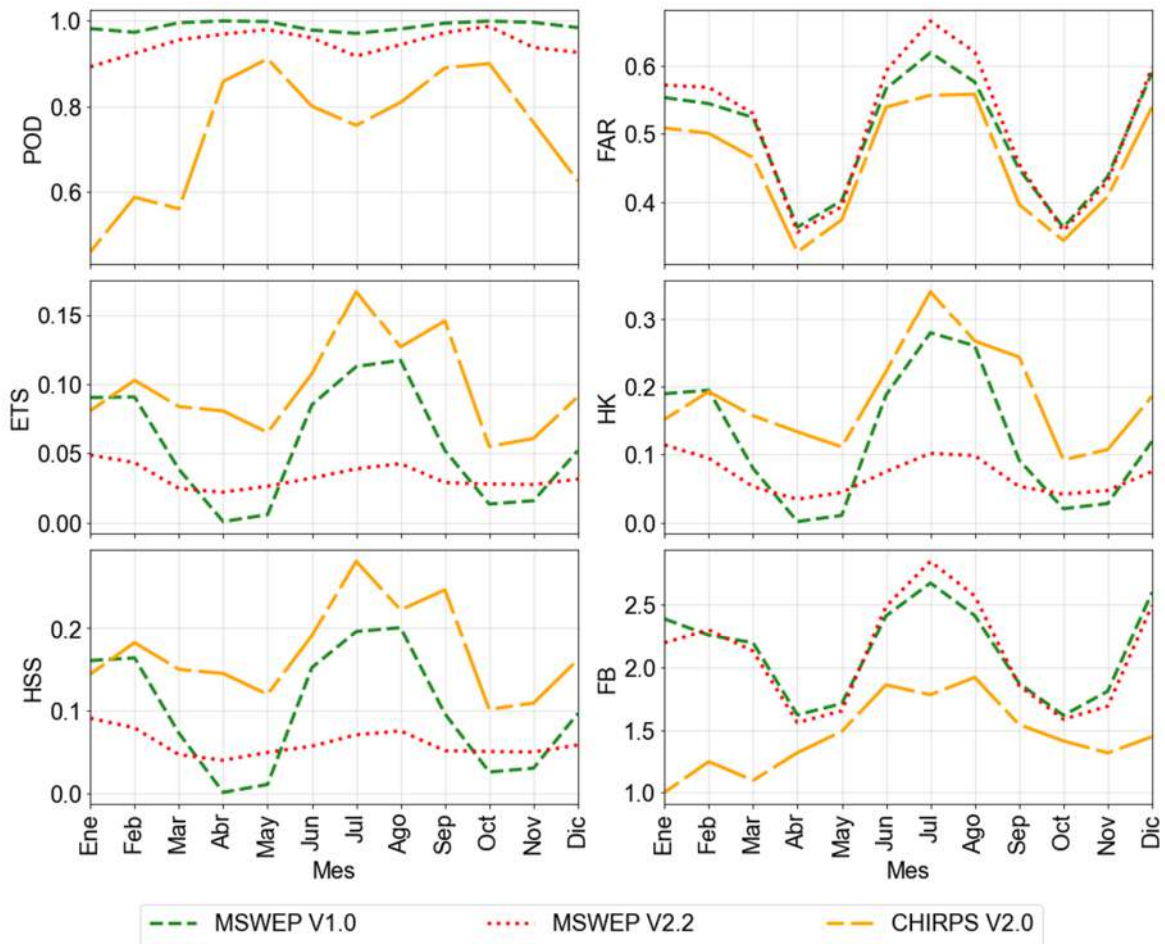


Figura 22. Resultados por mes del desempeño de los productos satelitales en la detección de eventos diarios de precipitación para el periodo entre 1982 y 1986. Fuente propia.

Además, se observa que CHIRPS V2.0 presenta un comportamiento bimodal para detectar los eventos de lluvia, ya que en los periodos secos los detecta en menor proporción y en los periodos lluviosos los detecta en mayor proporción. Los tres productos presentan comportamiento bimodal en el FAR ya que en los periodos lluviosos tienen menor cantidad de falsas alarmas (proporción de eventos estimados por el satélite, que no fueron medidos por la estación) y en los periodos secos, mayor.



---

## **5.3 Técnicas de relleno de datos faltantes de lluvia diaria**

A continuación, se presentan las técnicas o estrategias de relleno de datos faltantes de lluvia diaria seleccionadas para este estudio, y se incluyen los períodos seleccionados para realizar la calibración y validación de cada una de estas técnicas.

### **5.3.1 Selección de las metodologías para el relleno de series**

Inicialmente, dentro de los objetivos específicos del estudio se planteó la selección y aplicación de al menos una metodología por enfoque (estadístico, estocástico e inteligencia artificial). Sin embargo, en total se seleccionaron cinco metodologías para el estudio. Para el enfoque estadístico, se optó por las metodologías IDW, NRM y SBE; en cuanto al enfoque estocástico, se seleccionaron las Cadenas de Markov de primer orden; mientras que para el enfoque de inteligencia artificial se eligió la metodología MLP.

Estas selecciones se realizaron considerando que una de las principales razones del estudio, es evaluar si la información de los productos satelitales de precipitación contribuye a las metodologías utilizadas actualmente para completar series de lluvia diaria en el departamento de Caldas, Colombia, ya que normalmente, estas metodologías se basan en el uso de las series diarias de las estaciones vecinas como insumo para completar los datos faltantes. Por lo tanto, se eligieron metodologías representativas de cada enfoque que hayan sido tradicionalmente utilizadas en el campo.

En algunas áreas, el método de Kriging muestra un desempeño superior al método de IDW, ya que su comportamiento depende de las características específicas de la zona en la que se aplica. Un estudio realizado por Cifuentes, (2016) en el departamento de Caldas, comparó el rendimiento de diferentes metodologías de interpolación para precipitación anual, incluyendo IDW, Global Polynomial Interpolation (GPI), Kriging Universal y Kriging Ordinario, para la variable de precipitación anual. El estudio se basó en 21 estaciones de la red de Cenicafé, propiedad de la Federación Nacional de Cafeteros de Colombia. Los resultados mostraron que la metodología de IDW obtuvo mejores resultados para la variable evaluada.

Se seleccionó el IDW debido a que es el método más comúnmente empleado para estimar datos faltantes en estudios de hidrología. El NRM, aparte de ser ampliamente utilizado, fue desarrollado específicamente para zonas montañosas, como el área de estudio. Por su parte, el método SBE, además de ser uno de los más sencillos, se basa en la correlación con las estaciones vecinas, lo cual es relevante en regiones montañosas.

Las Cadenas de Markov, según algunos estudios, han demostrado un buen rendimiento en series con un alto porcentaje de datos faltantes, lo cual es una característica de las series de este estudio. Por último, el MLP, que es ampliamente utilizado en metodologías de inteligencia artificial, es un tipo de red neuronal con capacidad de mapeo de entrada-salida que es adecuado para la modelación de la precipitación, dado que las series de precipitación diaria no suelen estar correlacionadas.

### **5.3.2 Información utilizada en las metodologías de relleno**

Se cuantificó el desempeño las metodologías estadísticas, estocásticas y de inteligencia artificial que se presentan en el capítulo 4.1.1 utilizando 1) solamente la información de las estaciones en tierra del IDEAM, 2) solamente la información satelital y 3) utilizando la información de las estaciones junto con la de los productos satelitales, lo anterior para observar la diferencia obtenida en el relleno con las dos fuentes de información por separado (estaciones y satélites) y utilizando las dos en conjunto. Realizando la comparación anterior se puede identificar cuál de las dos fuentes de información es más apta para rellenar las series diarias de precipitación de las estaciones en tierra o si el uso de las dos fuentes combinadas es mejor que el uso de estas de manera individual.

Dentro de los productos satelitales de precipitación se usó solamente CHIRPS V2.0 (en el punto 2 y 3) por ser el producto satelital que detecta mejor los eventos diarios de precipitación, es el que mejor se ajusta a las estaciones en tierra cuando se observa la precipitación acumulada, y aunque fue el producto que peor desempeño tuvo en la estimación de las precipitaciones diarias, los tres productos evaluados obtuvieron un desempeño similar (bajo desempeño); mientras que CHIRPS V2.0 si destaca en la detección de los eventos diarios de lluvia. Adicionalmente en un estudio reciente, Valencia et al. (2023) evaluaron el desempeño de ocho productos satelitales de precipitación

---

(CHIRPSv2, CMORPHv1, ERA5, ERA5-Land, IMERG, MSWEPv2.8, PERSIANN-CCS, and PERSIANN-CDR) en los gradientes climáticos de Colombia y sus grandes pendientes; en donde concluyeron que dentro de los ocho productos evaluados, el mejor fue CHIRPSv2, tanto a nivel nacional como regional, y en todas las zonas de elevación y escalas temporales (3-días, mensual, estacional y anual).

En las metodologías de interpolación espacial la selección de las estaciones vecinas es decisivo para el buen desempeño de los resultados obtenidos en su aplicación (Eischeid et al., 1995). Diferentes autores recomiendan el uso de tres o cuatro estaciones vecinas para estimar datos faltantes de precipitación; otros aconsejan el uso de las estaciones vecinas que tengan un coeficiente de correlación mínimo con la estación objetivo (Suhaila et al., 2008). Por ejemplo, Eischeid et al. (1995) usaron un coeficiente de correlación mínimo de 0.35 de las estaciones vecinas en relación con la estación objetivo. Por otro lado, la inclusión de más de cuatro estaciones no mejora el desempeño de la estimación de datos faltantes y que por el contrario degrada el desempeño (Eischeid et al., 1995). Asimismo, el desempeño de las metodologías IDW y NRM no varía significativamente con la inclusión de diferentes estaciones variando el radio de búsqueda para la inclusión de estaciones (Suhaila et al., 2008). En este trabajo se usa una combinación de los dos enfoques mencionados; se usaron las cuatro estaciones vecinas con coeficientes de correlación mínimos de 0.35 (calculados a partir de mínimo un mes de concurrencia de datos entre las estaciones), cumpliendo así también el criterio de la geoestadística en el cual se plantea que entre más cercanas son las estaciones se dice que son más homogéneas. El enfoque mencionado se aplicó solamente para las metodologías que usan varias estaciones vecinas como insumo (IDW, NRM y MLP).

### **5.3.3 Periodos de calibración y validación**

Cuando se realiza cualquier estimación determinista es indispensable realizar validación de los resultados (OMM, 2018). Para este caso se utilizó la validación de muestra dividida (*Split-sample*) la cual consiste en utilizar dos subconjuntos tomados de los datos disponibles, uno para desarrollar los métodos y realizar las estimaciones que permitan el cálculo o estimación de parámetros y el otro para evaluar los datos estimados y su eficiencia o incertidumbre (OMM, 2018), el primer subconjunto es llamado periodo de calibración y el segundo, periodo de validación. El periodo de calibración se utilizó para

calcular los parámetros necesarios para desarrollar las metodologías NRM, SBE, Cadenas de Markov y MLP; mientras que el periodo de validación para evaluar el desempeño de todas las metodologías.

Se escogió como periodo de calibración desde 1994 a 2003 y como periodo de validación de 2004 a 2013; se eligieron teniendo en cuenta la variabilidad climática típica de la región, es decir, que comprendieran tanto periodos El Niño como periodos La Niña del ENOS. Para esto se revisó el Índice Niño Oceánico (ONI, por sus siglas en inglés), que es la principal medida para monitorear, evaluar y predecir el ENOS (Climate Prediction Center /NCEP, 2022), en donde periodos El Niño presentan valores del ONI mayor o igual a  $+0.5^{\circ}\text{C}$  y La Niña valores menor o igual a  $-0.5^{\circ}\text{C}$ . En la Figura 23 se observa que los dos periodos utilizados (resaltados con rojo) tienen periodos El Niño, La Niña y normal.

Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
1990	0.1	0.2	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.3	0.4	0.4
1991	0.4	0.3	0.2	0.3	0.5	0.6	0.7	0.6	0.6	0.8	1.2	1.5
1992	1.7	1.6	1.5	1.3	1.1	0.7	0.4	0.1	-0.1	-0.2	-0.3	-0.1
1993	0.1	0.3	0.5	0.7	0.7	0.6	0.3	0.3	0.2	0.1	0.0	0.1
1994	0.1	0.1	0.2	0.3	0.4	0.4	0.4	0.4	0.6	0.7	1.0	1.1
1995	1.0	0.7	0.5	0.3	0.1	0.0	-0.2	-0.5	-0.8	-1.0	-1.0	-1.0
1996	-0.9	-0.8	-0.6	-0.4	-0.3	-0.3	-0.3	-0.3	-0.4	-0.4	-0.4	-0.5
1997	-0.5	-0.4	-0.1	0.3	0.8	1.2	1.6	1.9	2.1	2.3	2.4	2.4
1998	2.2	1.9	1.4	1.0	0.5	-0.1	-0.8	-1.1	-1.3	-1.4	-1.5	-1.6
1999	-1.5	-1.3	-1.1	-1.0	-1.0	-1.0	-1.1	-1.1	-1.2	-1.3	-1.5	-1.7
Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
2000	-1.7	-1.4	-1.1	-0.8	-0.7	-0.6	-0.6	-0.5	-0.5	-0.6	-0.7	-0.7
2001	-0.7	-0.5	-0.4	-0.3	-0.3	-0.1	-0.1	-0.1	-0.2	-0.3	-0.3	-0.3
2002	-0.1	0.0	0.1	0.2	0.4	0.7	0.8	0.9	1.0	1.2	1.3	1.1
2003	0.9	0.6	0.4	0.0	-0.3	-0.2	0.1	0.2	0.3	0.3	0.4	0.4
2004	0.4	0.3	0.2	0.2	0.2	0.3	0.5	0.6	0.7	0.7	0.7	0.7
2005	0.6	0.6	0.4	0.4	0.3	0.1	-0.1	-0.1	-0.1	-0.3	-0.6	-0.8
2006	-0.9	-0.8	-0.6	-0.4	-0.1	0.0	0.1	0.3	0.5	0.8	0.9	0.9
2007	0.7	0.2	-0.1	-0.3	-0.4	-0.5	-0.6	-0.8	-1.1	-1.3	-1.5	-1.6
2008	-1.6	-1.5	-1.3	-1.0	-0.8	-0.6	-0.4	-0.2	-0.2	-0.4	-0.6	-0.7
2009	-0.8	-0.8	-0.6	-0.3	0.0	0.3	0.5	0.6	0.7	1.0	1.4	1.6
Year	DJF	JFM	FMA	MAM	AMJ	MJJ	JJA	JAS	ASO	SON	OND	NDJ
2010	1.5	1.2	0.8	0.4	-0.2	-0.7	-1.0	-1.3	-1.6	-1.6	-1.6	-1.6
2011	-1.4	-1.2	-0.9	-0.7	-0.6	-0.4	-0.5	-0.6	-0.8	-1.0	-1.1	-1.0
2012	-0.9	-0.7	-0.6	-0.5	-0.3	0.0	0.2	0.4	0.4	0.3	0.1	-0.2
2013	-0.4	-0.4	-0.3	-0.3	-0.4	-0.4	-0.4	-0.3	-0.3	-0.2	-0.2	-0.3

Figura 23. Índice ONI entre 1990 y 2013. Fuente: (NOAA-CPC, 2022)

### 5.3.4 Técnicas utilizadas para el relleno de datos faltantes de lluvia diaria

La selección de estaciones que se utilizaron para cuantificar el desempeño de las metodologías de relleno, se realizó de la siguiente manera:

- Las estaciones sin datos en los periodos de calibración (1994 a 2003) o validación (2004 a 2013) fueron descartadas.

- Se seleccionaron estaciones con por lo menos cuatro estaciones vecinas en los periodos de calibración y validación con coeficientes de correlación mínimos de 0.35; lo anterior con el fin de aplicar el criterio usado para rellenar con las metodologías IDW, NRM y MLP, ver 5.3.2.
- Se seleccionaron estaciones con por lo menos un mes o más datos seguidos sin faltantes en los periodos de calibración y validación.

Las 15 estaciones seleccionadas para aplicar las metodologías de relleno se presentan en la Tabla 11 y Figura 24, las cuales fueron seleccionadas siguiendo los tres parámetros expuestos anteriormente.

Tabla 11. Estaciones utilizadas aplicar las metodologías de relleno

<b>Código</b>	<b>Nombre</b>	<b>Municipio</b>	<b>Latitud</b>	<b>Longitud</b>	<b>Altitud (m)</b>
26155110	AEROPUERTO LA NUBIA	Manizales	5.030	-75.470	2104
26160120	CRISTALINA LA	Neira	5.192	-75.436	2411
26150160	LA ESPERANZA	Manizales	5.018	-75.357	3341
26155150	LAS BRISAS - AUT	Villamaría	4.935	-75.350	4133
23020080	MANZANARES	Manzanares	5.266	-75.144	1974
23020090	MARULANDA	Marulanda	5.278	-75.267	1700
26160090	PACORA PLAZA FERIA	Pácora	5.522	-75.455	712
26185010	PELADA LA	Aguadas	5.579	-75.345	2180
26170290	RIOSUCIO	Riosucio	5.425	-75.730	1946
26160080	SALAMINA CONC BARC	Salamina	5.395	-75.489	1890
23055040	SAMANA	Samaná	5.419	-74.999	1532
26185040	SAN FELIX	Pensilvania	5.373	-75.375	2690
23050230	SAN JOSE PENSILVAN	Pensilvania	5.374	-75.149	216
26155170	TESORITO FINCA - AUT	Samaná	5.032	-75.438	2325
26180250	VALLE ALTO - AUT	Manizales	5.350	-75.318	2800

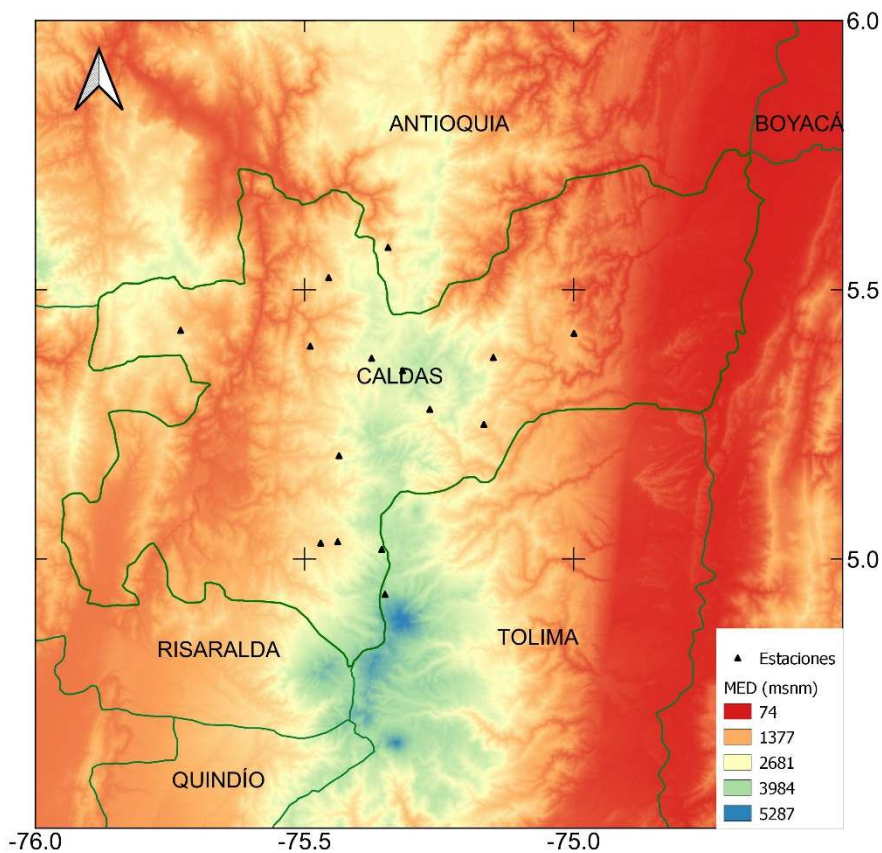


Figura 24. Estaciones utilizadas para aplicar las metodologías de relleno.

Las estaciones vecinas utilizadas para la estimación de cada dato faltante varían según las fechas, debido a los datos faltantes presentes en las series de las mismas.

Como se explica en el capítulo 5.3.2, el relleno de series se realizó de tres maneras 1) utilizando solamente la información de las estaciones en tierra del IDEAM, 2) solamente la información satelital y 3) utilizando la información de las estaciones junto con la de los productos satelitales. Para cada metodología la aplicación de cada uno de estos tres puntos se aplicó de manera diferente, por lo cual, a continuación, se explica de manera detallada la aplicación en cada metodología.

- **Distancia inversa ponderada (IDW)**

- Utilizando solamente la información de las estaciones en tierra del IDEAM:** se usaron las cuatro estaciones vecinas con coeficientes de correlación mínimos de 0.35 (calculados a partir de mínimo un mes de concurrencia de datos entre las estaciones).
- Utilizando solamente la información de CHIRPS V2.0:** se rellenaron las series utilizando los datos de CHIRPS V2.0 multiplicados por la pendiente de la regresión lineal calculada a partir de la curva de doble masa entre la estación y CHIRPS.
- Utilizando la información de las estaciones y la información de CHIRPS V2.0:** se rellenaron las series promediando la información de CHIRPS V2.0 (multiplicada por la pendiente de la regresión lineal calculada a partir de la curva de doble masa entre la estación y CHIRPS) con las series estimadas en el punto i (utilizando como insumo solamente las estaciones).

- **Método de la relación normal modificado, basado en la correlación (NRM)**

- Utilizando solamente la información de las estaciones en tierra del IDEAM:** al igual que en la metodología IDW se usaron las cuatro estaciones vecinas con coeficientes de correlación mínimos de 0.35 (calculados a partir de mínimo un mes de concurrencia de datos entre las estaciones).
- Utilizando solamente la información de CHIRPS V2.0:** al igual que en la metodología IDW se rellenaron las series utilizando los datos de CHIRPS V2.0 multiplicados por la pendiente de la regresión lineal calculada a partir de la curva de doble masa entre la estación y CHIRPS.
- Utilizando la información de las estaciones y la información de CHIRPS V2.0:** se rellenaron las series utilizando como insumo para la metodología, las cuatro estaciones vecinas con coeficientes de correlación mínimos de 0.35 (calculados a partir de mínimo un mes de concurrencia de datos entre las estaciones) y la serie de CHIRPS V2.0 con su respectivo índice de correlación (parámetro para ponderar).



- **Mejor estimador único (SBE)**

- Utilizando solamente la información de las estaciones en tierra del IDEAM:** se usó la estación con el mejor coeficiente de correlación (calculado a partir de mínimo un mes de concurrencia de datos entre las estaciones).
- Utilizando solamente la información de CHIRPS V2.0:** se rellenaron las series utilizando los datos de CHIRPS V2.0 multiplicados por la pendiente de la regresión lineal calculada a partir de la curva de doble masa entre la estación y CHIRPS.
- Utilizando la información de las estaciones y la información de CHIRPS V2.0:** se rellenaron las series promediando la información de CHIRPS V2.0 (multiplicada por la pendiente de la regresión lineal calculada a partir de la curva de doble masa entre la estación y CHIRPS) con las series estimadas en el punto i (utilizando como insumo solamente las estaciones).

- **Cadenas de Markov de primer orden con dos estados**

- Utilizando solamente la información de las estaciones en tierra del IDEAM:** se usa la información del periodo de calibración de la misma estación que se está rellenando para calcular los parámetros que se requieren en la metodología, ver 3.1.2.
- Utilizando solamente la información de CHIRPS V2.0:** se rellenaron las series utilizando los datos de CHIRPS V2.0 multiplicados por la pendiente de la regresión lineal calculada a partir de la curva de doble masa entre la estación y CHIRPS.
- Utilizando la información de las estaciones y la información de CHIRPS V2.0:** se rellenaron las series promediando la información de CHIRPS V2.0 (multiplicada por la pendiente de la regresión lineal calculada a partir de la curva de doble masa entre la estación y CHIRPS) con las series estimadas en el punto i (utilizando como insumo solamente las estaciones).

Con el propósito de evaluar el rendimiento del método para rellenar tanto un solo dato como varios faltantes consecutivos; se eliminó de manera iterativa un solo dato a lo largo de la serie de validación de la estación, así como cinco datos consecutivos de

manera iterativa. Lo anterior debido a que el resultado de las Cadenas de Markov de primer orden depende de la información del dato anterior. Se aplicó la metodologías para cinco datos faltantes consecutivos, ya que en (OMM, 2018) no se recomienda calcular un valor mensual de precipitación si faltan más de diez valores diarios o más de cinco consecutivos.

- **Redes de perceptrones multicapa (MLP)**

- iv. **Utilizando solamente la información de las estaciones en tierra del IDEAM:** para entrenar la red y estimar los datos faltantes se usaron las cuatro estaciones vecinas con coeficientes de correlación mínimos de 0.35 (calculados a partir de mínimo un mes de concurrencia de datos entre las estaciones). En este caso la estructura de la red neuronal fue de cuatro neuronas de entrada (igual al número de estaciones utilizadas como insumo), 20 neuronas ocultas (con base en las utilizadas en (Coulibaly & Evora, 2007), ya que ellos utilizaron varias arquitecturas obteniendo mejores resultados con 20 neuronas ocultas; en este estudio también se varió el número de neuronas ocultas y no se obtuvieron variaciones significativas) y una neurona de salida (la salida es la serie estimada).
- **Utilizando solamente la información de CHIRPS V2.0:** para el entrenamiento de la red, se usa como serie de entrada a la red, la información de CHIRPS V2.0 y como salida la información de la estación objetivo (series del periodo de calibración). Para estimar la información en el periodo de validación se usó como insumo las series de CHIRPS V2.0 en la ubicación correspondiente a cada estación.
- **Utilizando la información de las estaciones y la información de CHIRPS V2.0:** para el entrenamiento de la red, se usan como series de entrada a la red, la información de las cuatro estaciones vecinas con mayor correlación (mínimo de 0.35) y la serie de CHIRPS V2.0; como salida la información de la estación objetivo (series del periodo de calibración). En este caso la estructura de la red neuronal es de cinco neuronas de entrada (igual al número de estaciones utilizadas como

insumo más la serie de CHIRPS V2.0), 20 neuronas ocultas y una neurona de salida (la salida es la serie estimada).

## **5.4 Relleno de series**

Se evaluó el rendimiento de las diferentes metodologías utilizando como fuentes de información las estaciones, CHIRPS V2.0 y la combinación de ambos, en cuanto a la estimación de las magnitudes de precipitación y en la detección de los eventos de lluvia diarios.

Se realizaron curvas de doble masa entre las estaciones y CHIRPS V2.0 con el fin de analizar la relación entre las precipitaciones acumuladas de ambas fuentes de datos. Como se puede apreciar en la Figura 25, las curvas de doble masa presentan una relación lineal, lo que permite verificar la consistencia entre las dos fuentes de información, tal y como se presentó en 5.1.2. Se ajustaron modelos de regresión lineal a las curvas de doble masa de las 15 estaciones para calcular la relación entre las precipitaciones acumuladas de cada estación con respecto a CHIRPS V2.0. Con base en los anterior para las metodologías IDW, NRM, SBE y Cadenas de Markov se utilizó la información de CHIRPS multiplicada por la pendiente de la línea recta ajustada a la curva de doble masa, según se explica en

la sección 5.3.4, con el objetivo de estimar la precipitación puntual en la ubicación de las estaciones mediante la información espacial proporcionada por CHIRPS.

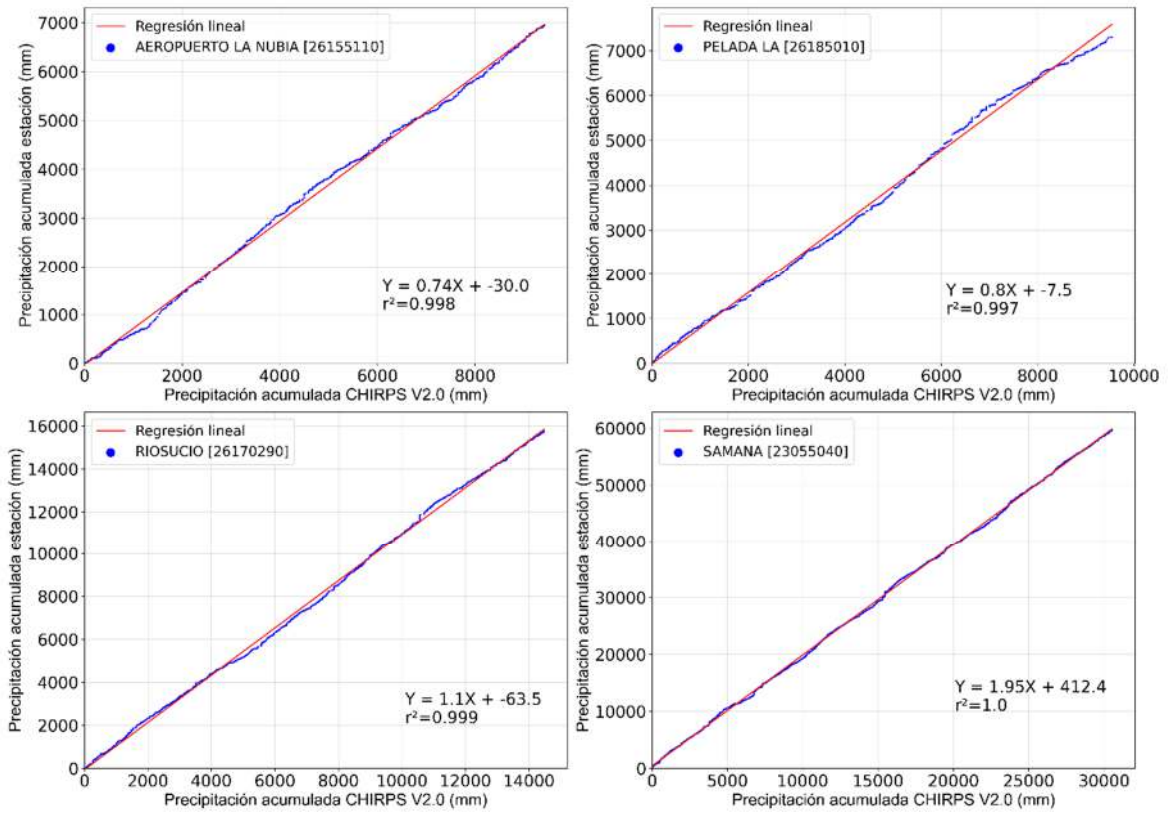


Figura 25. Curvas de doble masa de precipitación entre las estaciones y CHIRPS V2.0 entre los años 1982 y 1986

### 5.4.1 Metodologías estadísticas

La Tabla 12 muestra el desempeño promedio de las metodologías NRM, SBE e IDW con respecto a la magnitud de las precipitaciones diarias de las 15 estaciones que se presentan en la Figura 24.

Tabla 12. Desempeño promedio de las metodologías NRM, SBE e IDW con respecto a la magnitud de las precipitaciones diarias para el periodo 2004 a 2013

FUENTE	ME (mm)	RMAE	Bias	Nash-Sutcliffe
CHIRPS V2.0	-0.36	1.14	0.95	-0.39
NRM con estaciones	-0.56	0.75	1.02	0.30
NRM con estaciones y CHIRPS V2.0	-0.56	0.75	1.02	0.32
SBE con estaciones	-0.76	0.80	0.96	0.09
SBE con estaciones y CHIRPS V2.0	-0.56	0.84	0.95	0.20
IDW con estaciones	-0.75	0.73	0.98	0.28
IDW con estaciones y CHIRPS V2.0	-0.55	0.82	0.96	0.24

Se encontró que, para las tres metodologías estadísticas, el menor valor de error medio (ME) se registra cuando se utiliza solo CHIRPS V2.0 como fuente de información, lo cual es consistente con los resultados presentados en la Figura 25, donde se muestra la capacidad de CHIRPS V2.0 para estimar de forma precisa las precipitaciones acumuladas; lo anterior teniendo en cuenta que el ME mide el error promedio en la estimación de toda la serie evaluada. Sin embargo, cuando se utiliza CHIRPS V2.0 como única fuente de información en las tres metodologías, el RMAE presenta un valor mayor, lo que indica un peor desempeño en la estimación de las precipitaciones diarias. Este mismo patrón se evidenció en el coeficiente de Nash-Sutcliffe, donde el peor desempeño se presentó al utilizar únicamente CHIRPS V2.0. Además, se observa que el uso de CHIRPS V2.0 en conjunto con las estaciones en tierra, en algunos casos mejora el desempeño de las metodologías, pero no de manera significativa, y en otras ocasiones incluso empeora los resultados, lo que indica una falta de consistencia a la hora de utilizar esta combinación de fuentes de datos.

También se observa que según el coeficiente Bias, en promedio, cuando se utiliza solo CHIRPS V2.0, las precipitaciones acumuladas se subestiman, mientras que cuando se utilizan solo las estaciones y la combinación de estaciones con CHIRPS V2.0, en algunas ocasiones se subestiman y en otras se sobrestiman.

La metodología NRM presenta los mejores resultados en la estimación de la magnitud de las precipitaciones diarias en comparación con las metodologías IDW y SBE.

En la Tabla 13 se presenta el desempeño promedio de las metodologías NRM, SBE e IDW con relación a la detección de eventos diarios de precipitación de las 15 estaciones que se presentan en la Figura 24.

Tabla 13. Desempeño promedio de las metodologías NRM, SBE e IDW con respecto a la detección de eventos diarios de precipitación para el periodo 2004 a 2013

FUENTE	POD	FAR	ETS	HK	HSS	FB
CHIRPS V2.0	0.77	0.29	0.16	0.27	0.27	1.10
NRM con estaciones	0.94	0.24	0.34	0.47	0.50	1.26
NRM con estaciones y CHIRPS V2.0	0.97	0.32	0.18	0.27	0.30	1.44
SBE con estaciones	0.82	0.18	0.38	0.56	0.54	1.01
SBE con estaciones y CHIRPS V2.0	0.93	0.28	0.26	0.38	0.41	1.30
IDW con estaciones	0.93	0.23	0.36	0.50	0.52	1.23
IDW con estaciones y CHIRPS V2.0	0.96	0.29	0.24	0.35	0.39	1.37

Se puede observar que la proporción de eventos medidos por las estaciones que fueron estimados por las tres metodologías (POD) utilizando ambas fuentes de información y la combinación de las dos, es muy alta (valores mayores a 0.77 en todos los casos), lo que indica un buen desempeño en la detección de eventos de lluvia. Sin embargo, al ajustar el estadístico por aciertos relacionados con la probabilidad aleatoria (ETS), se observa que el valor es mucho más bajo (valores entre 0.16 y 0.38), lo que sugiere que en general las metodologías aciertan los días con lluvia, pero tienen dificultades para estimar los días sin lluvia y que la alta proporción de aciertos en los días de lluvia puede deberse a aciertos aleatorios. Lo anterior es consistente con los resultados del HK (proporción de días con lluvia y sin lluvia estimados correctamente), que muestra valores entre 0.27 y 0.56. Además, los valores de HSS reflejan una situación parecida, con valores similares a los de HK.

Los valores entre 1.01 y 1.44 de FB en todos los casos, indican la tendencia mencionada de las metodologías a sobrestimar la cantidad de días con lluvia en comparación con las observaciones de las estaciones.

La metodología estadística que presenta los mejores resultados para la detección de días con y sin precipitaciones es la SBE con el uso de estaciones en tierra, ya que en general

obtuvo los mejores ajustes, a excepción del POD, debido a que no sobrestima tantos días con lluvia de manera aleatoria (mejor valor de FAR, HK y HSS).

El desempeño en la detección de días con y sin precipitaciones cuando se utiliza CHIRPS V2.0 como única fuente de información para las metodologías estadísticas, así como cuando se combina con las estaciones, empeora con respecto a cuando se utilizan solamente las estaciones.

#### **5.4.2 Metodología estocástica**

Se evaluaron los ajustes a las funciones de distribución de probabilidad (FDP) mediante la valoración visual, la similitud en los valores extremos altos y los criterios de Akaike (AIC) y Bayesiano (BIC); los cuales se basan en el criterio de máxima verosimilitud. La Figura 26 muestra un ejemplo de los ajustes realizados para la estación AEROPUERTO LA NUBIA, donde se observa que el mejor ajuste lo presenta la función exponencial (coincidiendo con los criterios AIC y BIC), por lo que esta función se utilizó para estimar la magnitud de las precipitaciones en la metodología Cadenas de Markov de primer orden con dos estados. Los ajustes a estas FDP fueron realizados para las 15 estaciones utilizadas para el relleno, con el objetivo de elegir la FDP con mejor ajuste para cada una. Después de realizar los ajustes para las 15 estaciones, se utilizó la FDP exponencial para estimar la magnitud de las precipitaciones, ya que presentó el mejor ajuste para todas las estaciones según los cuatro criterios evaluados.

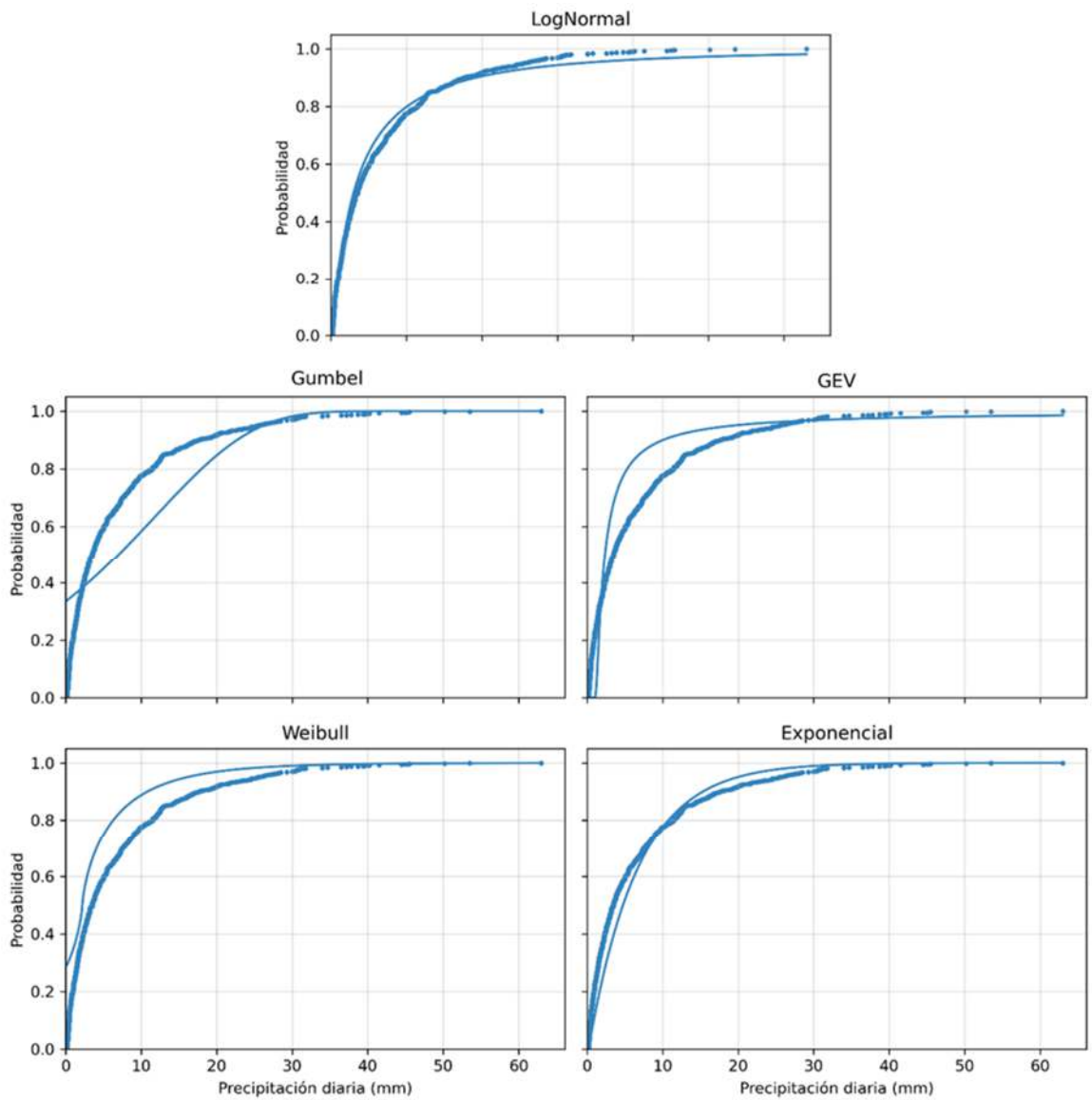


Figura 26. Ajustes de las funciones de distribución de probabilidad Gumbel, Generalized Extreme Value (GEV), Log Normal, Weibull y Exponencial para la estación AEROPUERTO LA NUBIA para el periodo 1994 a 2003

En la Tabla 14 se presentan los estadísticos utilizados para evaluar el rendimiento en la estimación de las magnitudes de las precipitaciones diarias mediante la metodología Cadenas de Markov de primer orden.



Tabla 14. Desempeño promedio en la estimación de la magnitud de las precipitaciones diarias utilizando la metodología Cadenas de Markov de primer orden para el periodo 2004 a 2013

FUENTE	ME (mm)	RMAE	Bias	Nash-Sutcliffe
CHIRPS V2.0	-0.33	1.14	0.95	-0.42
Markov con estaciones (relleno de un solo dato)	0.00	1.42	1.04	-0.84
Markov con estaciones (relleno de cinco datos consecutivos)	-0.07	1.37	1.02	-0.78
Markov con estaciones y CHIRPS V2.0 (relleno de un solo dato)	-1.50	1.19	0.82	-0.34
Markov con estaciones y CHIRPS V2.0 (relleno de cinco datos consecutivos)	-1.38	1.16	0.82	-0.31

Se logra un error medio (ME) promedio de cero al utilizar la metodología para rellenar un solo dato faltante, utilizando únicamente la información de la misma estación. Al emplear la misma metodología para rellenar cinco datos faltantes consecutivos a partir de la fuente de la estación, también se obtiene un error promedio muy bajo de -0.07 mm. CHIRPS V2.0 presenta un error mayor, aunque aun así se logran errores bajos (-0.33 mm). Sin embargo, al combinar CHIRPS V2.0 con las estaciones, se obtienen errores promedio mayores, siendo estos de -1.5 mm para un solo dato y de -1.38 mm para cinco datos consecutivos. Esto demuestra que las Cadenas de Markov de primer orden presentan un excelente desempeño al considerar precipitaciones acumuladas.

Los valores de RMAE para todos los casos son muy altos; ya que son mayores a 1, lo que refleja el bajo desempeño de las cadenas de Markov para estimar las precipitaciones diarias. Lo anterior también se observa en los valores del coeficiente de Nash-Sutcliffe, ya que para todos los casos los valores obtenidos son menores que 0, es decir, que el valor medio observado es mejor predictor que la aplicación de la metodología con cualquiera de los enfoques realizados.

En los casos que se usa CHIRPS V2.0 ya sea como única fuente de información o en combinación con las Cadenas de Markov, se subestiman los valores acumulados de las precipitaciones; valores de Bias menores a 1, mientras que cuando se usa como fuente de información las estaciones, se sobrestiman; valores mayores a 1.

En la Tabla 15 se presenta el desempeño promedio de la metodología Cadenas de Markov de primer orden con relación a la detección de eventos diarios de precipitación de las 15 estaciones que se presentan en la Figura 24.

Tabla 15. Desempeño promedio en la detección de eventos diarios de precipitación del relleno de series utilizando la Cadenas de Markov de primer orden para el periodo 2004 a 2013

FUENTE	POD	FAR	ETS	HK	HSS	FB
CHIRPS V2.0	0.77	0.29	0.16	0.27	0.27	1.10
Markov con estaciones (relleno de un solo dato)	0.64	0.41	-0.04	-0.08	-0.08	1.12
Markov con estaciones (relleno de cinco datos consecutivos)	0.67	0.38	0.00	0.00	0.00	1.11
Markov con estaciones y CHIRPS V2.0 (relleno de un solo dato)	0.64	0.41	-0.04	-0.08	-0.08	1.11
Markov con estaciones y CHIRPS V2.0 (relleno de cinco datos consecutivos)	0.67	0.38	0.00	0.00	0.00	1.11

El mejor caso de proporción de eventos medidos por la estación y que fueron estimados por las metodologías, se presenta cuando solamente se usa CHIRPS V2.0; de igual manera ocurre para las falsas alarmas (FAR), ya que en donde se tiene menor proporción de falsas alarmas es cuando se usa solamente CHIRPS V2.0. Al evaluar los aciertos de los eventos y no eventos medidos por las metodologías, CHIRPS V2.0 obtiene el mejor rendimiento con respecto a los otros casos evaluados (con estaciones y la combinación de estaciones con CHIRPS V2.0). Además, CHIRPS V2.0 sobrestima en menor proporción los eventos medidos por la estación con respecto a los otros enfoques.

A partir de lo mencionado anteriormente, se concluye que CHIRPS V2.0 presenta una mejor capacidad en la detección de eventos diarios en comparación con las Cadenas de Markov de primer orden utilizando únicamente las estaciones o combinándolas con CHIRPS V2.0. Sin embargo, al evaluar las magnitudes de las precipitaciones diarias, tanto las Cadenas de Markov (con estaciones y con la combinación con CHIRPS V2.0) como CHIRPS V2.0 tienen un rendimiento inferior. En contraste, al evaluar las precipitaciones

acumuladas, se observa un buen desempeño, especialmente cuando se utilizan únicamente las estaciones como fuente de información.

### 5.4.3 Metodología de inteligencia artificial

En la Tabla 16 se presentan los estadísticos utilizados para evaluar el rendimiento en la estimación de las magnitudes de las precipitaciones diarias mediante la metodología MLP.

Tabla 16. Desempeño promedio en la estimación de la magnitud de las precipitaciones diarias utilizando la metodología MLP para el periodo 2004 a 2013

FUENTE	ME (mm)	RMAE	Bias	Nash-Sutcliffe
MLP con CHIRPS V2.0	-0.67	1.03	0.91	0.05
MLP con estaciones	-0.13	0.71	0.98	0.42
MLP con estaciones y CHIRPS V2.0	-0.14	0.71	0.98	0.40

Se puede observar que el uso de MLP con las estaciones vecinas como único insumo para la metodología, presenta los mejores ajustes según los estadísticos evaluados. Por otro lado, el uso exclusivo de CHIRPS V2.0 como insumo para MLP, resulta ser el escenario con los peores ajustes según los estadísticos. La combinación de las estaciones vecinas y CHIRPS V2.0 en MLP no produce un cambio significativo en los resultados en comparación a cuando se usan solamente las estaciones vecinas.

En la Tabla 17 se presenta el desempeño promedio de la metodología MLP con relación a la detección de eventos diarios de precipitación de las 15 estaciones que se presentan en la Figura 24.

Tabla 17. Desempeño promedio en la detección de eventos diarios de precipitación del relleno de series utilizando la metodología MLP para el periodo 2004 a 2013

FUENTE	POD	FAR	ETS	HK	HSS	FB
MLP con CHIRPS V2.0	1.00	0.39	0.00	0.00	0.00	1.66
MLP con estaciones	1.00	0.39	0.01	0.01	0.02	1.65
MLP con estaciones y CHIRPS V2.0	0.99	0.37	0.04	0.06	0.07	1.60

Cuando se utiliza la metodología MLP combinando CHIRPS V2.0 y las estaciones vecinas, se observa una mejora en los resultados en comparación con los otros dos escenarios

(solo con CHIRPS V2.0 y solo con las estaciones vecinas). Sin embargo, esta mejora no es significativa en comparación con el uso exclusivo de las estaciones como insumo de MLP. Es importante destacar que la metodología MLP solo con CHIRPS V2.0 es la que presenta peores resultados en la detección de eventos diarios de precipitación, pero, aun así, muestra ajustes similares a los otros dos escenarios.

La metodología MLP en los tres escenarios propuestos muestra una excelente capacidad para detectar eventos de lluvia (POD) con valores entre 0.99 y 1.00. Sin embargo, existe una sobreestimación de la cantidad de días con lluvia, lo que se refleja en los bajos valores de HK (proporción de días con lluvia y sin lluvia estimados correctamente), que oscilan entre 0 y 0.06. Además, estos resultados son consistentes con los valores de ETS (entre 0 y 0.04) y HSS (entre 0 y 0.07), lo que indica una tendencia de la metodología a sobreestimar días con lluvia; tendencia que también se demuestra con los valores de FB mayores a 1 (entre 1.60 y 1.66).

#### **5.4.4 Resumen metodologías de relleno**

La Figura 27 presenta los errores del relleno relacionados con la magnitud de las precipitaciones diarias para todas las metodologías.

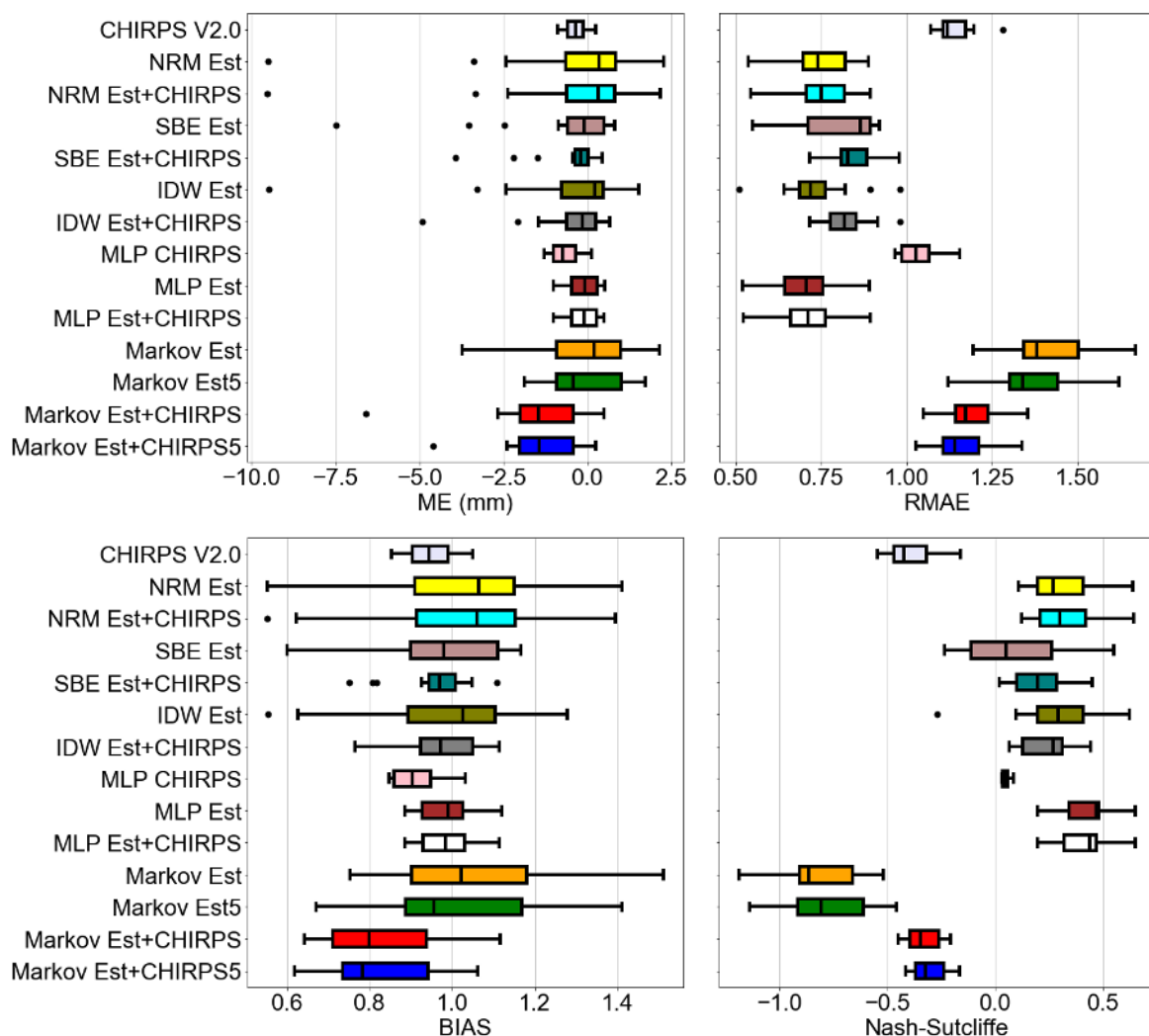


Figura 27. Desempeño en la estimación de la magnitud de las precipitaciones diarias para el periodo 2004 a 2013

Aunque en general, los errores promedio (ME) para las 15 estaciones y todas las metodologías son cercanos a cero, se destaca la consistencia de la metodología MLP con estaciones y con la combinación de estas con CHIRPS V2.0, ya que el rango de valores para las 15 estaciones evaluadas es cercano a cero. Lo cual refleja el buen comportamiento de la metodología MLP para precipitaciones acumuladas. También se destaca el comportamiento del producto CHIRPS V2.0. Es de resaltar que como se ha mencionado anteriormente, la combinación de CHIRPS V2.0 y las estaciones vecinas en

MLP no cambia significativamente los resultados, con respecto al uso de solamente las estaciones vecinas.

De igual manera el RMAE y el coeficiente de Nash-Sutcliffe, demuestran que el mejor comportamiento para rellenar precipitaciones diarias, lo presenta la metodología MLP con estaciones y con CHIRPS V2.0 combinado con estaciones, obteniendo los resultados más cercanos a cero para el RMAE y los valores más cercanos a uno para el coeficiente de Nash-Sutcliffe. Las metodologías con buenos valores de RMAE y Nash-Sutcliffe después de MLP son NRM con estaciones y con la combinación de CHIRPS V2.0 y estaciones e IDW con estaciones. MLP con estaciones y con la combinación de CHIRPS V2.0, también es la metodología que obtiene los mejores valores de Bias; cercanos a uno.

En la Figura 28 se presentan los errores del relleno con respecto la detección de eventos diarios de precipitación.

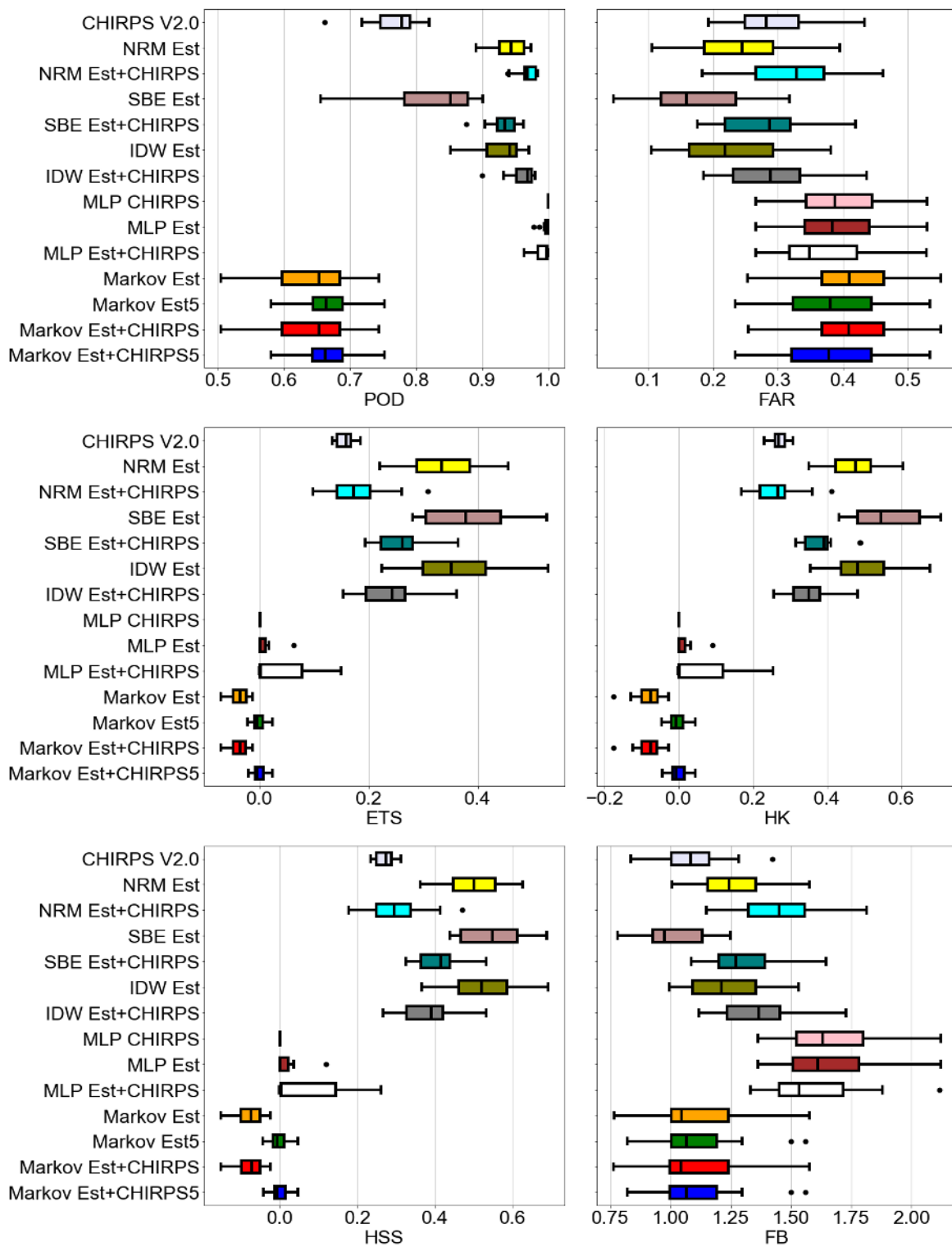


Figura 28. Desempeño en la detección de eventos diarios de precipitación del relleno de series para el periodo 2004 a 2013

MLP con las tres fuentes de información (solo estaciones, solo CHIRPS V2.0 y la combinación de ambas) es la metodología que presenta mejor proporción de eventos diarios medidos por la estación que fueron calculados por la metodología (POD). Sin embargo, la metodología que presenta menor proporción de falsas alarmas es SBE con estaciones.

Los mejores valores de ETS, HK y HSS los presenta la metodología SBE con estaciones, reflejando el buen comportamiento de esta metodología para estimar eventos diarios de precipitación y sin precipitación, siendo consistente con el resultado de FAR, ya que es la metodología que presenta menor proporción de falsas alarmas. También se destaca la metodología IDW con estaciones. De igual manera SBE con estaciones es la metodología que presenta la mejor proporción de eventos estimados por la metodología y medidos por la estación.

En términos generales, MLP con estaciones demuestra un rendimiento superior al estimar con mayor precisión la magnitud de las series diarias de precipitación. Por otro lado, la metodología SBE con estaciones logra una reproducción más precisa de los eventos diarios de precipitación.

En la Figura 29 se presenta un mapa comparativo del error medio (ME) para CHIRPS V2.0 (multiplicado por la pendiente de la regresión lineal calculada a partir de la curva de doble masa entre la estación y CHIRPS), la metodología MLP con estaciones, con CHIRPS V2.0 y con la combinación de ambas. Este mapa fue calculado a partir de la interpolación mediante IDW a partir del error puntual en la ubicación de las estaciones utilizadas con una resolución igual a la de CHIRPS V2.0; 0.05°.



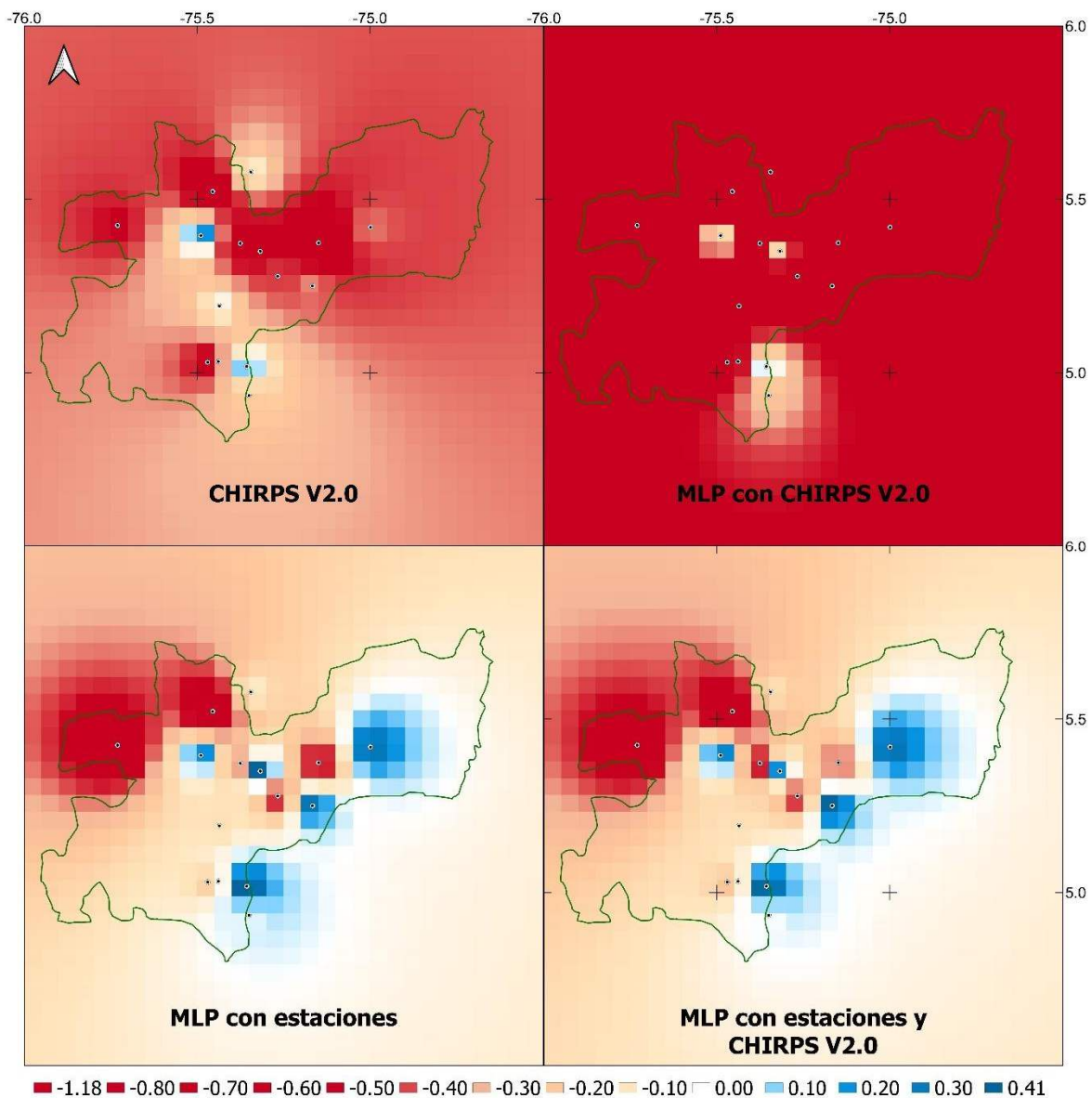


Figura 29. Mapa del ME (mm) para la metodología MLP con las diferentes fuentes de información, para el periodo 2004 a 2013. La resolución de cuadrícula es de  $0.05^\circ$

Al igual que la Figura 27, se puede notar que cuando CHIRPS V2.0 es la única fuente de información (tanto en CHIRPS V2.0 como en MLP con CHIRPS V2.0), se tiende a subestimar las precipitaciones acumuladas en la mayoría del área del departamento. Esta tendencia es más pronunciada en el caso de MLP con CHIRPS V2.0. Además, al considerar MLP con estaciones y MLP con estaciones y CHIRPS V2.0, no se observan

diferencias significativas. En estas situaciones, se subestiman las precipitaciones en algunas partes del departamento y se sobrestiman en otras, con una inclinación a sobrestimar en el oriente del departamento.

En la Figura 30 y Figura 31 se muestran los diagramas de caja y bigotes que representan los promedios y las desviaciones estándar de las series estimadas mediante las metodologías aplicadas, en comparación con CHIRPS V2.0 y las estaciones del IDEAM.

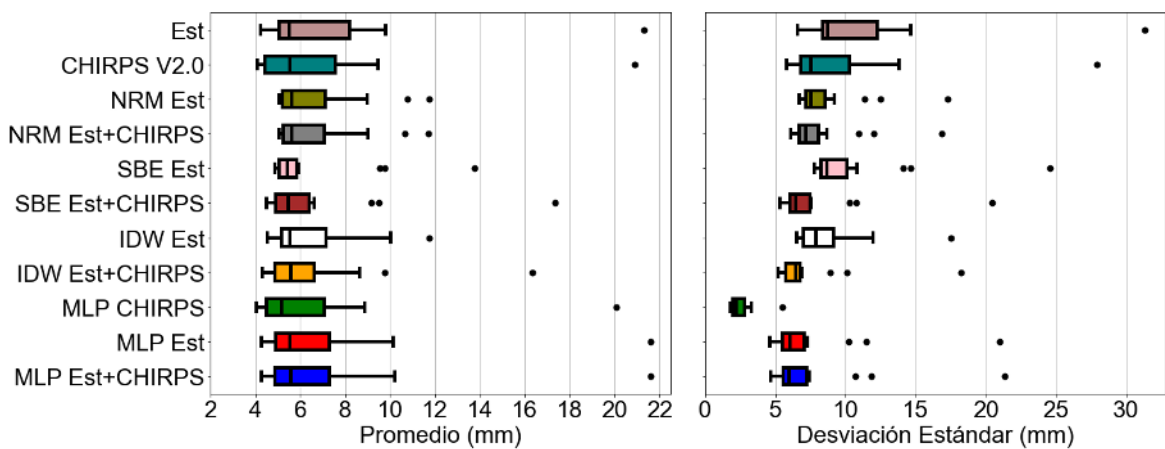


Figura 30. Diagrama de caja y bigotes de las series rellenadas con las metodologías estadísticas y MLP, para el periodo 2004 a 2013

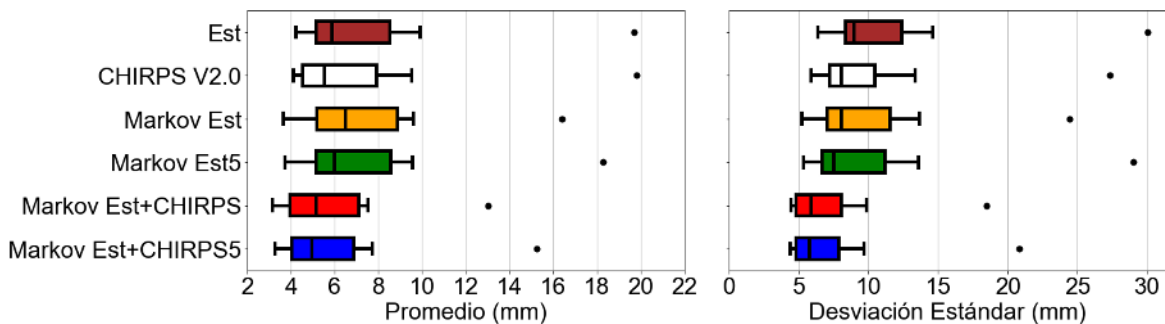


Figura 31. Diagrama de caja y bigotes de las series rellenadas con la metodología Cadenas de Markov de primer orden, para el periodo 2004 a 2013

En lo que respecta a los promedios, destaca la metodología MLP en los tres escenarios de fuentes de información, junto con CHIRPS V2.0. En cuanto a la desviación estándar, resalta CHIRPS V2.0 con los resultados más similares a las estaciones del IDEAM, en comparación con todas las metodologías utilizadas. Es importante señalar que, al igual que todas las comparaciones realizadas en el relleno de las series, los datos de CHIRPS V2.0 se han multiplicado por la pendiente de la regresión lineal calculada a partir de la curva de doble masa entre la estación correspondiente y CHIRPS V2.0.

Según lo expuesto anteriormente y considerando que la metodología que sobresale en los estadísticos descriptivos es MLP, se presentan en la Figura 32 y Figura 33 el mapa de promedios y el mapa de desviaciones estándar, respectivamente. Estos mapas corresponden a las series rellenas utilizando la metodología MLP, junto con las series originales de las estaciones y las series de CHIRPS V2.0.

Se puede observar que los promedios de las series rellenas son muy similares a los de CHIRPS V2.0 y a los de las estaciones en tierra. Como resultado, los mapas muestran una gran similitud en todos los casos. Además, en el diagrama de cajas y bigotes se aprecia una similitud marcada, con la misma cantidad de valores atípicos (específicamente, en la ubicación de la estación SAMANÁ [23055040] al oriente de Caldas).

En cuanto al diagrama de cajas y bigotes de las desviaciones estándar y al mapa de desviaciones estándar, se puede observar que la metodología MLP presenta desviaciones estándar significativamente inferiores a las presentadas por las estaciones y por CHIRPS V2.0. Estas últimas presentan desviaciones estándar similares entre sí. Es importante destacar que las desviaciones estándar son notablemente inferiores en el oriente de Caldas para la metodología MLP utilizando las tres fuentes de información.

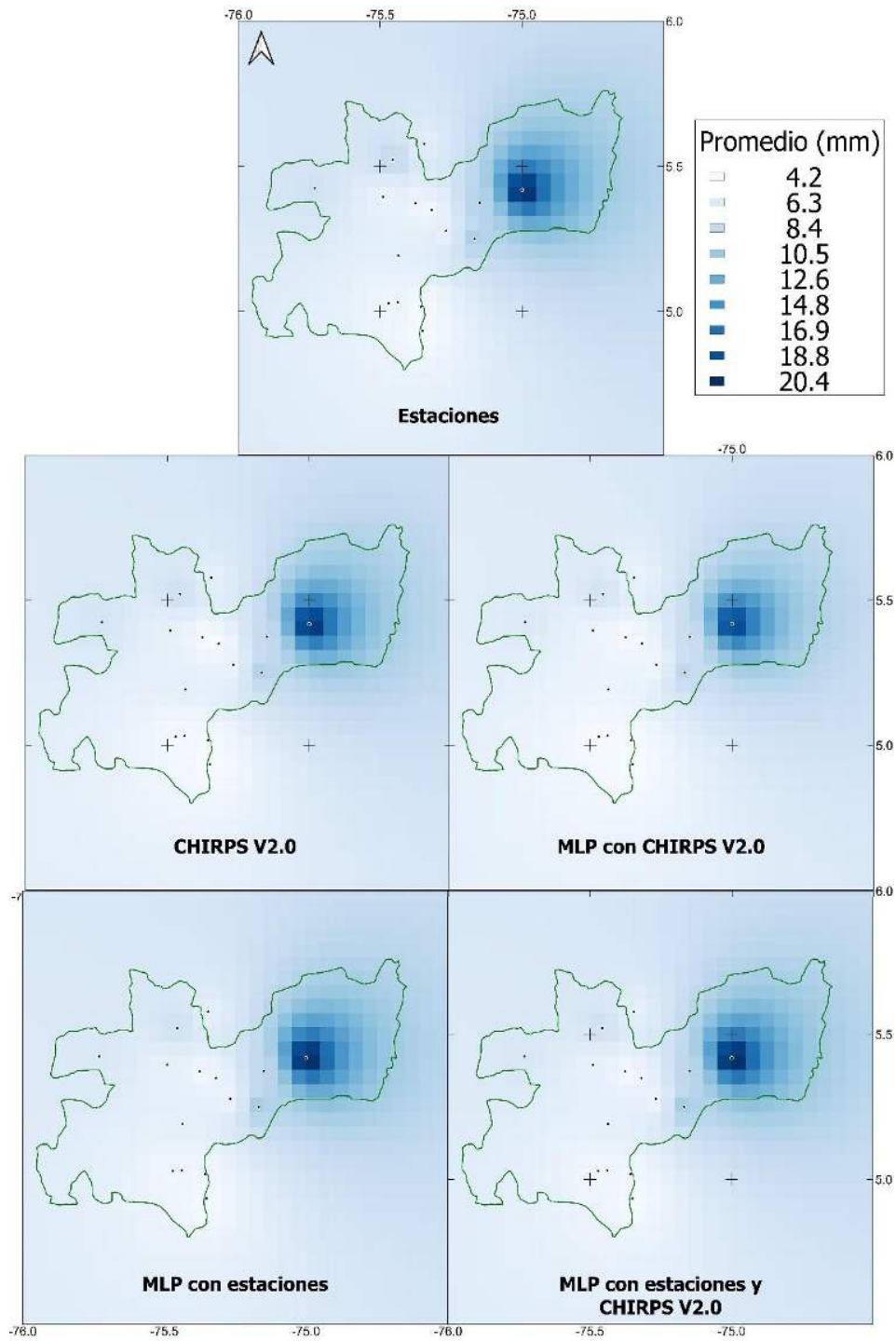


Figura 32. Mapa del promedio de las series rellenadas con la metodología MLP, para el periodo 2004 a 2013

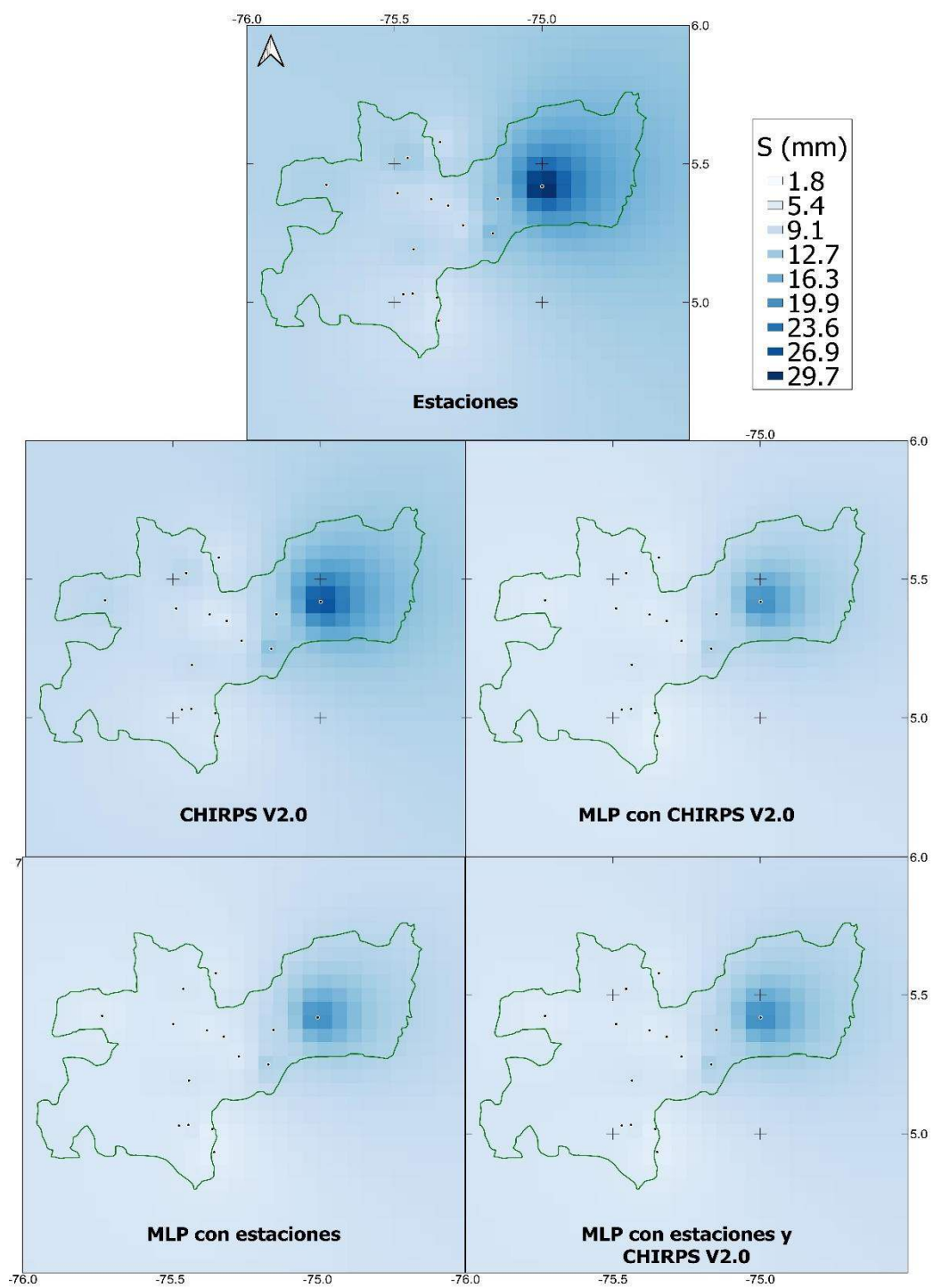


Figura 33. Mapa de la desviación estándar de las series rellenadas con la metodología MLP, para el periodo 2004 a 2013

## 5.5 Metodologías propuestas

Después de realizar un exhaustivo análisis y presentar los resultados previamente expuestos, se puede concluir que la metodología más efectiva para completar datos faltantes de lluvia diaria en el departamento de Caldas, Colombia, es MLP. Durante el análisis de los tres escenarios propuestos, donde se utilizaron únicamente las estaciones en tierra, solo las series de CHIRPS V2.0, y una combinación de las series de CHIRPS V2.0 y las estaciones en tierra, se pudo evidenciar claramente que el desempeño de la metodología con solo CHIRPS V2.0 es notablemente inferior en comparación con el uso exclusivo de las estaciones en tierra o la combinación de estas con CHIRPS V2.0. En consecuencia, se descarta la alternativa de utilizar MLP con CHIRPS V2.0 como la mejor metodología para el relleno de series de lluvia diaria.

Los resultados obtenidos mediante MLP utilizando únicamente las estaciones y MLP utilizando tanto las estaciones como CHIRPS V2.0 no presentan diferencias significativas. Por lo tanto, se considera que MLP con las estaciones en tierra es la mejor alternativa para el relleno de series diarias de lluvia, ya que la incorporación de CHIRPS V2.0 no produce modificaciones notables en los resultados y, en algunos casos, incluso obtiene peores resultados en los indicadores de desempeño evaluados en comparación con el uso exclusivo de MLP con las estaciones, como por ejemplo en el coeficiente de Nash-Sutcliffe.

## 6. Conclusiones

Resulta paradójico considerar el uso de datos de sensores remotos para complementar los registros de lluvia, dado que algunos productos satelitales de precipitación utilizan estaciones terrestres de precipitación para obtener estimaciones confiables de las precipitaciones. Sin embargo, la escala espacial en los sensores remotos abarca múltiples estaciones terrestres, lo que mitiga los errores causados por la alta variabilidad espacial de la lluvia. Por lo tanto, el empleo de sensores remotos como una variable adicional podría resultar beneficioso para mejorar el llenado de datos faltantes de precipitación, especialmente en zonas con una alta variabilidad espacial de la precipitación, como es el caso de los Andes tropicales.

Existen numerosos estudios en donde se comparan las precipitaciones estimadas por productos satelitales, pero a menudo no se evalúan ni aplican técnicas de relleno de datos para analizar el rendimiento de estas fuentes de información en dicha aplicación. Por lo tanto, en este estudio se realizó una evaluación de las técnicas de relleno de series diarias más comúnmente utilizadas, recurriendo a los productos satelitales de precipitación como fuente de información, además de las estaciones en tierra. Para lo cual se recopiló y analizó la información de las estaciones del IDEAM ubicadas en el departamento de Caldas, además de la información de los productos CHIRPS V2.0, MSWEP V1.0 y MSWEP V 2.8; utilizando finalmente como insumo para las metodologías de relleno la información de las estaciones del IDEAM y de CHIRPS V2.0. Al hacerlo, se pudo obtener una visión más completa y precisa de la calidad y fiabilidad de estos productos para su implementación en diversas aplicaciones relacionadas con la precipitación; en este caso para el relleno de series de lluvia diarias.

A partir de lo expuesto, se estudiaron y se llevaron a cabo evaluaciones de diversas metodologías de relleno (estadísticas, estocásticas y de inteligencia artificial) utilizando únicamente información de estaciones vecinas, únicamente CHIRPS V2.0, así como la combinación de ambas fuentes de información. El propósito de estas evaluaciones era determinar si la incorporación de la información proveniente de productos satelitales resultaba significativa para rellenar series de lluvia diaria, en situaciones donde la disponibilidad de datos de estaciones vecinas era limitada, o si la incorporación de los

productos satelitales mejoraba el desempeño del relleno de las series diarias en situaciones donde si hay información de estaciones vecinas.

Los resultados revelaron que el uso exclusivo de CHIRPS V2.0 arrojó de manera significativa peores resultados en comparación con el uso de estaciones terrestres. Además, se constató que la combinación de las estaciones y CHIRPS V2.0, dentro de las metodologías evaluadas, no generó cambios notorios en los resultados. En conclusión, a partir de este estudio se pudo constatar que, en situaciones donde la información proveniente de estaciones vecinas es escasa, la utilización exclusiva de CHIRPS V2.0 no ofrece buenos resultados en términos de relleno de series de lluvia diaria. Asimismo, la combinación de las estaciones y CHIRPS V2.0 no representa un cambio significativo. Por lo tanto, actualmente se considera que los productos satelitales de precipitación no son una opción adecuada para llevar a cabo el relleno de estas series de lluvia diarias en el departamento de Caldas, Colombia.

Por el contrario, se observó que en términos de los estadísticos de medición de precipitación acumulada, como el error medio (ME), el desempeño de las metodologías que emplean CHIRPS V2.0, particularmente aquellas basadas exclusivamente en CHIRPS V2.0, fue satisfactorio y consistente en las diferentes zonas del departamento de Caldas. Estos resultados indican que CHIRPS V2.0 es efectivo en la estimación de precipitación acumulada y presenta una consistencia notable en sus resultados. Además, se destaca como un indicador positivo del desempeño de CHIRPS V2.0 en la medición de precipitaciones acumuladas, su capacidad para seguir la distribución bimodal de las precipitaciones mensuales. Esta observación plantea una perspectiva interesante y sugiere la posibilidad de llevar a cabo un estudio similar, pero considerando series de precipitación con una escala temporal más amplia, como mensual o anual.

Entre las metodologías evaluadas, se destaca que la metodología MLP utilizando estaciones meteorológicas ha demostrado un mejor desempeño en la reproducción de la magnitud de las precipitaciones diarias. Sin embargo, es importante mencionar que no es la metodología que reproduce de manera más precisa los eventos diarios de precipitación. Se observó que la metodología SBE utilizando estaciones meteorológicas es la que logra una reproducción más fiel de los eventos diarios de precipitación (días con lluvia o sin



lluvia). Esto significa que la SBE con estaciones es capaz de capturar de manera más acertada los patrones y características específicas de los eventos de precipitación diarios.

En resumen, aunque la metodología MLP con estaciones sobresale en la reproducción de la magnitud de las precipitaciones diarias, la metodología SBE con estaciones se destaca por su capacidad para reproducir de manera más fehaciente los eventos diarios de precipitación. Ambas metodologías presentan fortalezas distintas en cuanto a la precisión en diferentes aspectos de las series de precipitación diaria.

En caso de llevar a cabo un análisis de cambio climático, es probable que los factores de calibración de las metodologías de interpolación sufran modificaciones, lo cual podría tener un impacto en el desempeño evaluado con los parámetros actuales. Es importante destacar que las metodologías evaluadas en este estudio no están siendo utilizadas para predecir escenarios futuros, sino para completar datos pasados. Por lo tanto, no se realizó un análisis específico sobre el cambio climático en el contexto de esta investigación.

La diferencia horaria entre las mediciones diarias de los productos satelitales (realizadas desde las 00:00 hasta las 24:00 en hora UTC) y las mediciones de las estaciones del IDEAM (realizadas desde las 7:00 de un día hasta las 7:00 del día siguiente, en hora colombiana) impacta los resultados obtenidos. Esto se debe a que la corrección aplicada se basa en la suposición de una precipitación constante a lo largo de cada día, lo cual no refleja la realidad. Es importante reconocer que esta disparidad horaria al realizar evaluaciones en escalas de tiempo acumuladas, es menos significativa.

Es importante destacar que el desempeño de los productos satelitales de precipitación puede variar dependiendo de la ubicación geográfica en la cual sean evaluados. Por lo tanto, es fundamental aclarar que los resultados obtenidos en este estudio son confiables y aplicables específicamente en la zona específica que fue evaluada. En caso de que se desee extrapolar estos resultados a otras áreas, es necesario tener especial cuidado y considerar las particularidades y características propias de cada región.

## 7.Recomendaciones y trabajos futuros

Los análisis realizados son útiles para estudios relacionados con el balance hidrológico, la planificación y la gestión de recursos hidráulicos. Sin embargo, es importante tener en cuenta que no son aplicables a eventos extremos. Esto se debe a que los productos satelitales de precipitación tienden a subestimar los valores de lluvias máximas, ya que representan celdas de gran tamaño en contraste con los datos puntuales de las estaciones de tierra.

Se recomienda tener en cuenta que factores como la topografía, la cobertura de nubes, la densidad de estaciones meteorológicas y otros elementos geográficos y climáticos pueden influir en el rendimiento de los productos satelitales de precipitación. Por lo tanto, es fundamental realizar evaluaciones adicionales y considerar las condiciones específicas de cada lugar antes de generalizar los resultados obtenidos en este estudio.

Dado que el monitoreo de la lluvia se ve afectado por muchas contingencias, es importante seguir explorando nuevas técnicas y estrategias que permitan el uso de la información proveniente de sensores remotos para el apoyo en el relleno de datos de lluvia diaria, dado el potencial de esta nueva información disponible de forma gratuita en internet.

La definición del umbral de lluvia sugerida en esta investigación es de 0.2 mm, pero no explora esta variabilidad, por lo que es interesante observar si los indicadores de desempeño para las diferentes técnicas se ven afectados de forma significativa para valores diferentes de ese umbral, teniendo en cuenta que la OMM sugiere un valor de 1 mm.

Puede ser importante estudiar para la zona de estudio nuevas variables provenientes de sensores remotos que permitan la mejora en los indicadores de relleno de datos de lluvia diaria.

Ampliar el análisis a escalas temporales más extensas (mensuales o anuales) permitiría evaluar de manera más completa el desempeño de las metodologías y los productos satelitales en términos de su aplicabilidad en el relleno de las series de precipitación

mensuales o anuales. Dado patrones de precipitación pueden variar a lo largo de períodos más largos, se podrían obtener perspectivas valiosas sobre la aplicabilidad y fiabilidad de las fuentes de información satelital en diferentes escalas temporales.

Este trabajo es importante ya que, generalmente en los estudios se comparan las estimaciones de precipitaciones proporcionadas por productos satelitales con respecto a las estaciones en tierra. Sin embargo, estos estudios a menudo no consideran ni aplican técnicas de relleno de datos para evaluar el rendimiento de estas fuentes de información en dicha aplicación. Por lo tanto, en este trabajo, se llevó a cabo una evaluación de las técnicas más utilizadas para rellenar series de datos diarios utilizando como fuente de información productos satelitales de precipitación, así como estaciones meteorológicas en tierra, específicamente en el departamento de Caldas. Este enfoque permitió obtener una visión más completa y precisa de la calidad y confiabilidad de estos productos para su aplicación en diversas áreas relacionadas con la precipitación. Este estudio, particularmente, se centró en el relleno de series de datos diarios de lluvia.

## 8. Bibliografía

Aghakouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research Atmospheres*, 116(2).

<https://doi.org/10.1029/2010JD014741>

American Society of Civil Engineers., T. C. on H. H. (1996). *Hydrology handbook*. ASCE.

Amjad, M., Yilmaz, M. T., Yucel, I., & Yilmaz, K. K. (2020). Performance evaluation of satellite- and model-based precipitation products over varying climate and complex topography. *Journal of Hydrology*, 584(February), 124707.

<https://doi.org/10.1016/j.jhydrol.2020.124707>

Arkin, P. A. (1990). State-of-the-art precipitation estimates from satellites. *ECMWF/WCRP*

- Workshop on Clouds, Radiative Transfer and the Hydrological Cycle, 12-15 November 1990*, 87–106. <https://www.ecmwf.int/node/7801>
- Arkin, P. A., & Meisner, B. N. (1987). The relationship between large-scale convective rainfall and cold cloud over the Western Hemisphere during 1982–84. *Mon. Wea. Rev.*, *115*, 51–74. [https://doi.org/10.1175/1520-0493\(1987\)115<0051:TRBLSC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<0051:TRBLSC>2.0.CO;2)
- Armanuos, A. M., Al-Ansari, N., & Yaseen, Z. M. (2020). Cross assessment of twenty-one different methods for missing precipitation data estimation. *Atmosphere*, *11*(4), 1–35. <https://doi.org/10.3390/ATMOS11040389>
- ASF. (n.d.). *ALOS PALSAR Digital Elevation Model*. <https://search.asf.alaska.edu/#/?dataset=ALOS>
- Bai, L., Shi, C., Li, L., Yang, Y., & Wu, J. (2018). Accuracy of CHIRPS satellite-rainfall products over mainland China. *Remote Sensing*, *10*(3). <https://doi.org/10.3390/rs10030362>
- Basri, M., Shaharudin, S., Kismiantini, Tan, M. L., Mohd Najib, S., Zainuddin, N., & Andayani, S. (2021). Regionalization of Rainfall Regimes Using Hybrid RF-Bs Couple with Multivariate Approaches. *ISPRS International Journal of Geo-Information*, *10*, 689. <https://doi.org/10.3390/ijgi10100689>
- Beck, H E, Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A. I. J. M., Weedon, G. P., Brocca, L., Pappenberger, F., Huffman, G. J., & Wood, E. F. (2017). Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrology and Earth System Sciences*, *21*(12), 6201–6217. <https://doi.org/10.5194/hess-21-6201-2017>
- Beck, Hylke E., Van Dijk, A. I. J. M., Levizzani, V., Schellekens, J., Miralles, D. G., Martens, B., & De Roo, A. (2017). MSWEP: 3-hourly 0.25° global gridded precipitation (1979-2015) by merging gauge, satellite, and reanalysis data. *Hydrology and Earth System Sciences*, *21*(1), 589–615. <https://doi.org/10.5194/hess-21-589-2017>
- Beck, Hylke E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Van Dijk, A. I. J. M., McVicar, T. R., & Adler, R. F. (2019). MSWep v2 Global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bulletin of the American Meteorological*

- Society*, 100(3), 473–500. <https://doi.org/10.1175/BAMS-D-17-0138.1>
- Campozano, L., Sánchez, E., Avilés, Á., & Samaniego, E. (2014). Evaluation of infilling methods for time series of daily precipitation and temperature: The case of the Ecuadorian Andes. *Maskana*, 5(1), 99–115. <https://doi.org/10.18537/mskn.05.01.07>
- Chahine, M. T. (1992). The hydrological cycle and its influence on climate. *Nature*.
- Chang, C.-L., Lo, S.-L., & Yu, S.-L. (2006). The Parameter Optimization in the Inverse Distance Method by GeneticAlgorithm for Estimating Precipitation. *Environmental Monitoring and Assessment*, 117(1), 145–155. <https://doi.org/10.1007/s10661-006-8498-0>
- Chen, F.-W., & Liu, C.-W. (2012). Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. *Paddy and Water Environment*, 10(3), 209–222. <https://doi.org/10.1007/s10333-012-0319-1>
- Chen, F., Liu, Y., Liu, Q., & Li, X. (2014). Spatial downscaling of TRMM 3B43 precipitation considering spatial heterogeneity. *International Journal of Remote Sensing*, 35(9), 3074–3093. <https://doi.org/10.1080/01431161.2014.902550>
- CHIRPS: Rainfall Estimates from Rain Gauge and Satellite Observations | Climate Hazards Center - UC Santa Barbara. (n.d.). Retrieved July 30, 2022, from <https://www.chc.ucsb.edu/data/chirps>
- Cifuentes, A. (2016). *EVALUACIÓN DE DIFERENTES MÉTODOS DE INTERPOLACIÓN PARA LA VARIABLE PRECIPITACIÓN EN EL DEPA*. UNIVERSIDAD DE MANIZALES.
- Climate Prediction Center /NCEP. (2022). *ENSO Cycle : Recent Evolution, Current Status and Predictions*. NOAA. [https://www.cpc.ncep.noaa.gov/products/analysis\\_monitoring/lanina/enso\\_evolution-status-fcsts-web.pdf](https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/lanina/enso_evolution-status-fcsts-web.pdf)
- CORPOCALDAS. (2001). *No Title*. <https://2.bp.blogspot.com/-duHTHLOHzTw/WQ26iTG3MUI/AAAAAAAAAWIU/EKi9z-VND9MHtR2FT2avlh8gYcciYpGlgCLcB/s1600/CALDAS%2B--caldas-precipit.gif>
- Coulibaly, P., & Evora, N. D. (2007). Comparison of neural network methods for infilling missing daily weather records. *Journal of Hydrology*, 341(1–2), 27–41. <https://doi.org/10.1016/j.jhydrol.2007.04.020>
- Cruz-Roa, A. F., & Barrios, M. I. (2018). Estimación de datos faltantes de lluvia mensual a través de la asimilación de información satelital y pluviométrica en una cuenca

- andina tropical. *Idesia (Arica), ahead*, 0–0. <https://doi.org/10.4067/s0718-34292018005001601>
- di Piazza, A., Conti, F. Lo, Noto, L. V., Viola, F., & La Loggia, G. (2011). Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. *International Journal of Applied Earth Observation and Geoinformation*, 13(3), 396–408. <https://doi.org/10.1016/j.jag.2011.01.005>
- Ebert, E. E., Janowiak, J. E., & Kidd, C. (2007). Comparison of Near-Real-Time Precipitation Estimates from Satellite Observations and Numerical Models. *Bulletin of the American Meteorological Society*, 88(1), 47–64. <https://doi.org/10.1175/BAMS-88-1-47>
- Eischeid, J. K., Baker, C. B., Karl, T. R., & Diaz, H. F. (1995). The Quality Control of Long-Term Climatological Data Using Objective Data Analysis. *Journal of Applied Meteorology and Climatology*, 34(12), 2787–2795. [https://doi.org/10.1175/1520-0450\(1995\)034<2787:TQCOLT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1995)034<2787:TQCOLT>2.0.CO;2)
- Eischeid, J. K., Pasteris, P. A., Diaz, H. F., Plantico, M. S., & Lott, N. J. (2000). Creating a Serially Complete, National Daily Time Series of Temperature and Precipitation for the Western United States. *JORNAL OF APPLIED METEOROLOGY*, 1580–1591.
- Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive Logistic Regression: A Statistical View of Boosting. *The Annals of Statistics*, 28, 337–407. <https://doi.org/10.1214/aos/1016218223>
- Funk, C., Verdin, A., Michaelsen, J., Peterson, P., Pedreros, D., & Husak, G. (2015). A global satellite-assisted precipitation climatology. *Earth System Science Data*, 7(2), 275–287. <https://doi.org/10.5194/essd-7-275-2015>
- Funk, Chris, Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data*, 2(1), 150066. <https://doi.org/10.1038/sdata.2015.66>
- Gobierno de Caldas - Geografía de Caldas*. (n.d.). Retrieved March 28, 2022, from <https://caldas.gov.co/index.php/portfolio-2/informacion-general/geografia-de-caldas>
- Guevara Ochoa, C., Briceño, N., Zimmermann, E., Vives, L., Blanco, M., Cazenave, &

- Ares, G. (2017). RELLENO DE SERIES DE PRECIPITACIÓN DIARIA PARA LARGOS PERIODOS DE TIEMPO EN ZONAS DE LLANURA. CASO DE ESTUDIO CUENCA SUPERIOR DEL ARROYO DEL AZUL. *Geoacta*, 42(1), 38–62.  
[http://www.scielo.org.ar/scielo.php?script=sci\\_arttext&pid=S1852-77442017000100004&lng=es&nrm=iso&tlng=es](http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S1852-77442017000100004&lng=es&nrm=iso&tlng=es)
- Hasana, M., & Crokea, B. F. W. (2013). *Filling gaps in daily rainfall data : a statistical approach*.
- Huffman, G. J., Adler, R. F., Arkin, P., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., McNab, A., Rudolf, B., & Schneider, U. (1997). The Global Precipitation Climatology Project (GPCP) Combined Precipitation Dataset. *Bulletin of the American Meteorological Society*, 78(1), 5–20. [https://doi.org/10.1175/1520-0477\(1997\)078<0005:TGPCPG>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<0005:TGPCPG>2.0.CO;2)
- Hurtado Montoya, A. F., & Mesa Sánchez, Ó. J. (2014). Reanalysis of monthly precipitation fields in Colombian territory. *Dyna*, 81(186), 251.  
<https://doi.org/10.15446/dyna.v81n186.40419>
- IDEAM. (2013). *Clasificación de los climas*. <http://atlas.ideam.gov.co/basefiles/climatext.pdf>
- IDEAM. (2014). *Cobertura de la Tierra Metodología CORINE Land Cover adaptada para Colombia durante el periodo 2010-2012*. Bogotá. <http://www.siac.gov.co/catalogo-de-mapas>
- IDEAM. (2019). *FICHA METODOLÓGICA OPERACIÓN ESTADÍSTICA VARIABLES METEOROLÓGICAS* Subdirección de Meteorología.
- IDEAM. (2021a). *Manual para el control de calidad de los datos meteorológicos en la preverificación, captura y verificación*.  
<http://sgi.ideam.gov.co/documents/412030/561097/M-GDI-M-M005+Manual+control+de+calidad+verificacion+V1>
- IDEAM. (2021b). *REDISEÑO DE LOS FORMATOS DE REGISTRO DE LAS VARIABLES METEOROLÓGICAS*.
- IDEAM - Instituto de Hidrología Meteorología y Estudios. (2019). *REGLAS DE VALIDACIÓN Y CONSISTENCIA DE LAS ESTADÍSTICAS DE MONITOREO Y SEGUIMIENTO DE LA CALIDAD DEL AIRE (EMSCA)*.  
<http://www.ideam.gov.co/documents/24189/102263008/M-GCI-EA-M019+MANUAL+DE+CRÍTICA+EMSCA.pdf/1406128d-4dd8-4845-a8ab->

1273bc3e0513?version=1.0

- IGAC. (1998). *Mapa Digital de Suelos del Departamento de Caldas, República de Colombia. Escala 1:100.000. Año 1998.*
- Jaramillo, Á., & Chaves, B. (2000). Distribución de la precipitación en Colombia analizada mediante conglomeración estadística. *Cenicafé*, 51(2), 102–113.
- Kajornrit, J., Wai Wong, K., & Che Fung, C. (2011). *Estimation of missing rainfall data in northeast region of Thailand using kriging methods: A comparison study. In: International Workshop on Bio-inspired Computing for Intelligent, Environments and Logistic Systems Estimation of Missing Rainfall Data in N.*  
<http://researchrepository.murdoch.edu.au/11390/>
- Kalteh, A. M., & Hjorth, P. (2009). Imputation of missing values in a precipitation–runoff process database. *Hydrology Research*, 40(4), 420–432.  
<https://doi.org/10.2166/nh.2009.001>
- Kanda, N., Negi, H. S., Rishi, M. S., & Shekhar, M. S. (2018). Performance of various techniques in estimating missing climatological data over snowbound mountainous areas of Karakoram Himalaya. *Meteorological Applications*, 25(3), 337–349.  
<https://doi.org/10.1002/met.1699>
- Kidd, C., & Levizzani, V. (2011). Status of satellite precipitation retrievals. *Hydrology and Earth System Sciences*, 15(4), 1109–1116. <https://doi.org/10.5194/hess-15-1109-2011>
- Kim, J., & Ryu, J. H. (2016). A heuristic gap filling method for daily precipitation series. *Water Resources Management*, 30(7), 2275–2294. <https://doi.org/10.1007/s11269-016-1284-z>
- Kim, J. W., & Pachepsky, Y. A. (2010). Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. *Journal of Hydrology*, 394(3–4), 305–314.  
<https://doi.org/10.1016/j.jhydrol.2010.09.005>
- Lei, H., Zhao, H., & Ao, T. (2022). Ground validation and error decomposition for six state-of-the-art satellite precipitation products over mainland China. *Atmospheric Research*, 269(July 2021), 106017. <https://doi.org/10.1016/j.atmosres.2022.106017>
- Longman, R., Newman, A., Giambelluca, T., & Lucas, M. (2020). Characterizing the



- Uncertainty and Assessing the Value of Gap-Filled Daily Rainfall Data in Hawaii. *Journal of Applied Meteorology and Climatology*, 59, 1261–1276.  
<https://doi.org/10.1175/JAMC-D-20-0007.1>
- Luis Morales, J., Horta -Rangel, F. A., Segovia-Domínguez, I., Robles Morua, A., Hernández, J. H., Morales Martínez, J. L., Horta-Rangel, F. A., Segovia-Domínguez, I., Robles Morua, A., & Hernández, J. H. (2019). Analysis of a new spatial interpolation weighting method to estimate missing data applied to rainfall records. *Atmósfera*, 32(3), 237–259. <https://doi.org/10.20937/ATM.2019.32.03.06>
- Mehan, S., Guo, T., Gitau, M. W., & Flanagan, D. C. (2017). Comparative Study of Different Stochastic Weather Generators for Long-Term Climate Data Simulation. *Climate*, 5(2). <https://doi.org/10.3390/cli5020026>
- Miró, J. J., Caselles, V., & Estrela, M. J. (2017). Multiple imputation of rainfall missing data in the Iberian Mediterranean context. *Atmospheric Research*, 197, 313–330. <https://doi.org/https://doi.org/10.1016/j.atmosres.2017.07.016>
- Moeletsi, M., Shabalala, Z., Nysschen, G., & Walker, S. (2016). Evaluation of an inverse distance weighting method for patching daily and dekadal rainfall over the Free State Province, South Africa. *Water S.A*, 42. <https://doi.org/10.4314/wsa.v42i3.12>
- Mora, D., Wyseure, G., & Willems, P. (2014). *GAP FILLING BASED ON A QUANTILE PERTURBATION FACTOR TECHNIQUE*. <https://doi.org/10.13140/2.1.3446.4006>
- NOAA-CPC. (2022). *ENSO Cold & Warm Episodes by Season*.  
[https://origin.cpc.ncep.noaa.gov/products/analysis\\_monitoring/ensostuff/ONI\\_v5.php](https://origin.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php)
- Okamoto, K. i., Ushio, T., Iguchi, T., Takahashi, N., & Iwanami, K. (2005). The global satellite mapping of precipitation (GSMaP) project. *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05.*, 5, 3414–3416. <https://doi.org/10.1109/IGARSS.2005.1526575>
- OMM. (2018). Guía de prácticas climatológicas. In *Organización Meteorológica Mundial* (Vol. 100).  
[http://www.wmo.int/pages/prog/wcp/ccl/guide/documents/wmo\\_100\\_es.pdf](http://www.wmo.int/pages/prog/wcp/ccl/guide/documents/wmo_100_es.pdf)
- Organización Meteorológica Mundial. (2011). Guía de prácticas climatológicas. In *Organización Meteorológica Mundial* (Vol. 100).  
[http://www.wmo.int/pages/prog/wcp/ccl/guide/documents/wmo\\_100\\_es.pdf](http://www.wmo.int/pages/prog/wcp/ccl/guide/documents/wmo_100_es.pdf)
- Organización Meteorológica Mundial. (2018). *Guía de prácticas climatológicas*.
- Panofsky, H. A., & Brier, G. W. (1958). *Some Applications of Statistics to Meteorology*.

- Mineral Industries Extension Services, College of Mineral Industries, Pennsylvania State University. <https://books.google.com.co/books?id=UyRRAAAAMAAJ>
- Paulhus, J. L. H., & Kohler, M. A. (1952). Interpolation of Missing Precipitation Records. *Monthly Weather Review*, 80(8), 129–133. [https://doi.org/10.1175/1520-0493\(1952\)080<0129:iompr>2.0.co;2](https://doi.org/10.1175/1520-0493(1952)080<0129:iompr>2.0.co;2)
- Paulhus, Joseph L H, & Kohler, M. A. (1952). INTERPOLATION OF MISSING PRECIPITATION RECORDS. *Monthly Weather Review*, 80, 129–133.
- Poveda, G. (2004). La Hidroclimatología de Colombia: Una síntesis desde la escala interdecadal hasta la escala diurna. *Revista Academia Colombiana de Ciencias*, 28(10), 201–222. [https://www.researchgate.net/publication/284691636\\_La\\_hidroclimatologia\\_de\\_Colombia\\_Una\\_sintesis\\_desde\\_la\\_escalainterdecadal\\_hasta\\_la\\_escaladiurna](https://www.researchgate.net/publication/284691636_La_hidroclimatologia_de_Colombia_Una_sintesis_desde_la_escalainterdecadal_hasta_la_escaladiurna)
- Poveda, Mesa, O. J., Agudelo, P. A., Álvarez, J. F., Arias, P. A., Moreno, H. A., Salazar, L. F., Toro, V. G., Vieira, S. C., Jaramillo, A., & Guzman, O. (2002). Diagnóstico del ciclo diurno de la precipitación en los Andes tropicales de Colombia. *Meteorología Colombiana*, 5, 23–30.
- Rivera, R. M. (2008). *Estimación estadística de valores faltantes en una serie de lluvia*. 59(3), 260–273. <https://doi.org/10.1017/CBO9781107415324.004>
- Satgé, F., Defrance, D., Sultan, B., Bonnet, M. P., Seyler, F., Rouché, N., Pierron, F., & Paturel, J. E. (2020). Evaluation of 23 gridded precipitation datasets across West Africa. *Journal of Hydrology*, 581(July 2019), 124412. <https://doi.org/10.1016/j.jhydrol.2019.124412>
- Sattari, M., & Rezazadeh Joudi, A. (2016). Assessment of different methods for estimation of missing data in precipitation studies. *Hydrology Research*, 48. <https://doi.org/10.2166/nh.2016.364>
- Shen, G., Chen, N., Wang, W., & Chen, Z. (2019). WHU-SGCC: A novel approach for blending daily satellite (CHIRP) and precipitation observations over the Jinsha River basin. *Earth System Science Data*, 11(4), 1711–1744. <https://doi.org/10.5194/essd-11-1711-2019>
- Suhaila, J., Sayang, M. D., & Jemain, A. A. (2008). Revised spatial weighting methods for estimation of missing rainfall data. *Asia-Pacific Journal of Atmospheric Sciences*,

- 44(2), 93–104.
- Tapiador, F. J., Turk, F. J., Petersen, W., Hou, A. Y., García-Ortega, E., Machado, L. A. T., Angelis, C. F., Salio, P., Kidd, C., Huffman, G. J., & de Castro, M. (2012). Global precipitation measurement: Methods, datasets and applications. *Atmospheric Research*, 104–105, 70–97.  
<https://doi.org/https://doi.org/10.1016/j.atmosres.2011.10.021>
- Teegavarapu, R. (2009). Estimation of missing precipitation records integrating surface interpolation techniques and spatio-temporal association rules. *Journal of Hydroinformatics - J HYDROINFORM*, 11. <https://doi.org/10.2166/hydro.2009.009>
- Teegavarapu, R. S. V. (2012). Interpolation spatiale à l'aide de modèles non linéaires de programmation mathématique pour l'estimation de précipitations manquantes. *Hydrological Sciences Journal*, 57(3), 383–406.  
<https://doi.org/10.1080/02626667.2012.665994>
- Teegavarapu, R. S. V., & Chandramouli, V. (2005). Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*, 312(1–4), 191–206.  
<https://doi.org/10.1016/j.jhydrol.2005.02.015>
- Telesca, V., Caniani, D., Calace, S., Marotta, L., & Mancini, I. M. (2017). Daily temperature and precipitation prediction using neuro-fuzzy networks and weather generators. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10409 LNCS, 441–455.  
[https://doi.org/10.1007/978-3-319-62407-5\\_31](https://doi.org/10.1007/978-3-319-62407-5_31)
- Tian, Y., & Peters-Lidard, C. D. (2010). A global map of uncertainties in satellite-based precipitation measurements. *Geophysical Research Letters*, 37(24).  
<https://doi.org/10.1029/2010GL046008>
- Toro Trujillo, A. M., Arteaga Ramírez, R., Vázquez Peña, M. A., & Ibáñez Castillo, L. A. (2017). Relleno de series diarias de precipitación, temperatura mínima, máxima de la región norte del Urabá Antioqueño. *Revista Mexicana de Ciencias Agrícolas*, 6(3), 577–588. <https://doi.org/10.29312/remexca.v6i3.640>
- Toté, C., Patricio, D., Boogaard, H., van der Wijngaart, R., Tarnavsky, E., & Funk, C. (2015). Evaluation of satellite rainfall estimates for drought and flood monitoring in Mozambique. *Remote Sensing*, 7(2), 1758–1776.  
<https://doi.org/10.3390/rs70201758>

- Twumasi, Y. A., Annan, J. B., Merem, E. C., Namwamba, J. B., Ayala-Silva, T., Ning, Z. H., Asare-Ansah, A. B., Oppong, J., Frimpong, D. B., Loh, P. M., Owusu, F., Kangwana, L. A., Mwakimi, O. S., Petja, B. M., Okwemba, R., Akinrinwoye, C. O., Mosby, H. J., & McClendon-Peralta, J. (2021). Time Series Analysis on Selected Rainfall Stations Data in Louisiana Using ARIMA Approach. In *Open Journal of Statistics* (Vol. 11, Issue 05, pp. 655–672). <https://doi.org/10.4236/ojs.2021.115039>
- Urrea, V., Ochoa, A., & Mesa, O. (2016). Validación de la base de datos de precipitación CHIRPS para Colombia a escala diaria, mensual y anual en el período 1981-2014. *XXVII Congreso Latinoamericano de Hidráulica*, 11. <http://ladhi2016.org/>
- Valencia, S., Marín, D. E., Gómez, D., Hoyos, N., Salazar, J. F., & Villegas, J. C. (2023). Spatio-temporal assessment of Gridded precipitation products across topographic and climatic gradients in Colombia. *Atmospheric Research*, 285, 106643. <https://doi.org/https://doi.org/10.1016/j.atmosres.2023.106643>
- Vergara, S., Alvarez, O., & Arias, P. (2018). *Metodología para la desagregación espacial de campos diarios de precipitación con modelos regresivos aplicada para Colombia*.
- Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists*. Wiley. <https://books.google.com.co/books?id=WBwSyvIvNY8C>
- Wilks, D. S. (1999). Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology*, 93(3), 153–169. [https://doi.org/10.1016/S0168-1923\(98\)00125-7](https://doi.org/10.1016/S0168-1923(98)00125-7)
- World Meteorological Organization. (2003). *OBSERVATION NETWORKS AND SYSTEMS* (Issue 52).
- Xia, Y., Fabian, P., Stohl, A., & Winterhalter, M. (1999). Forest climatology: Estimation of missing values for Bavaria, Germany. *Agricultural and Forest Meteorology*, 96(1–3), 131–144. [https://doi.org/10.1016/S0168-1923\(99\)00056-8](https://doi.org/10.1016/S0168-1923(99)00056-8)
- Xu, G., Xu, X., Liu, M., Sun, A. Y., & Wang, K. (2015). Spatial Downscaling of TRMM Precipitation Product Using a Combined Multifractal and Regression Approach: Demonstration for South China. *Water*, 7(6), 3083–3102. <https://doi.org/10.3390/w7063083>
- Young, K. C. (1992). A three-way model for interpolating for monthly precipitation values. *Monthly Weather Review*, 120(11), 2561–2569. <https://doi.org/10.1175/1520->

0493(1992)120<2561:ATWMFI>2.0.CO;2

# 9. Anexos

## 9.1 Series utilizadas

- Series estaciones del IDEAM en Caldas:

<https://drive.google.com/drive/folders/10BXC5B53bk7C6WSA2apnnn-7EyRJAOUq?usp=sharing>

- Series información satelital:

<https://drive.google.com/drive/folders/1qPOgX9Ls7bvX574WGfM8VA01zaLC4szL?usp=sharing>

## 9.2 Programas en Python

[https://drive.google.com/drive/folders/1rxTx\\_fxKFQzyyvYddgWqLy63ZsnfFhwG?usp=sharing](https://drive.google.com/drive/folders/1rxTx_fxKFQzyyvYddgWqLy63ZsnfFhwG?usp=sharing)

## 9.3 Gráficas de la comparación de la precipitación media mensual entre los años 1982 y 1986 para las estaciones del IDEAM y los productos satelitales

[https://drive.google.com/drive/folders/1U29zjJIC6\\_XxX\\_TBywEGP86qbbMaiBfS?usp=sharing](https://drive.google.com/drive/folders/1U29zjJIC6_XxX_TBywEGP86qbbMaiBfS?usp=sharing)

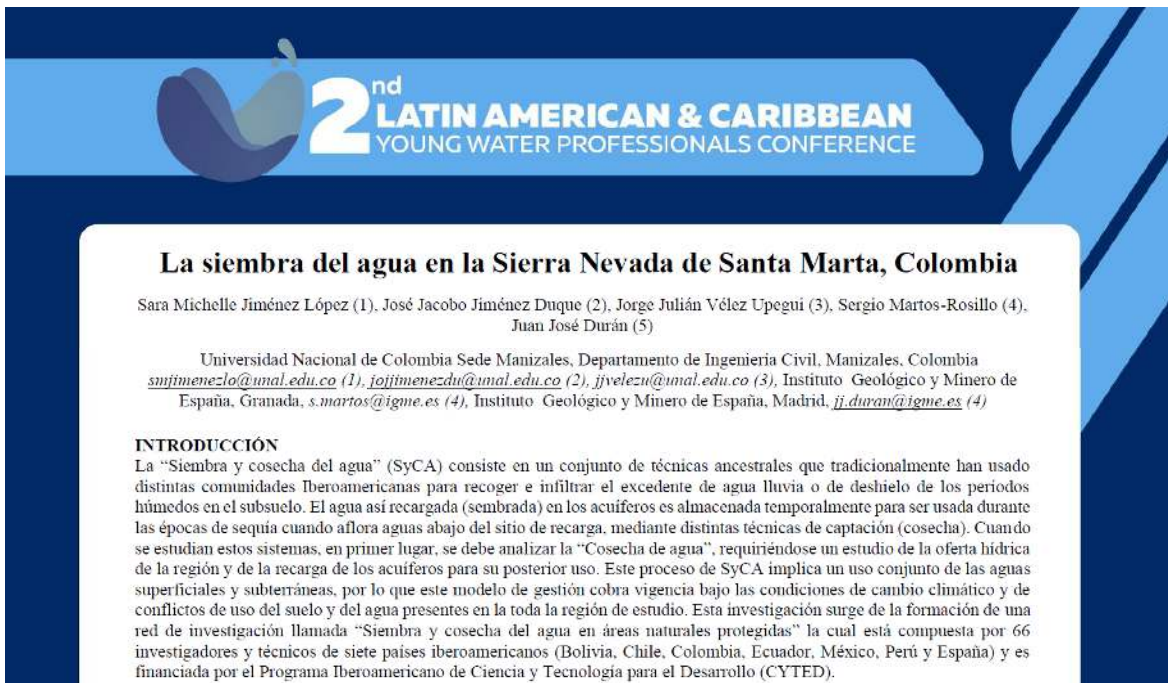
## 9.4 Curvas de masa doble

<https://drive.google.com/drive/folders/1GnM3RXCifqsaUz3oSgRd9NRluBDmpZZ?usp=sharing>

## 9.5 Curvas de masa simple

<https://drive.google.com/drive/folders/1NBWzSqsjYkIVjHa22-3dBj9ymM9F3IZq?usp=sharing>

## 9.6 Póster presentando en 2nd LATIN AMERICAN & CARIBBEAN YOUNG WATER PROFESSIONALS CONFERENCE



**2<sup>nd</sup> LATIN AMERICAN & CARIBBEAN YOUNG WATER PROFESSIONALS CONFERENCE**

**La siembra del agua en la Sierra Nevada de Santa Marta, Colombia**

Sara Michelle Jiménez López (1), José Jacobo Jiménez Duque (2), Jorge Julián Vélez Upegui (3), Sergio Martos-Rosillo (4), Juan José Durán (5)

Universidad Nacional de Colombia Sede Manizales, Departamento de Ingeniería Civil, Manizales, Colombia [smjimenezlo@unal.edu.co](mailto:smjimenezlo@unal.edu.co) (1), [jojjimenezdu@unal.edu.co](mailto:jojjimenezdu@unal.edu.co) (2), [jvvelezu@unal.edu.co](mailto:jvvelezu@unal.edu.co) (3), Instituto Geológico y Minero de España, Granada, [s.martos@igme.es](mailto:s.martos@igme.es) (4), Instituto Geológico y Minero de España, Madrid, [jj.duran@igme.es](mailto:jj.duran@igme.es) (4)

**INTRODUCCIÓN**

La "Siembra y cosecha del agua" (SyCA) consiste en un conjunto de técnicas ancestrales que tradicionalmente han usado distintas comunidades Iberoamericanas para recoger e infiltrar el excedente de agua lluvia o de deshielo de los periodos húmedos en el subsuelo. El agua así recargada (sembrada) en los acuíferos es almacenada temporalmente para ser usada durante las épocas de sequía cuando aflora aguas abajo del sitio de recarga, mediante distintas técnicas de captación (cosecha). Cuando se estudian estos sistemas, en primer lugar, se debe analizar la "Cosecha de agua", requiriéndose un estudio de la oferta hídrica de la región y de la recarga de los acuíferos para su posterior uso. Este proceso de SyCA implica un uso conjunto de las aguas superficiales y subterráneas, por lo que este modelo de gestión cobra vigencia bajo las condiciones de cambio climático y de conflictos de uso del suelo y del agua presentes en la toda la región de estudio. Esta investigación surge de la formación de una red de investigación llamada "Siembra y cosecha del agua en áreas naturales protegidas" la cual está compuesta por 66 investigadores y técnicos de siete países iberoamericanos (Bolivia, Chile, Colombia, Ecuador, México, Perú y España) y es financiada por el Programa Iberoamericano de Ciencia y Tecnología para el Desarrollo (CYTED).

## 9.7 Certificado ponencia XXIX CONGRESO LATINOAMERICANO DE HIDRÁULICA – MÉXICO 2021



ASOCIACIÓN  
MEXICANA  
DE  
HIDRÁULICA



XXIX CONGRESO LATINOAMERICANO DE  
HIDRÁULICA MÉXICO 2021



IAHR  
Hosted by  
South Water  
and IWH, China

**La Asociación Mexicana de Hidráulica (AMH) y la Asociación Internacional de Ingeniería e Investigación Hidro-Ambiental (IAHR)**

Otorgan el presente Certificado a:

**José Jacobo Jiménez Duque**

Por su participación como Ponente del Artículo COMPARACIÓN DE LA LLUVIA EXTREMA UTILIZANDO INFORMACIÓN SATELITAL EN LA ZONA CAFETERA COLOMBIANA con ID 44 en el marco del XXIX Congreso Latinoamericano de Hidráulica "México 2021" Celebrado del 15 al 19 de Noviembre en formato VIRTUAL



**PROF. PABLO DANIEL SPALLETTI**  
Presidente del Comité División Latinoamericana IAHR



**M. EN I. DANIEL MARTINEZ BAZUA**  
Presidente Asociación Mexicana de Hidráulica

No del Certificado: CATCLHM21011



## 9.8 Póster presentando en Spatial Statistics 2023: Climate and the Environment



### Strategies for gap-filling daily rainfall for water management in a tropical Andes region

José Jacobo Jiménez-Duque, Jorge Julián Vélez-Upegui

Universidad Nacional de Colombia Sede Manizales, Civil Engineering Department.

Universidad Nacional de Colombia Sede Manizales, Civil Engineering Department. IDEA.

#### Abstract:

The gap-filling rainfall data is required for water resources management, hydrological modelling and land planning, among others. Recently remote sensors data from different reanalysis are available for free, therefore the use of remote sensing data as a support for filling missing precipitation data can be a good strategy, and this paper explore this capability. Firstly, selection of the best satellite source data for the Andes range of Colombia is essential, and CHIRPS data performs better than other platforms, and it is the most appropriate for gap-filling rainfall daily data in the tropical Andes region studied. Secondly, different strategies for gap-filling daily precipitation records as statistics analysis, stochastic approaches and using artificial intelligence techniques are compared with or without satellite data. The results indicate that the use of Satellite Precipitation Products (SPPs) data with artificial neural networks performs better than statistics and stochastic method, but the simple k-nearest method is also a good approach for data filling in Colombian Andes region.

**Keywords:** remote sensors, filling missing data, daily precipitation, artificial intelligence, stochastic methods.

**Research question:** Does the joint use of remote sensors and ground data improve the gap-filling of daily precipitation in the tropical Andes region?

#### Study region -

##### CALDAS:

Elevation: 147-5200 m a.s.l.

PPT: 1400-5600 mm/year

Mean annual temperature:



#### Materials and Methods:

Selected remote sensor platforms, and downloaded satellite data refers to daily precipitation. The Table shows

#### Introduction

Daily precipitation data without missing data are essential for environmental, meteorological, and hydrological studies. The main idea is to complete the missing data, estimating the closest values to the actual value. In general, the input for the methodologies for filling a series of rain gauge stations are the nearby stations (Sattari & Rezaazadeh Joudi, 2016). Remote sensing information sources such as satellites or radars are complementary to ground stations since they have greater spatial coverage, which makes them interesting for filling in the ground station series (World Meteorological Organization, 2018). Furthermore, estimates of global precipitation with uniform precision are not observed with any measurement method other than satellite remote sensing (Okamoto et al., 2005).

The use of satellite precipitation products is attractive due to the characteristics of remote sensing (spatial and temporal coverage and uniform precision in the estimation), these sources of information must be validated in some way (Ebert et al., 2007); for which the estimates of the on-site rain gauges are usually used (Urrea et al., 2016).

The IDW method is probably the most widely used method to estimate missing data in the field of hydrology (Di Piazza et al., 2011; Teegavarapu & Chandramouli, 2005; R. Teegavarapu, 2009). Martins et al. (2022) compared seven gap-filling methods to evaluate the occurrence of the maximum number of gaps, considering random and sequential gaps. The results indicated that the type of gaps affects the performance of the methods for gap-filling rainfall data at a daily scale.

1. Gathering information:  
Remote sensing and ground  
station data analysis  
Description of study area

2. Review of  
methodologies: statistics,  
stochastics, artificial  
intelligence and hybrid