




IDEAM-METEO / 001-2025  
**NOTA TÉCNICA DEL IDEAM**  
*septiembre de 2025*



# **VERIFICACIÓN DE BASES DE DATOS RETICULADOS CHIRPS, IRE Y CUBO DE DATOS FRENTE A REGISTROS**

de precipitación en Colombia durante el periodo 1991-2020

---

**Autores**

Alexander Martínez Pedraza / Julieta Serna Cuenca  
*Grupo de Climatología y Agrometeorología*

**Apoyo técnico**

Paula Castañeda / Ángela Amaya

## Resumen

La disponibilidad de datos de precipitación precisos es fundamental para la gestión de recursos hídricos en Colombia. Este estudio evalúa rigurosamente el desempeño de productos de precipitación reticulada frente a los registros históricos de 1567 estaciones pluviométricas del IDEAM para el período 1991-2020. Los productos analizados fueron CHIRPS V2, CHIRPS V3, el “Cubo de Datos” del IDEAM, y nueve variantes del producto IRE, generado con el software GeoClim. La metodología empleó un conjunto exhaustivo de métricas de bondad de ajuste (p. ej., RMSE, NSE, MIoAd) y un análisis inferencial de las propiedades estadísticas de los errores (pruebas de Jarque-Bera, t-Student, Wilcoxon) para identificar sesgos sistemáticos. Los resultados establecen una clara jerarquía de rendimiento. Los productos IRE, evaluados mediante una verificación dependiente, exhiben la mayor precisión y exactitud, posicionándose como una excelente herramienta de interpolación. Le sigue CHIRPS V3, que demostró ser el producto de estimación independiente más robusto y confiable, a pesar de una tendencia a la sobreestimación. CHIRPS V2 presenta un rendimiento intermedio con bajas métricas de eficiencia, mientras que el Cubo de Datos muestra el desempeño más bajo, con una marcada sobreestimación y errores de alta magnitud. Se concluye que los productos IRE son la opción preferente para aplicaciones que requieran alta precisión. CHIRPS V3 es una alternativa secundaria robusta, desaconsejándose el uso de CHIRPS V2 y, especialmente, del Cubo de Datos en estudios donde la sobreestimación de la precipitación sea un factor crítico.

# CONTENIDO

|   |    |
|---|----|
| 1. INTRODUCCIÓN .....   | 4  |
| 2. DATOS Y METODOLOGÍA.....   | 7  |
| 2.1. Fuentes de Datos.....  | 8  |
| 2.2. Metodología de Verificación .....  | 10 |
| 2.3. Estrategia de Categorización para el Análisis Comparativo .....                        | 16 |
| 3. RESULTADOS Y DISCUSIÓN .....   | 19 |
| 3.1. Verificación de los productos reticulados CHIRPS V2, CHIRPS V3 y Cubo .....            | 20 |
| 3.2. Verificación de los productos reticulados IRE.....                                     | 21 |
| 3.2.1. Determinación de la configuración de GeoClim más eficiente .....                     | 22 |
| 3.2.2. Determinación de la interacción más eficiente de la configuración 3 de GeoClim ..... | 24 |
| 3.3. Verificación de los productos reticulados CHIRPS, Cubo e IRE ..                        | 25 |
| 4. CONCLUSIONES .....   | 29 |
| Referencias Bibliográficas .....  | 33 |
| Anexos.....   | 34 |
| Anexo A: Tablas Completas de Métricas de Verificación .....                                 | 34 |
| Anexo B: Distribución espacial de las métricas de verificación por mes ....                 | 35 |



# 1. INTRODUCCIÓN

La comprensión de los patrones climáticos en Colombia y la gestión efectiva del recurso hídrico dependen fundamentalmente de la disponibilidad de datos de precipitación precisos y confiables. Dada la compleja geografía del país y, la inherente heterogeneidad espacial y temporal de la precipitación, la Red Nacional de Estaciones, si bien es invaluable, a menudo presenta limitaciones en su densidad y cobertura, especialmente en zonas de difícil acceso. Esta realidad impulsa el uso de fuentes complementarias de información, como los productos de precipitación reticulados derivados de la observación satelital y modelos de interpolación.

En este contexto y como antecedente del presente trabajo, la nota técnica del IDEAM –METEO/002-2018 (Martínez & Serna, 2018) ya había documentado las limitaciones en el desempeño de las estimaciones CHIRPS para la estimación de la precipitación en Colombia. Por ello, el proyecto actual tuvo como objetivo evaluar rigurosamente la calidad de diversos productos de precipitación reticulada para su aplicación en estudios hidrológicos y climáticos desarrollados por el IDEAM. Para ello, se realizó una verificación exhaustiva de las series de precipitación obtenidas de CHIRPS V2, CHIRPS V3, el “Cubo de Datos”—un conjunto de datos reticulados generado por el IDEAM con el software CDT a partir del procesamiento de sus series históricas—, y las salidas del software GeoClim conocidas como “IRE”, bajo diferentes escenarios de parametrización. Todas estas fuentes fueron comparadas sistemáticamente con los registros históricos de precipitación observada por parte de la Red Nacional de Estaciones del IDEAM para el período 1991-2020.

El objetivo principal de esta verificación consistió en presentar los resultados y las conclusiones del análisis detallado de la calidad de estos productos reticulados. En la metodología se empleó un conjunto exhaustivo de métricas de bondad de ajuste y un análisis de las propiedades estadísticas de los errores.

Los hallazgos detallados en este informe proporcionan una base sólida para la comprensión del desempeño de cada producto reticulado en el contexto colombiano, permitiendo la identificación de sus fortalezas y debilidades. Esta información es crucial para guiar el uso adecuado, la selección y la posible mejora de estas fuentes de datos en la labor institucional del IDEAM, contribuyendo a la toma de decisiones informadas en materia climática e hidrológica.



## 2. DATOS Y METODOLOGÍA



La verificación de los productos de precipitación reticulados se fundamentó en un conjunto específico de datos observados de la Red Nacional de Estaciones y en los productos reticulados derivados de fuentes satelitales y modelos numéricos. La metodología aplicada para el procesamiento, la verificación y el análisis se estructuró en un flujo de trabajo sistemático y reproducible.

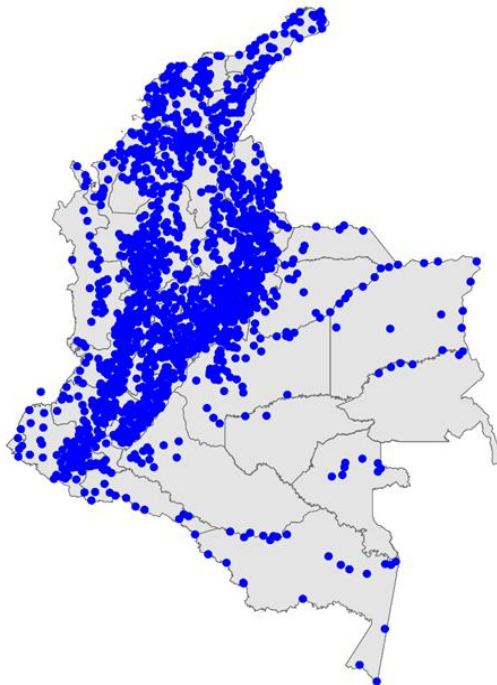
### 2.1. Fuentes de Datos

Las fuentes de datos utilizadas en este estudio para el período 1991-2020 comprendieron:

- **Datos Observados:**

Se emplearon las series históricas de precipitación mensual de la Red Nacional de Estaciones operada por el IDEAM. Estas series corresponden específicamente a las 1567 estaciones que constituyeron el insumo para la creación del conjunto de datos reticulados "Cubo de Datos" en la Subdirección de Meteorología, las cuales cuentan con periodos largos de información dentro del periodo 1991-2020 y distribución espacial como se presenta en la Figura 1.

Este mismo conjunto de estaciones fue, a su vez, la única referencia para la verificación de todos los productos reticulados analizados en este estudio. La información de geolocalización y metadatos de estas estaciones se obtuvo del Catálogo Nacional de Estaciones.



**Figura 1.** Ubicación de 1567 estaciones meteorológicas utilizadas para la verificación de los productos reticulados

## ● Productos Reticulados:

- **CHIRPS V2:** Correspondiente a la versión 2 del producto Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS), una estimación cuasi-global de precipitación con una resolución espacial de  $0.05^\circ \times 0.05^\circ$ . Los archivos de precipitación mensual (.tif.gz) fueron obtenidos desde la ruta:

[https://data.chc.ucsb.edu/products/CHIRPS-2.0/global\\_monthly/](https://data.chc.ucsb.edu/products/CHIRPS-2.0/global_monthly/).

- **CHIRPS V3:** Correspondiente a la versión 3 del producto CHIRPS, con las mismas características de resolución espacial de  $0.05^\circ \times 0.05^\circ$ . Los archivos de precipitación mensual (.tif) fueron obtenidos desde la ruta:

<https://data.chc.ucsb.edu/products/CHIRPS/v3.0/monthly/>.

- **Cubo de Datos:** Este producto representa una salida del software Climate Data Tools (CDT), generado por el IDEAM, el cual integra y procesa las series históricas de precipitación para generar un conjunto de datos reticulados con resolución de  $0.1^\circ$  por  $0.1^\circ$  ( $\sim 11$  km). La cuadrícula tiene 188 por 178 celdas, entre los  $82.5^\circ\text{W}$  y  $65.0^\circ\text{W}$  y entre los  $4.5^\circ\text{S}$  y  $14.0^\circ\text{N}$ .

### **Aspecto importante al analizar los resultados de las comparaciones:**

El **Cubo de Datos** tiene una resolución espacial diferente a la de los otros conjuntos reticulados. Cada píxel cubre un área de  $0.1^\circ \times 0.1^\circ$ , mientras que en **CHIRPSv2**, **CHIRPSv3** e **IRE** la resolución es de  $0.05^\circ \times 0.05^\circ$ .

Esta condición limita la valoración en condiciones equivalentes y tiende a desplazar los resultados de los estadísticos hacia posiciones lejanas.

Para futuros análisis, se recomienda usar el Cubo de Datos con una resolución de  $0.05^\circ \times 0.05^\circ$ , lo que permitirá obtener comparaciones más precisas y homogéneas.



- **IRE (Improved Rainfall Estimates):** Son datos de precipitación que resultan de la combinación de datos ráster con información de estaciones climáticas, utilizando el software GeoCLIM (USGS & CHC-UCSB, 2023). Para este análisis, se generaron productos IRE a partir de tres fuentes de datos de precipitación distintas, a las cuales se les aplicaron tres configuraciones diferentes de parámetros de interpolación dentro de GeoCLIM. Esto resultó en un total de nueve salidas de IRE (3 fuentes × 3 configuraciones = 9 productos IRE finales) que fueron objeto de análisis.

Para la generación de estos productos, se utilizó el mismo conjunto de 1567 estaciones empleado como referencia en la verificación.

Las fuentes de datos de precipitación ráster utilizadas para generar las salidas IRE fueron:

- Valores mensuales de CHIRPS V2.
- Valores mensuales de CHIRPS V3.
- Acumulados mensuales resultantes de la suma de la información decadiaria de CHIRPS V3.

Las tres configuraciones (escenarios de parametrización de interpolación) aplicadas dentro de GeoCLIM para cada una de las fuentes se definieron de la siguiente manera:

**Tabla 1.** Parámetros de interpolación en GeoClim.

| Parámetros de interpolación | Configuración 1 | Configuración 2 | Configuración 3 |
|-----------------------------|-----------------|-----------------|-----------------|
| Weight Power                | 2               | 2               | 3               |
| Max Stations                | 10              | 10              | 12              |
| Search Radius               | 100             | 500             | 500             |
| Factor Fuzz                 | 0               | 1               | 0               |
| Effective Dist              | 100             | 100             | 200             |
| Long Range Value            | 1               | 1               | 1               |
| Max Ratio                   | 3               | 3               | 3               |
| Interpolation Style         | Simple          | Simple          | Simple          |

## 2.2. Metodología de Verificación

*El análisis de datos fue desarrollado con el software libre R (R Core Team, 2022).*

Para verificar la calidad de los productos reticulados de precipitación, se siguieron los lineamientos propuestos por el Grupo de Trabajo Conjunto sobre Investigación de Verificación de Pronósticos (JWGFVR, 2017), asociado al Programa Mundial de Investigaciones Meteorológicas (WWRP, por sus siglas en inglés) de la Organización Meteorológica Mundial (OMM). Específicamente, los métodos de verificación asociados a variables continuas. También se tuvieron en

cuenta las métricas de verificación de pronósticos hidrológicos expuestas en (Dawson et al., 2007).

Un aspecto fundamental del diseño metodológico es la distinción entre los enfoques de verificación aplicados a los diferentes productos, lo cual es crucial para la correcta interpretación de los resultados:

- **Verificación Independiente:** Este enfoque se aplicó a los productos CHIRPS V2 y CHIRPS V3. En este caso, el conjunto de datos de referencia (las 1567 estaciones del IDEAM) no fue un insumo directo en la construcción del producto reticulado. Por lo tanto, la evaluación constituye una prueba rigurosa de la capacidad del producto para estimar la precipitación de forma autónoma.

- **Verificación Dependiente:** Este enfoque se utilizó para los productos IRE y el Cubo de Datos. En este escenario, el mismo conjunto de 1567 estaciones de referencia fue el insumo principal para la generación de los productos. Esta evaluación, por lo tanto, mide la eficacia del algoritmo de interpolación y procesamiento para reproducir los datos de entrada, en lugar de su capacidad de estimación puramente independiente.

Adicionalmente, se realizó un análisis inferencial sobre el error (definido como la diferencia entre los valores reticulados y observados) para identificar y caracterizar la existencia de sesgo sistemático.

Para poder identificar las diferencias asociadas al ciclo anual, la verificación se realizó para cada uno de los meses

del año, ya que se evidenció que el análisis sobre las series completas esconde características que solo se observan mediante la desagregación propuesta.

La verificación se estructuró en dos fases principales:

**Fase 1: Cuantificación de Métricas de Bondad de Ajuste:** Se calcularon métricas ampliamente reconocidas para evaluar el desempeño de las fuentes de información reticulada. Las métricas utilizadas se describen a continuación, donde  $X_i$  es el valor observado,  $Y_i$  es el valor del producto reticulado (estimación),  $\bar{X}$  es la media de los valores observados y  $N$  es el total de datos.

Error Medio (ME) o Sesgo (Bias) Aditivo

Permite realizar una comparación entre los valores observados y los reticulados, indicando si estos últimos presentan una tendencia a subestimar o sobreestimar las variables simuladas por los modelos (Venegas, 2020). Específicamente, estima el promedio del error, un producto reticulado perfecto arrojaría un valor  $ME=0$ . Sin embargo, es posible obtener una puntuación perfecta para una mala estimación si existen errores compensatorios. Un valor positivo  $ME>0$  indica tendencia de sobreestimación y un valor negativo  $ME<0$  indica tendencia de subestimación.

$$ME = \frac{1}{N} \sum_{i=1}^N (Y_i - X_i)$$

### Raíz del Error Cuadrático Medio (RMSE)

Mide el error "promedio", ponderado según el cuadrado del error. Un modelo perfecto tendría un valor  $RMSE=0$ .

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - X_i)^2}$$

### Eficiencia de Nash-Sutcliffe (NSE)

Mide la capacidad predictiva del modelo en relación con la media de los datos observados. Un valor de 1 indica un ajuste perfecto, 0 indica que el modelo tiene la misma habilidad que la media de las observaciones, y valores negativos indican que la media de las observaciones es un mejor predictor que el modelo. (Nash & Sutcliffe, 1970).

El NSE es muy sensible a los valores altos de la serie (picos) y puede ser engañoso en periodos de baja precipitación o cuando hay una gran varianza en los datos observados (Legates & McCabe, 1999).

$$NSE = 1 - \frac{\sum_{i=1}^N (Y_i - X_i)^2}{\sum_{i=1}^N (X_i - \bar{X})^2}$$

### Coefficiente de Eficiencia Modificado (CEM)

El Coeficiente de Eficiencia Modificado (CEM) expuesto en (Legates & McCabe, 1999) es una modificación del Coeficiente de Eficiencia de Nash

– Sutcliffe que busca abordar algunas de sus limitaciones, especialmente en presencia de valores atípicos (Krause et al., 2005). Esta es una métrica muy usada en hidrología, que permite comparar el Error Absoluto Medio con el promedio de los errores absolutos de las observaciones con respecto a la media. Esta medida tiene valor máximo uno (1) y no tiene cota inferior, el valor máximo corresponde al modelo perfecto, el valor de cero (0) indica que no hay diferencia entre el modelo y la media de las observaciones, y los valores negativos indican que el modelo no es apropiado.

$$CEM = 1 - \frac{\sum_{i=1}^N |Y_i - X_i|}{\sum_{i=1}^N |X_i - \bar{X}|}$$

### Índice Modificado de Acuerdo (MIoAd)

El Índice Modificado de Acuerdo (MIoAd) es una métrica estadística empleada para evaluar el grado de concordancia entre los valores observados y los valores simulados o estimados por un modelo o producto. Fue propuesto por Legates & McCabe (1999) como una mejora al Índice de Acuerdo original de Willmott (1981). La principal motivación para esta modificación fue la reconocida sensibilidad del índice original a los valores extremos, lo que podía resultar en una sobreestimación del grado de acuerdo incluso con la presencia de errores significativos. El MIoAd, al igual que el índice original, varía en un rango de 0 a 1. Un valor de 1 indica un acuerdo perfecto entre

los datos observados y los valores del producto reticulado, mientras que valores cercanos a 0 sugieren un bajo nivel de acuerdo. Es una métrica ampliamente utilizada, especialmente en el campo de la hidrometeorología (Krause et al., 2005).

$$MloAd = 1 - \frac{\sum_{i=1}^N |Y_i - X_i|}{\sum_{i=1}^N (|Y_i - \bar{X}| + |X_i - \bar{X}|)}$$

#### Coeficiente de Correlación de Spearman

Esta medida no paramétrica mide la relación lineal entre dos variables. "Este coeficiente, propuesto por Spearman en 1904, es una versión del coeficiente de correlación de Pearson, pero calculado sobre una sucesión transformada de rangos en las dos variables" (Corzo, 2005) y puede ser expresado de la siguiente forma:

$$r_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n d_i^2$$

Donde  $d_i$  es la diferencia entre los rangos de las dos variables comparadas. Para la interpretación es suficiente tener en cuenta que los valores de  $r_s$  cercanos a uno (1) indican asociación positiva o directa, los valores cercanos a menos uno (-1) indican asociación negativa o inversa y los valores cercanos a cero (0) revelan que no hay ningún tipo de asociación.

#### Prueba de Kolmogorov-Smirnov (KS)

La prueba de Kolmogorov-Smirnov es una prueba no paramétrica que se utiliza para determinar si dos distribuciones de probabilidad son similares o si una muestra sigue una distribución teórica específica (Corzo, 2005). Formalmente lo que se plantea es un contraste de la hipótesis.

$$H_0: F_X = F_Y$$

Donde  $F_x$  es la distribución de las observaciones y  $F_x$  es la distribución de los datos reticulados. El contraste se basa en las diferencias entre las frecuencias relativas acumuladas, el estadístico de prueba es:

$$D_n = \max |F_{X_n}(x) - F_{Y_n}(y)|$$

Donde  $F_{x_n}(x)$  es la frecuencia relativa acumulada de las observaciones y  $F_{x_n}(x)$  es la frecuencia relativa acumulada de los datos reticulados.

#### **Fase 2: Análisis Detallado de las Propiedades de los Errores:**

Se profundizó en las características estadísticas de los errores (diferencia entre el valor reticulado y el observado) para comprender su comportamiento. En general, en los trabajos sobre verificación de modelos los esfuerzos se centran en cuantificar el error mediante las métricas descritas en la sección anterior; estas permiten evaluar la exactitud e identificar qué modelos implican menos error.

En este trabajo, se realizó un análisis complementario sobre los errores, tomando como base la Guía para la expresión de la incertidumbre en la medición (JCGM, 2008) donde se plantea que una medición presenta imperfecciones que dan lugar a un error en el resultado de la medición. Tradicionalmente, se considera que un error tiene dos componentes: un componente aleatorio ( $e_R$ ) y un componente sistemático ( $e_S$ ).

$$e_i = e_{R,i} + e_{S,i}$$

Donde  $e_{R,i}$  es el componente del error que en una sucesión de estimaciones varía de forma aleatoria, y  $e_{S,i}$  es el componente del error que en una sucesión de estimaciones varía de forma predecible y es constante en el tiempo. Idealmente se busca que este tipo de error sea nulo o que se pueda reducir lo máximo posible. En este contexto, dado que los valores reticulados analizados en este trabajo son sucesiones temporales, se plantea un enfoque de análisis sobre los errores en una serie temporal, asumiendo que los errores son netamente aleatorios y siguen una distribución de ruido blanco gaussiano, es decir, que los errores se distribuyen normal con media cero y varianza constante ( $e_t \sim N(0, \sigma^2)$ ) (Cowpertwait & Metcalfe, 2009).

Por lo tanto, para verificar los supuestos, se utilizan los estadísticos de prueba que se describen a continuación:

#### Prueba de normalidad de Jarque-Bera

Esta prueba se realiza para verificar si los errores siguen una distribución normal, es decir, se contrasta la hipótesis

*$H_0$ : Los errores se ajustan a una distribución Normal*

El estadístico de prueba Jarque-Bera se describe de forma detallada en Jarque & Bera (1980). Bajo  $H_0$  el estadístico se distribuye asintóticamente como una distribución chi cuadrado con dos grados de libertad.

De forma complementaria a la prueba Jarque – Bera se calculó el coeficiente de asimetría de los errores de tal manera que:

- Si se rechaza  $H_0$  y coeficiente de asimetría es mayor que cero da indicios de sesgo positivo en la distribución.
- Si se rechaza  $H_0$  y coeficiente de asimetría es menor que cero da indicios de sesgo negativo en la distribución.

#### Prueba t-Student

Para verificar si la media de los errores de pronóstico es estadísticamente cero o por el contrario es mayor o menor, asumiendo el supuesto de distribución normal, se contrastan las siguientes pruebas de hipótesis a una y dos vías de la siguiente manera:

- Dos vías:  $H_0: \mu_{error} = 0$  (Si se rechaza  $H_0$ , da indicios que la media de los errores es diferente de cero)

- Una vía:  $H_0: \mu_{error} \leq 0$  (Si se rechaza  $H_0$ , da indicios que la media de los errores es mayor que cero, indicando que el pronóstico tiende a sobreestimar)
- Una vía:  $H_0: \mu_{error} \geq 0$  (Si se rechaza  $H_0$ , da indicios que la media de los errores es menor que cero, indicando que el pronóstico tiende a subestimar)

En este caso, el estadístico de prueba es

$$T_c = \frac{\bar{x}_{error}}{s / \sqrt{n}}$$

Donde  $s$  es la desviación estándar de los errores y  $n$  es el número de pronósticos. Bajo  $H_0$  cierta,  $T_c$  tiene una distribución t - Student con  $n-1$  grados de libertad  $T_c \sim t_{(n-1)}$ .

#### Prueba de rangos signados de Wilcoxon

Es una prueba no paramétrica, utilizada en lugar de la t de Student cuando los datos no cumplen el requisito de normalidad, como en el caso de la precipitación. En este análisis se utiliza para comprobar si la tendencia central, la mediana, de los errores de pronóstico es cero. Aquí también se contrastaron tres hipótesis de la siguiente manera:

- Dos vías:  $H_0: Mediana_{error} = 0$
- Una vía:  $H_0: Mediana_{error} \leq 0$  (Si se rechaza  $H_0$ , da indicios que la mediana de los errores es mayor que cero,

indicando que el pronóstico tiende a sobreestimar)

- Una vía:  $H_0: Mediana_{error} \geq 0$ , (Si se rechaza  $H_0$ , da indicios que la mediana de los errores es menor que cero, indicando que el pronóstico tiende a subestimar)

Para el contraste de la hipótesis se asignan rangos desde 1 hasta  $n$  atendiendo únicamente al valor absoluto del error en el tiempo  $i$  ( $e_i$ ) y se suman los rangos correspondientes a los errores positivos y a los errores negativos por separado. Si la hipótesis nula es cierta, es de esperar que los rangos se distribuyan aleatoriamente entre las diferencias positivas y negativas y, por tanto, que ambas sumas de rangos sean aproximadamente iguales. El estadístico de prueba,  $T$ , es la menor de las dos sumas de rangos. Cuando  $n > 15$  la distribución muestral de  $T$  bajo el supuesto de que  $H_0$  es cierta se aproxima a una normal de parámetros:

$$\mu_T = \frac{n(n+1)}{4} \quad \sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$$

El estadístico de prueba es el valor  $Z$ :

$$Z = \frac{T - \mu_T}{\sigma_T}$$

Que se distribuye según una normal estándar.

### 2.3. Estrategia de Categorización para el Análisis Comparativo

Para facilitar una comparación sistemática y cuantitativa del desempeño entre los distintos productos reticulados, se implementó una estrategia de categorización sobre los resultados de las métricas de bondad de ajuste y las pruebas de hipótesis. Este proceso permitió consolidar los resultados de las 1567 estaciones en tablas de resumen porcentuales y representaciones espaciales (mapas). La lógica de categorización se adaptó a la naturaleza de cada métrica, como se describe a continuación:

- **Categorización por Intervalos Fijos:** Para métricas de eficiencia y correlación como la Eficiencia de Nash-Sutcliffe (NSE), el Coeficiente de Eficiencia Modificado (CEM), el Índice Modificado de Acuerdo (MIoAd) y el Coeficiente de Correlación de Spearman, se utilizaron umbrales predefinidos y fijos. Estos umbrales se basan en rangos de desempeño comúnmente aceptados en la literatura, permitiendo clasificar los resultados en categorías como "Bajo", "Moderado", "Bueno" o "Excelente".
- **Categorización Adaptativa por Cuantiles:** Para métricas de error como el Error Medio (Bias) y la Raíz del Error Cuadrático Medio (RMSE), se empleó un método adaptativo que genera los intervalos (categorías) a partir de la distribución de los propios datos. Para métricas de error simétricas como el Bias, se utilizó una función que crea categorías divergentes en torno a un intervalo central cercano a cero  $(-0.05, 0.05]$ , asegurando una representación equitativa de la subestimación y la sobreestimación. Para métricas de error siempre positivas como RMSE, los cortes se definieron usando cuantiles para reflejar la densidad de los datos.




• **Categorización de Pruebas de Hipótesis:** Los resultados de las pruebas estadísticas se agruparon según sus conclusiones directas. Por ejemplo:

- Para las pruebas t-Student y Wilcoxon, los resultados se clasificaron en tres categorías: "Sobrestima" (si se rechaza  $H_0: \mu \leq 0$ ), "Subestima" (si se rechaza  $H_0: \mu \geq 0$ ) o "H0" (si no se rechaza la hipótesis nula de que la media/mediana del error es cero).
- Para la prueba de normalidad de Jarque-Bera, las categorías fueron: "Dist. Normal" (no rechazo de  $H_0$ ), "Sesgo positivo" o "Sesgo negativo" (rechazo de  $H_0$ , clasificado según el signo del coeficiente de asimetría).
- Para las pruebas basadas en valores p (como Kolmogorov-Smirnov), los resultados se clasificaron como "Significativo ( $<0.05$ )" y "No Significativo ( $\geq 0.05$ )".

Una vez que cada estación fue clasificada, los resultados se agregaron para calcular el porcentaje de casos que caen dentro de cada categoría para cada producto reticulado. Estos resúmenes porcentuales se consolidaron en tablas comparativas (ver **Anexo A: Tablas Completas de Métricas de Verificación**) y se utilizaron para simbolizar los puntos de las estaciones en mapas (ver **Anexo B: Distribución espacial de las métricas de verificación por mes**), permitiendo un análisis visual de los patrones espaciales de desempeño de cada producto.



A tropical landscape featuring several tall palm trees in the foreground and middle ground. A river flows through the scene, reflecting the warm, golden light of a sunset or sunrise. The background shows a dense forest of palm trees under a hazy, orange sky.

# 3. RESULTADOS Y DISCUSIÓN



### 3.1. Verificación de los productos reticulados CHIRPS V2, CHIRPS V3 y Cubo

Los siguientes párrafos resumen los hallazgos de la comparación entre los productos CHIRPS V2, CHIRPS V3 y Cubo de Datos. El soporte numérico completo para esta fase del análisis se encuentra en el archivo Anexo\_A\_1.xlsx, mientras que la representación gráfica de la distribución espacial de los resultados se presenta en el Anexo B.

El análisis de los resultados de la verificación de estos productos reticulados revela las siguientes características en su desempeño:

- **Error Medio (Bias):** CHIRPS V3 presenta una clara predominancia de valores de error cercanos a cero. Específicamente, en el intervalo  $(-5.91, 8.18]$ , CHIRPS V3 duplica el porcentaje de valores en comparación con CHIRPS V2 y Cubo. Por su parte, el Cubo muestra una marcada evidencia de sobreestimación, con más del 53% de sus valores en categorías de error positivo de mayor magnitud. Aunque CHIRPS V3 también tiende a sobreestimar, lo hace con magnitudes de error más bajas que el Cubo.
- **Raíz del Error Cuadrático Medio (RMSE):** Esta métrica confirma la predominancia de CHIRPS V3 en las magnitudes de error más bajas, con mejor desempeño en los meses de enero y febrero. También se confirma que el conjunto Cubo es el que presenta mayores magnitudes de error. Para las tres fuentes comparadas los meses de mayores magnitudes de error son mayo, octubre y noviembre.
- **Eficiencia de Nash-Sutcliffe (NSE):** El conjunto Cubo muestra una baja eficiencia, con más del 50% de sus valores en el intervalo  $[-\text{Inf}, 0]$  en el periodo de diciembre a marzo. En general, el Cubo presenta un menor rendimiento que los otros conjuntos de datos a lo largo de todo el año. Por el contrario, CHIRPS V3 indica una mayor eficiencia, con porcentajes promedio de 41.5% en el intervalo  $(0.5, 1]$ .
- **Coeficiente de Eficiencia Modificado (CEM):** El coeficiente revela una falta de eficiencia en CHIRPS V2, que presenta más del 62.5% de sus valores en el intervalo  $[-\text{Inf}, 0]$  en todos los meses. En ese mismo intervalo, los conjuntos CHIRPS V3 y Cubo también muestran baja eficiencia, con más del 50% de sus valores en los meses de abril a junio y de septiembre a noviembre.
- **Índice Modificado de Acuerdo (MIoAd):** Este índice confirma una menor concordancia de CHIRPS V2 con los datos observados. Le sigue el Cubo, y CHIRPS V3 muestra la mayor concordancia. Sin embargo, es importante resaltar que los tres conjuntos de datos se concentran los valores en el intervalo  $(0.5, 0.75]$ , lo que indica un nivel de concordancia moderado.

- **Coefficiente de Correlación de Spearman:** En cuanto a la asociación con los datos observados, CHIRPS V2 y CHIRPS V3 demuestran un alto grado de correlación, con porcentajes acumulados superiores al 62% en el intervalo (0.6, 1] durante todos los meses del año. El conjunto **Cubo** tiene una mayor incidencia en el intervalo (0.4, 0.8].
- **Prueba de Kolmogorov-Smirnov (KS):** La prueba de igualdad de distribuciones muestra para CHIRPS V3 porcentajes de no rechazo mayores a 76%. Mientras que CHIRPS V2 tiene en promedio porcentajes de no rechazo de 61% y Cubo porcentajes promedio de no rechazo de 67%.
- **Prueba de Jarque-Bera:** Los tres conjuntos de datos exhiben un comportamiento normal en sus errores durante los meses de abril a mayo y de septiembre a noviembre, superando el 72% de los casos. Los meses con mayor rechazo de la normalidad son enero y febrero. CHIRPS V3 muestra mejores características de normalidad (71.6% de no rechazo promedio), seguido de CHIRPS V2 (68%) y Cubo (64%). En los casos de rechazo de la normalidad, el coeficiente de asimetría indica sesgo negativo para las dos versiones de CHIRPS, mientras que para el Cubo muestra un alto sesgo positivo en los meses de diciembre a marzo.
- **Prueba t-Student:** El producto reticulado que mejor se ajusta a la hipótesis nula ( $H_0: \mu_{error} = 0$ ) es CHIRPS V3, con un promedio de no rechazo del 66%. Le siguen CHIRPS V2 (41%), y el Cubo (40%). En cuanto al contraste de las hipótesis de una vía ( $H_0: \mu_{error} \leq 0$  y  $H_0: \mu_{error} \geq 0$ ), existe evidencia estadística de que el Cubo y CHIRPS V3 tienden a sobreestimar los valores observados, con el Cubo mostrando porcentajes superiores al 42%.
- **Prueba Wilcoxon:** La prueba de Wilcoxon para la hipótesis nula  $H_0: Mediana_{error} = 0$  muestra que CHIRPS V3 es el producto con el mayor promedio de no rechazo por mes (60%). Le siguen CHIRPS V2 (40%), y el Cubo (36%). El contraste de las hipótesis de una vía ( $H_0: Mediana_{error} \leq 0$  y  $H_0: Mediana_{error} \geq 0$ ) proporciona evidencia estadística para afirmar que el Cubo y CHIRPS V3 tienen una tendencia a la sobreestimación de los valores observados. Además, CHIRPS V2 también muestra indicios de sobreestimación, a excepción del mes de marzo.

En resumen, los resultados indican que, de estos tres productos, CHIRPS V3 es la opción más robusta y confiable, dada su superioridad en precisión, eficiencia y concordancia.

### 3.2. Verificación de los productos reticulados IRE

En esta fase se evaluó el desempeño de los nueve productos de precipitación IRE generados con el software GeoClim para determinar la configuración y la interacción más eficiente. Los resultados numéricos completos que respaldan el análisis de esta sección están disponibles en el archivo Anexo\_A\_2.xlsx.

### 3.2.1. Determinación de la configuración de GeoClim más eficiente

El análisis comparativo de la interacción entre las tres configuraciones y las fuentes de datos (CHIRPS V2 - Mensual, CHIRPS V3 - Mensual y CHIRPS V3 - Decadal) demostró que la Configuración 3 exhibió el mejor desempeño, seguida por la Configuración 1 y, en último lugar, la Configuración 2. Esto sugiere que la configuración implementada en GeoClim ejerce una influencia más significativa sobre las diferencias observadas en el rendimiento, a pesar de que el desempeño base de las fuentes sea semejante. A continuación, se resumen sus características principales:

- **Error Medio (Bias):** La Configuración 3 presenta un predominio de valores cercanos a cero, con una distribución simétrica entre valores negativos y positivos. La Configuración 1 también muestra una distribución simétrica, mientras que la Configuración 2 tiene un menor porcentaje de valores cercanos a cero y su distribución evidencia una tendencia a la sobreestimación.

- **Raíz del Error Cuadrático Medio (RMSE):** En las tres configuraciones, los valores de RMSE más cercanos a cero se presentaron en el periodo de diciembre a febrero. Por el contrario, las magnitudes más altas se registraron en los periodos de abril-mayo y octubre-noviembre. Específicamente, la Configuración 3 tuvo el mayor porcentaje de valores cercanos a cero, mientras que la Configuración 2 presentó el mayor porcentaje de valores en las categorías de magnitudes altas.

- **Eficiencia de Nash-Sutcliffe (NSE):** Las configuraciones 1 y 3 mostraron ajustes casi perfectos, con porcentajes superiores al 70% en el intervalo (0.75, 1]. No obstante, la Configuración 3 demostró ser la más robusta, con la mayor incidencia de valores en el intervalo (0.9, 1]. En contraste, la mayoría de los valores de la Configuración 2 se ubicaron en el intervalo (0.5, 0.9], y la distribución de los valores negativos (intervalo [-Inf,0]) fue similar en las tres configuraciones.

- **Coeficiente de Eficiencia Modificado (CEM):** Las configuraciones 1 y 3 mostraron un alto porcentaje de valores en el intervalo (0.5, 0.9], superando el 50% en todos los meses. Se destaca que la Configuración 3 demostró una mayor eficiencia, con más del 20% de sus valores en el intervalo de rendimiento superior (0.9, 1]. Por el contrario, la Configuración 2 tuvo la mayor incidencia en el intervalo inferior (0, 0.5] con porcentajes superiores al 50% y también presentó el mayor porcentaje de valores negativos (intervalo [-Inf, 0]) a lo largo de todos los meses.

- **Índice Modificado de Acuerdo (MIoAd):** Los valores de este índice indicaron un mayor grado de concordancia para la Configuración 3, la cual tiene una incidencia superior al 40% en el intervalo (0.9,1]. La Configuración 1, por su lado, presentó la mayor incidencia en el intervalo (0.75,0.9], con

porcentajes que superan el 45%. En contraste, la Configuración 2 exhibió el menor grado de concordancia, con porcentajes superiores al 50% en el intervalo (0.5, 0.75].

- **Coeficiente de Correlación de Spearman:** Las configuraciones 1 y 3 mostraron distribuciones similares en términos de correlación, con más del 90% de sus valores ubicados en el intervalo (0.8,1]. En contraste, para ese mismo intervalo, la Configuración 2 presentó porcentajes alrededor del 70%.

- **Prueba de Kolmogorov-Smirnov (KS):** Desde el enfoque de igualdad entre las distribuciones de los datos, la Configuración 2 fue la que más rechazó la hipótesis nula, con un máximo de 13.6% en enero.

- **Prueba de Jarque-Bera:** Indica que los errores, en general, tienen una distribución normal. El promedio de no rechazo de la hipótesis nula es de 71% para la Configuración 2, 66% para la Configuración 1 y 63% para la Configuración 3.

- **Prueba t-Student:** para la hipótesis nula de que la media del error es cero ( $H_0 : \mu_{error} = 0$ ), muestra promedios mensuales de no rechazo de 57% para la Configuración 2, y 45% para las configuraciones 1 y 3. El contraste con las hipótesis de una vía ( $H_0 : \mu_{error} \leq 0$  y  $H_0 : \mu_{error} \geq 0$ ) proporciona evidencia estadística de que la Configuración 2 tiene una tendencia a sobreestimar los valores observados.

- **Prueba de Wilcoxon:** para la hipótesis nula de que la mediana del error es cero ( $H_0 : Mediana_{error} = 0$ ), indica promedios mensuales de no rechazo de 54% para la Configuración 2, y 43% para las configuraciones 1 y 3. Las pruebas de una vía  $H_0 : Mediana_{error} \leq 0$  y  $H_0 : Mediana_{error} \geq 0$  dan también evidencia estadística para afirmar que la Configuración 2 tiende a sobreestimar los valores observados.

A partir de estos resultados, se puede concluir que la Configuración 3, en interacción con las tres fuentes de análisis, demostró la mayor precisión y porcentajes de exactitud aceptable para representar los valores observados. Por lo tanto, se tomó la decisión de continuar el análisis de los productos IRE exclusivamente con aquellos asociados a esta configuración, priorizando así un menor error general. Los productos seleccionados para la siguiente etapa de análisis fueron:

- IRE\_V2\_MENSUAL\_Configuracion\_3 (IRE - CHIRPS V2 – Mensual – Configuración 3)
- IRE\_V3\_MENSUAL\_Configuracion\_3 (IRE - CHIRPS V3 – Mensual – Configuración 3)
- IRE\_V3\_DECADAL\_Configuracion\_3 (IRE - CHIRPS V3 – Decadal – Configuración 3)

### 3.2.2. Determinación de la interacción más eficiente de la configuración 3 de GeoClim

El siguiente paso consistió en comparar estas tres interacciones entre sí, cuyos resultados se describen a continuación.

- **Error Medio (Bias):** Las distribuciones del Error Medio (Bias) son similares entre las tres interacciones. Sin embargo, se observó una ligera ventaja de IRE\_V3\_MENSUAL, seguido por IRE\_V2\_MENSUAL y, por último, IRE\_V3\_DECADAL. Los periodos con valores de error medio más cercanos a cero fueron de diciembre a marzo, mientras que los de mayores magnitudes fueron mayo, octubre y noviembre.
- **Raíz del Error Cuadrático Medio (RMSE):** En los tres escenarios, los valores de RMSE más cercanos a cero se presentaron entre diciembre y febrero. En contraste, las magnitudes más altas se registraron en los periodos de abril-mayo y octubre-noviembre. Específicamente, IRE\_V2\_MENSUAL tuvo el mayor porcentaje de valores cercanos a cero, mientras que IRE\_V3\_DECADAL presentó el mayor porcentaje en las categorías de magnitudes altas.
- **Eficiencia de Nash-Sutcliffe (NSE):** Los tres escenarios tuvieron ajustes casi perfectos, con porcentajes superiores al 70% en el intervalo (0.75,1]. Sin embargo, IRE\_V2\_MENSUAL se destacó con una mayor incidencia en el intervalo de rendimiento superior (0.9,1], mientras que IRE\_V3\_MENSUAL e IRE\_V3\_DECADAL mostraron distribuciones muy similares entre sí.
- **Coeficiente de Eficiencia Modificado (CEM):** Aunque las distribuciones fueron muy semejantes y los tres escenarios acumularon más del 50% de sus valores en el intervalo (0.75,1], se logró identificar que IRE\_V2\_MENSUAL tuvo mayor incidencia en el intervalo superior (0.9,1]. Por su parte, IRE\_V3\_DECADAL tuvo una mayor incidencia en el intervalo (0.5,0.75]. De acuerdo con este coeficiente, el orden de eficiencia propuesto es IRE\_V2\_MENSUAL, IRE\_V3\_MENSUAL y, en último lugar, IRE\_V3\_DECADAL.
- **Índice Modificado de Acuerdo (MIoAd):** Los resultados de este índice mostraron una correspondencia con los hallazgos del Coeficiente de Eficiencia Modificado (CEM), confirmando la superioridad de IRE\_V2\_MENSUAL.
- **Coeficiente de Correlación de Spearman:** No se observaron diferencias significativas entre los tres escenarios, ya que los porcentajes en el intervalo de correlación (0.8,1] superaron el 91% en todos los casos.

El análisis de las tres fuentes bajo la Configuración 3 de GeoClim permite concluir que el orden de relevancia es, en primer lugar, IRE\_V2\_MENSUAL, seguido por IRE\_V3\_MENSUAL y, por último, IRE\_V3\_DECADAL. Sin embargo, los tres escenarios son considerados válidos para su uso, dado que están ajustados con información de estaciones y la influencia de la Configuración 3 por sí misma garantiza la robustez de los productos.

Finalmente, para la verificación general, se decidió continuar solo con los productos IRE\_V2\_MENSUAL e IRE\_V3\_MENSUAL. Esta decisión se fundamenta en el anuncio de que la versión CHIRPS V2 será descontinuada, lo que convierte a CHIRPS V3 en la fuente principal de información a futuro.

### 3.3. Verificación de los productos reticulados CHIRPS, Cubo e IRE

En esta sección se presenta el análisis comparativo final entre los productos reticulados seleccionados: CHIRPS V2, CHIRPS V3, Cubo e IRE (considerando las versiones IRE\_V2\_MENSUAL e IRE\_V3\_MENSUAL). Los datos que sustentan esta discusión se encuentran de forma exhaustiva en las tablas del archivo Anexo\_A\_3.xlsx y en los mapas del Anexo B.

- **Error Medio (Bias):** En el análisis del Error Medio (Bias), se observa que los dos productos IRE tienen un mejor rendimiento en comparación con CHIRPS V2, CHIRPS V3 y el Cubo. En el intervalo  $(-0.05, 0.05]$ , los productos IRE concentran un promedio del 5.8% de sus valores, mientras que los otros productos no superan el 0.7%. Ampliando el rango al intervalo  $(-3.06, 4.67]$ , los productos IRE acumulan alrededor del 45% de los valores, seguidos por CHIRPS V3 (21%), CHIRPS V2 (11.8%) y el Cubo (9%). Los meses de enero a marzo tienen las magnitudes de error más bajas, mientras que abril-mayo y octubre-noviembre tienen las más altas. Se identificó que CHIRPS V3 y el Cubo muestran una tendencia marcada a la sobreestimación. El Cubo sobreestima con mayor magnitud de diciembre a marzo, mientras que CHIRPS V3 lo hace en los meses de septiembre, mayo y octubre.

- **Raíz del Error Cuadrático Medio (RMSE):** Los resultados de esta métrica confirman que los productos IRE tienen el menor error, con una ligera ventaja de IRE\_V2\_MENSUAL sobre IRE\_V3\_MENSUAL. Estos productos demuestran mayor precisión de junio a agosto y de diciembre a febrero. Por otro lado, aunque las distribuciones de CHIRPS V3 y CHIRPS V2 son muy similares, CHIRPS V3 muestra una mejor bondad de ajuste. Ambos productos presentan errores más altos de abril a mayo y de octubre a noviembre, y más bajos de enero a febrero (especialmente en la región del Caribe y el piedemonte llanero) y de junio a septiembre (en los piedemontes llanero y amazónico y el sur de la región Andina). El producto Cubo tiene los valores más altos de error, especialmente de noviembre a marzo.

- **Eficiencia de Nash-Sutcliffe (NSE):** Los productos IRE mostraron un ajuste casi perfecto, con más del 70% de sus valores ubicados en el intervalo  $(0.75, 1]$ . Los demás productos reticulados fueron menos eficientes. En el intervalo  $(0.5, 1]$ , CHIRPS V3 acumuló en promedio el 41.5%, CHIRPS V2 el 27% y el Cubo el 21%. Según este índice, no se recomienda el uso del Cubo, especialmente de diciembre a febrero, ya que en este periodo acumula más del 54% de sus valores en el intervalo  $[-\text{Inf}, 0]$ .

- **Coefficiente de Eficiencia Modificado (CEM):** Los valores de este coeficiente evidencian la robustez de los productos IRE, con acumulados superiores al 80% en el intervalo de alta eficiencia (0.5,1]. En ese mismo intervalo, los otros productos no acumulan más del 3%. Según esta métrica, no se considera apropiado el uso de CHIRPS V2, dado que sus acumulados en el intervalo  $[-\text{Inf},0]$  superan el 62% de los casos por mes.
- **Índice Modificado de Acuerdo (MIoAd):** Este índice confirma la superioridad de los productos IRE, a la vez que califica a CHIRPS V2 como el conjunto reticulado de menor rendimiento. En el intervalo de alto rendimiento (0.5,1]: IRE\_V2\_MENSUAL e IRE\_V3\_MENSUAL acumulan el 97% de los casos evaluados por mes. Por su parte, CHIRPS V3 acumula alrededor del 78%, el Cubo cerca del 67% y CHIRPS V2 alrededor del 58%.
- **Coefficiente de Correlación de Spearman:** Los productos IRE\_V2\_MENSUAL e IRE\_V3\_MENSUAL son los mejor correlacionados, con acumulados superiores al 92.8% y 91.5% respectivamente, en el intervalo de correlación (0.8,1]. Para los otros productos, la correlación en el intervalo (0.6,1] es considerablemente menor con promedios de porcentaje (CHIRPS V3: 73.5%, CHIRPS V2: 72.2%, Cubo: 61.1%).
- **Prueba de Kolmogorov-Smirnov (KS):** Esta prueba, que evalúa la igualdad de distribución entre los datos observados y los conjuntos reticulados, indica que los productos IRE tienen un rendimiento superior. IRE\_V2\_MENSUAL e IRE\_V3\_



MENSUAL muestran acumulados de no rechazo de la hipótesis nula superiores al 93%, mientras que CHIRPS V3 tiene un porcentaje superior al 76%.

- **Prueba Jarque-Bera y Análisis de Asimetría:** La prueba de normalidad de los errores indica que CHIRPS V3 es el producto que mejor se ajusta a una distribución normal, con porcentajes de no rechazo de la hipótesis nula alrededor del 71.5%. El mejor ajuste se observa en los periodos de septiembre a noviembre y de abril a mayo, mientras que el menor ajuste se presenta entre enero y febrero. En los casos de rechazo, el análisis de la asimetría revela que CHIRPS V2 y CHIRPS V3 tienen un sesgo negativo bien definido. Por otro lado, IRE\_V2\_MENSUAL e IRE\_V3\_MENSUAL muestran una leve tendencia a un sesgo positivo. Por su parte, en el Cubo, los porcentajes más altos indican un sesgo positivo, especialmente de noviembre a marzo.

- **Prueba t-Student:** El producto reticulado que mejor se ajusta a la hipótesis nula ( $H_0: \mu_{error} = 0$ ) es CHIRPS V3, con un promedio de no rechazo del 66%. Le siguen IRE\_V3\_MENSUAL (45%), IRE\_V2\_MENSUAL y CHIRPS V2 (41% cada uno), y el Cubo (40%).

- **Prueba Wilcoxon:** La prueba de Wilcoxon para la hipótesis nula  $H_0: Mediana = 0$  muestra que CHIRPS V3 es el producto con el mayor promedio de no rechazo por mes (60%). Le siguen IRE\_V3\_MENSUAL (44%), IRE\_V2\_MENSUAL y CHIRPS V2 (40% cada uno), y el Cubo (36%).



# 4. CONCLUSIONES

Desempeño General de los Productos



Los resultados de la verificación demuestran una clara jerarquía en el rendimiento de los productos reticulados. Los productos IRE (IRE\_V2\_MENSUAL e IRE\_V3\_MENSUAL) superan consistentemente a CHIRPS V3, que a su vez se desempeña mejor que CHIRPS V2. Finalmente, el conjunto de datos Cubo se ubica en el último lugar en la mayoría de las métricas.

Para interpretar correctamente esta jerarquía, es crucial considerar el enfoque de verificación. Los productos IRE y Cubo fueron evaluados usando el mismo conjunto de 1567 estaciones que se utilizó para su construcción, lo que se conoce como una verificación dependiente. En contraste, los productos CHIRPS fueron sometidos a una verificación independiente.

Desglosando el desempeño de cada producto bajo este contexto:

- **Productos IRE:** En la cima de este rendimiento, exhiben un desempeño superior en la mayoría de los análisis. Presentan una mayor precisión, con valores de error más cercanos a cero (Bias, RMSE), una eficiencia casi perfecta (NSE, CEM) y una alta concordancia y correlación con los datos observados (MIoAd, Spearman). Su capacidad para aproximarse a la distribución de los datos de las estaciones también es la más alta (Prueba KS). Este contexto confirma su excelencia como herramienta de interpolación de muy alta precisión.

- **CHIRPS V3:** En una meritoria segunda posición, este producto demuestra un rendimiento robusto y confiable, superando a sus contrapartes CHIRPS V2 y Cubo. Aunque presenta una tendencia a la sobreestimación, la magnitud de sus errores es significativamente menor que la del Cubo. Se destaca por ser el producto cuyos errores se ajustan mejor a una distribución normal (Jarque-Bera) y a las hipótesis de error cero (t-Student, Wilcoxon).

Cabe destacar que, si bien los productos IRE muestran errores de menor magnitud (RMSE) y un sesgo general más bajo (Bias), sus errores residuales se alejan con mayor frecuencia de una media/mediana estadísticamente igual a cero en comparación con CHIRPS V3, cuyos errores, aunque de mayor magnitud, presentan una distribución más centrada. Todo esto lo posiciona como el producto de estimación independiente más confiable para el país.

- **CHIRPS V2:** Se sitúa en un rendimiento intermedio. Si bien muestra una alta correlación, sus métricas de eficiencia (CEM) y concordancia (MIoAd) son las más bajas entre todos los productos, lo que sugiere limitaciones en su capacidad para estimar con precisión la precipitación.

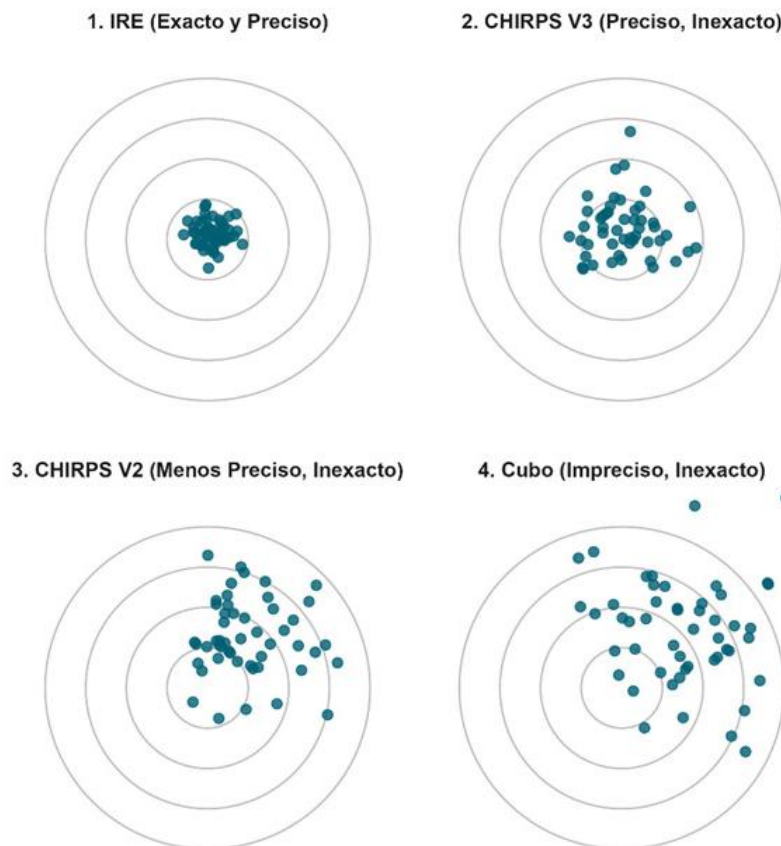
- **Cubo de Datos:** Finalmente, este conjunto presenta el rendimiento más bajo. A pesar de contar con la ventaja metodológica de la verificación dependiente, su principal debilidad es una fuerte tendencia a la sobreestimación con errores de alta magnitud. Esto evidencia limitaciones significativas tanto en su resolución espacial ( $0.1^\circ$ ), considerablemente más gruesa que la de los productos CHIRPS, como en su algoritmo de procesamiento.

Como complemento a la evaluación cuantitativa, se realizó una simulación de datos para representar visualmente el comportamiento característico de cada producto analizado. Basándose en los hallazgos de precisión, exactitud y distribución de errores, se generaron las ilustraciones de las Figuras Figura 2 y Figura 3. Estas gráficas ejemplifican de manera intuitiva la clara superioridad de los productos IRE, el rendimiento robusto pero sesgado de CHIRPS V3, y la considerable imprecisión e inexactitud del conjunto Cubo, reforzando así la jerarquía de desempeño establecida en este estudio.

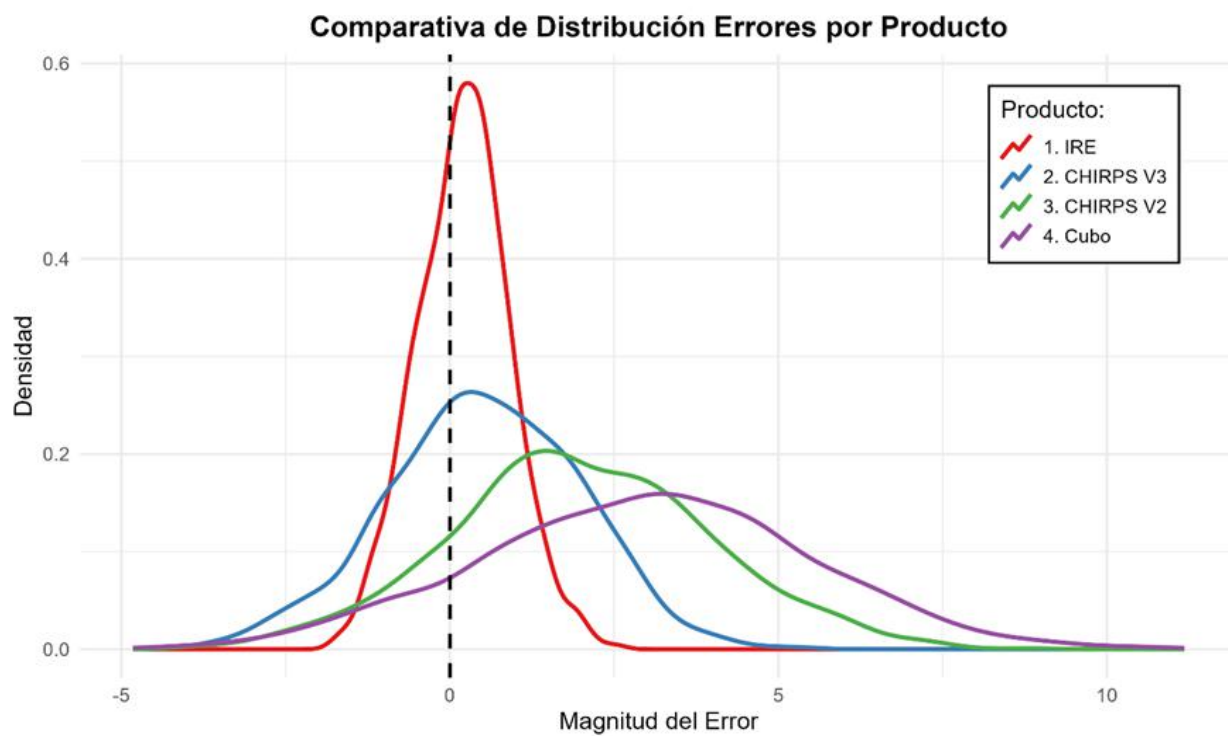
### Recomendaciones de Uso

En base a estos hallazgos, se recomienda priorizar el uso de los productos IRE (específicamente IRE\_V2\_MENSUAL e IRE\_V3\_MENSUAL) para estudios hidrológicos y climáticos que requieran la mayor precisión posible. CHIRPS V3 se presenta como una alternativa secundaria robusta. Por el contrario, se desaconseja el uso de CHIRPS V2 y, especialmente, del Cubo de Datos para aplicaciones donde la sobreestimación de la precipitación pueda generar resultados críticos".

#### Comparativa de Precisión y Exactitud por Producto



**Figura 2.** Simulación de las características identificadas en los productos reticulados (precisión y exactitud)



**Figura 3.** Simulación de las características identificadas en los productos reticulados (funciones de distribución)



## Referencias Bibliográficas

Corzo, J. A. (2005). *Notas de clase. Estadística no paramétrica (Métodos basados en rangos)*. Univ. Nacional de Colombia.

Cowpertwait, P. S. P., & Metcalfe, A. V. (2009). *Introductory time series with R* (1st ed.). Springer-Verlag New York.

**<https://doi.org/10.1007/978-0-387-88698-5>**

Dawson, C. W., Abrahart, R. J., & See, L. M. (2007). HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling and Software*, 22(7), 1034–1052.

**<https://doi.org/10.1016/j.envsoft.2006.06.008>**

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters*, 6, 255–259.

JCGM. (2008). *Evaluation of measurement data-Guide to the expression of uncertainty in measurement - JCGM 100:2008*. **[www.bipm.org](http://www.bipm.org)**

JWGFVR. (2017). *7th International Verification Methods Workshop*.

**<https://www.cawcr.gov.au/projects/verification/>**

Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences*, 5, 89–97.

**<https://doi.org/10.5194/adgeo-5-89-2005>**

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, 35(1), 233–241. **<https://doi.org/10.1029/1998WR900018>**

Martinez, A., & Serna, J. (2018). *Validación de las estimaciones de precipitación con CHIRPS e IRE/IDEAM – METEO/002-2018*.

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, 10(3), 282–290.

**[https://doi.org/https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/https://doi.org/10.1016/0022-1694(70)90255-6)**

USGS, & CHC-UCSB. (2023). *GeoCLIM 3.1.0 Manual de Usuario*.

**<https://edcintl.cr.usgs.gov/downloads/sciweb1/shared/fews/geoclim/>**

Venegas, T. (2020). *Verificación de los modelos globales GFS y ECMWF, y del modelo regional WRF para las variables precipitación, viento y temperatura en Chile durante los años 2018-2019*.

Willmott, C. J. (1981). ON THE VALIDATION OF MODELS. *Physical Geography*, 2(2), 184–194. **<https://doi.org/10.1080/02723646.1981.10642213>**

## Anexos

### Anexo A: Tablas Completas de Métricas de Verificación

Los resultados numéricos detallados y la categorización completa para todas las métricas y pruebas de hipótesis se encuentran en los archivos de Microsoft Excel adjuntos.

Cada archivo de Excel contiene 10 hojas de cálculo, una para cada una de las siguientes métricas y pruebas analizadas en el cuerpo del documento:

- Error Medio (Bias)
- Raíz del Error Cuadrático Medio (RMSE)
- Eficiencia de Nash-Sutcliffe (NSE)
- Coeficiente de Eficiencia Modificado (CEM)
- Índice Modificado de Acuerdo (MIoAd)
- Coeficiente de Correlación de Spearman
- Prueba de Kolmogorov-Smirnov (KS)
- Prueba de Normalidad de Jarque-Bera
- Prueba t-Student para la media del error
- Prueba de rangos signados de Wilcoxon para la mediana del error

El contenido de los anexos se organiza de la siguiente manera, correspondiendo a cada una de las fases del análisis presentado en la sección de "Resultados y discusión":

- **Anexo\_A\_1.xlsx:** Contiene los resultados del análisis comparativo inicial entre los productos CHIRPS V2, CHIRPS V3 y Cubo de Datos. Estos datos son la base del análisis presentado en la sección 3.1.
- **Anexo\_A\_2.xlsx:** Contiene los resultados de la verificación de los nueve productos IRE generados con GeoClim. Estos datos respaldan el análisis de la sección 3.2 para determinar la configuración más eficiente.
- **Anexo\_A\_3.xlsx:** Contiene los resultados de la comparación final entre los productos seleccionados: CHIRPS V2, CHIRPS V3, Cubo de Datos y las dos mejores versiones de IRE. Estos datos son el fundamento del análisis general presentado en la sección 3.3.

## **Anexo B: Distribución espacial de las métricas de verificación por mes**

Este anexo contiene el compendio gráfico completo de los resultados de la verificación, presentando la distribución espacial de cada métrica para todos los productos analizados a escala mensual. El conjunto completo de mapas se encuentra consolidado en el documento PDF adjunto:

Anexo\_B\_Distribucion\_espacial\_metricas.pdf.

El documento está organizado en 10 secciones principales, cada una dedicada a una métrica de verificación o prueba de hipótesis específica, permitiendo un análisis visual detallado del comportamiento de los productos a lo largo del año.

La estructura es la siguiente:

- Anexo B1. Distribución espacial del Error Medio (Bias)
- Anexo B2. Distribución espacial de la Raíz del Error Cuadrático Medio (RMSE)
- Anexo B3. Distribución espacial de la Eficiencia de Nash-Sutcliffe (NSE)
- Anexo B4. Distribución espacial del Coeficiente de Eficiencia Modificado (CEM)
- Anexo B5. Distribución espacial del Índice Modificado de Acuerdo (MIoAd)
- Anexo B6. Distribución espacial del Coeficiente de Correlación de Spearman
- Anexo B7. Distribución espacial de la Prueba de Kolmogorov-Smirnov (KS)
- Anexo B8. Distribución espacial de la Prueba de Normalidad de Jarque-Bera
- Anexo B9. Distribución espacial de la Prueba t-Student para la media del error
- Anexo B10. Distribución espacial de la Prueba de rangos signados de Wilcoxon para la mediana del error

Dentro de cada una de estas secciones, se presentan 12 gráficos, uno por cada mes del año. A su vez, cada gráfico mensual está compuesto por un panel de seis mapas que muestran la distribución espacial de la métrica para los siguientes productos reticulados, facilitando su comparación directa:

- |                  |                   |
|------------------|-------------------|
| 1. CHIRPS V2     | 4. IRE_V2_MENSUAL |
| 2. CHIRPS V3     | 5. IRE_V3_MENSUAL |
| 3. Cubo de Datos | 6. IRE_V3_DECADAL |



**IDEAM**

Instituto de Hidrología, Meteorología  
y Estudios Ambientales

