

# Estimación de datos faltantes de lluvia mensual a través de la asimilación de información satelital y pluviométrica en una cuenca andina tropical

*Estimation of monthly rainfall missing data through rain gauge and satellite data assimilation in a tropical Andean river basin*

A.F. Cruz-Roa<sup>1\*</sup>, M.I. Barrios<sup>1</sup>

## RESUMEN

Las series de datos hidrometeorológicos generalmente presentan datos faltantes o erróneos por falta de continuidad en los registros o errores instrumentales, surgiendo la necesidad de desarrollar métodos numéricos para calcular estimaciones cercanas al valor real, contribuyendo a mejorar la consistencia de los estudios de manejo y ordenación de los recursos hídricos. El objetivo de la investigación es comparar cuatro métodos para estimar datos faltantes de lluvia mensual y evaluar los errores de estimación en cuatro escenarios desde un 10% a un 40% de datos faltantes. Como caso de estudio se ha tomado la red de estaciones pluviométricas de la cuenca hidrográfica del río Coello durante el periodo de 1998 al 2005. Se compararon el método tradicional de la relación normal y el krigado y dos derivaciones que utilizan la elevación y datos de lluvia estimada mediante satélite como información auxiliar. Los métodos fueron evaluados por medio de cuatro índices de ajuste (Raíz cuadrada del error cuadrático medio, Error medio absoluto, Error de sesgo promedio y Coeficiente de correlación de Pearson) calculados entre los valores observados y estimados en los cuatro escenarios de datos faltantes. De acuerdo con los resultados, el método que minimiza el error de estimación de los datos faltantes de lluvia corresponde al cokrigado con la elevación como información auxiliar (RMSE = 15,54, MAE = 7,21, MBE = -0,23 y R = 0,97). La investigación concluye que los métodos que consideran la dependencia espacial proporcionan mejores estimaciones de la lluvia a escala mensual, recomendando los métodos geoestadísticos.

**Palabras Clave:** Datos faltantes de lluvia; Cokrigado, Elevación, Cuenca del río Coello.

## ABSTRACT

*Hydrometeorological data sets usually present missing or erroneous data due to lack of continuity in the records or instrumental errors, which imply the need to develop numerical methods for calculating a reliable estimation, which contributes to improving the consistency in studies of water resources management. The objective of the research is to compare four methods to estimate missing monthly rainfall data and to evaluate the estimation errors in four scenarios (from 10% to 40% of missing data). As a case study, the rainfall gauge stations of the Coello river basin were analyzed during the period from 1998 to 2005. We applied the traditional method of the normal ratio, the ordinary kriging technique and cokriging using elevation and rainfall data estimated by satellite as auxiliary information. The performance of the four methods was evaluated by applying four indices (Root Mean Square Error, Mean Absolute Error, Mean Bias Error, and Pearson correlation coefficient), which were calculated for the four scenarios of missing data. According to results, cokriging with elevation as auxiliary information was the best method to minimize the estimation error (RMSE = 15.54, MAE = 7.21, MBE = -0.23 and R = 0.97). The research concludes that methods that consider spatial dependence provide better estimates of rainfall in the monthly scale.*

**Key Words:** Missing rainfall data; Cokriging; Elevation; Coello river basin.

<sup>1</sup> Facultad de Ingeniería Forestal, Universidad del Tolima, Barrio Santa Helena, Ibagué, Colombia.

\*Autor por correspondencia: afcruz@ut.edu.co

Fecha de recepción: 19 enero, 2018.

Fecha de aceptación: 13 agosto, 2018.

## Introducción

Es habitual requerir series continuas y completas de lluvia a escala mensual para la realización de análisis hidrológicos y climáticos tendientes a soportar instrumentos de planificación del recurso hídrico, tales como análisis y caracterización de sequías, estudios de cambio y variabilidad climática, aprovechamiento de recursos hídricos, diseño de sistemas de utilización conjunta de aguas superficiales y subterráneas, entre otros (Adhikary, Muttill, & Yilmaz, 2017; Cruz-Roa, Olaya-Marín, & Barrios, 2017; Mair & Fares, 2010). Sin embargo, las series de datos de lluvia presentan datos faltantes por mal funcionamiento de los pluviómetros o falta de medición que interrumpe la consistencia y continuidad de los registros históricos (Di Piazza, Conti, Noto, Viola, & La Loggia, 2011; Mair & Fares, 2010). Es común encontrar limitaciones logísticas, económicas y de accesibilidad en terreno que dificultan el establecimiento y mantenimiento de redes de medición de variables meteorológicas, por lo que es necesario desarrollar, aplicar y evaluar métodos de estimación de la lluvia para representar su comportamiento espacio-temporal a nivel mensual (Goovaerts, 2000; Suhaila, Sayang, & Jemain, 2008).

Varios autores (Adhikary *et al.*, 2017; de Amorim Borges, Franke, da Anunciação, Weiss, & Bernhofer, 2016; Di Piazza *et al.*, 2011; Goovaerts, 2000; Mair & Fares, 2010; Suhaila *et al.*, 2008; Teegavarapu & Chandramouli, 2005) sugieren el uso de métodos no-geoestadísticos para completar las series de datos faltantes de lluvia mensual (generalmente regresiones lineales, ponderación de distancia inversa) y métodos geoestadísticos (como los modelos de la familia del krigeado). El método de Relación Normal (RN) es un método no-geoestadístico ampliamente usado para completar datos faltantes de lluvia a escala mensual; este método estima los valores de lluvia erróneos o faltantes en la estación meteorológica bajo revisión con el promedio ponderado de las estaciones de la región hidrológica homogénea (De Silva, Dayawansa, & Ratnasiri, 2007; Radi, Zakaria, & Azman, 2015; Suhaila *et al.*, 2008). El krigeado y sus derivaciones obtienen una estimación de la variable estudiada mediante el ajuste de un modelo de dependencia espacial, el cual es expresado por un variograma experimental con la información disponible, obteniendo valores con mínima varianza de la estimación (Bayat, Nasser, & Zahraie, 2015; Goovaerts, 2000). El krigeado puede considerar variables secundarias, que son utilizadas como

covariables en el modelo, y comúnmente permiten estimar un valor más aproximado a las condiciones reales, en algunos casos la elevación y datos de precipitación estimada por satélite se han utilizado como variables auxiliares (Cruz-Roa *et al.*, 2017; Goovaerts, 2000; Mair & Fares, 2010). En este contexto, Di Piazza *et al.* (2011) y de Amorim Borges *et al.* (2016) sugieren la incorporación de información auxiliar (radar y datos satelitales) de factores geográficos considerados en los métodos geoestadísticos de interpolación espacial de la lluvia. La lluvia y la elevación se encuentran estrechamente relacionadas por la influencia de la orografía del paisaje, siendo la elevación la información auxiliar más usada, como potencial factor para estimar datos faltantes de lluvia en áreas de baja densidad de registros de observación meteorológica (Adhikary *et al.*, 2017; Cruz-Roa *et al.*, 2017; Goovaerts, 2000).

Goovaerts (2000) realizó la comparación de siete algoritmos para interpolar espacialmente datos de lluvia por medio de técnicas geoestadísticas y no-geoestadísticas en la región de Algarve (sur de Portugal), encontrando que los algoritmos geoestadísticos que incluyen variables secundarias (e.g. elevación) se aproximan más a los valores observados, sobrepasando el rendimiento de los algoritmos no-geoestadísticos. Teegavarapu y Chandramouli (2005) compararon cinco métodos para estimar datos de lluvia en veinte estaciones meteorológicas ubicadas en el estado de Kentucky (USA), reconfirmando la versatilidad de los métodos geoestadísticos (principalmente el Krigeado) sobre los métodos no-geoestadísticos tradicionales (e.g. IDW y RN) para estimar datos faltantes de lluvia. De Silva *et al.* (2007) compararon cuatro métodos determinísticos para estimar datos faltantes de lluvia en la isla de Sir Lanka; caracterizada por una topografía plana y costera, en dicho estudio determinaron que el IDW y RN fueron los métodos con mejor desempeño. Suhaila *et al.* (2008) realizaron comparaciones entre los métodos IDW y RN, y modificaciones en las ecuaciones tradicionales incorporando principalmente la dependencia de las distancias Euclidianas entre las estaciones pluviométricas para la península de Malasia; en la investigación no encontraron diferencias significativas entre los métodos no-geoestadísticos comparados para estimar datos faltantes de lluvia, más sí sugiere la importancia de considerar la topografía y factores geográficos que limitan los métodos evaluados. Mair y Fares (2010) compararon métodos de interpolación tradicionales

(polígonos de Thiessen, IDW y Regresión Lineal) y geoestadísticos (krigeado ordinario y Krigeado simple) incorporando la información secundaria de la elevación e isoyetas de máximos de lluvia regional, en los registros históricos de la isla de O'ahu (Hawái, USA), concluyendo que el método de los polígonos de Thiessen tiene alto error en la estimación de datos faltantes de lluvia y el krigeado ordinario produce estimaciones más precisas; confirmando su versatilidad y superioridad sobre los métodos tradicionales.

En el contexto de las condiciones orográficas y climáticas andinas de Colombia, se han reportado estudios recientes que evalúan métodos geoestadísticos para la interpolación de la lluvia anual (Álvarez-Villa, Vélez, & Poveda, 2011; Manz *et al.*, 2016; Poveda *et al.*, 2007; Rogelis & Werner, 2013). Álvarez-Villa *et al.* (2011), plantean que el análisis espacial de la lluvia andina tropical es complejo, y no solo son relevantes los aspectos topográficos de control orográfico sino también la interacción de procesos climático-atmosféricos que inciden en la variabilidad de la lluvia. Algunas investigaciones han desarrollado campos de lluvia a escala anual y mensual (Mejía *et al.*, 1999; Poveda *et al.*, 2007; Vélez, Poveda, & Mesa, 2000), sin embargo, no hay estudios reportados que estén orientados a la evaluación de diferentes métodos para la estimación de datos faltantes de lluvia mensual.

Esta investigación evalúa los errores de estimación de datos faltantes de lluvia mensual a través del tradicional método de la relación normal, el krigeado ordinario, y el cokrigeado ordinario usando como información auxiliar la elevación y datos de lluvia estimada mediante satélite obtenidos del proyecto Tropical Rainfall Measuring Mission Multi-satellite Precipitation Analysis (TMPA), en escenarios desde un 10% a un 40% de datos faltantes. Como caso de estudio se ha tomado la red de estaciones pluviométricas de la cuenca hidrográfica del río Coello.

## Materiales y métodos

### Área de estudio

El área considerada corresponde a la cuenca hidrográfica del río Coello, ubicada en la región andina del departamento del Tolima, de la cordillera central de Colombia. Ésta abarca 1899,31 Km<sup>2</sup> de superficie en el rango altitudinal de 200 a 5300 msnm. Representa una región orográfica compleja desde su nacimiento (Nevado del Tolima) hasta su desembocadura (Río

Magdalena), siendo un área de importancia hídrica para los sistemas de riego en la producción de cultivos agrícolas (Cruz-Roa *et al.*, 2017). La variabilidad de la precipitación a escala mensual en los Andes colombianos están altamente influenciada por la dinámica macroclimática de los fenómenos de El Niño/Southern Oscillation (ENSO), la Zona de Convergencia Intertropical (ZCIT), los vientos alisios y del chorro del Chocó, reflejado en cambios del comportamiento climático bimodal y cambios notables en el promedio anual de la precipitación (Manz *et al.*, 2016; Poveda, 2004; Poveda, Álvarez, & Rueda, 2011).

### Estaciones pluviométricas

El análisis se realizó con base en la red pluviométrica de 16 estaciones operada por el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM) de Colombia, a efecto de este estudio se utilizó el periodo de enero de 1998 a diciembre de 2005 (8 años), la tabla 1 presenta la ubicación geográfica de cada estación, la cantidad de datos faltantes, la referencia y resolución de cada instrumento de medición. Estas estaciones fueron consideradas por su consistencia, distribución espacial y continuidad de los registros pluviométricos, necesarios para la comparación de los métodos de estimación de datos faltantes de lluvia (Figura 1).

### Métodos de estimación de datos faltantes de lluvia

#### Método de la relación normal:

El método de la relación normal calcula la precipitación del mes  $j$  en una estación determinada a través del promedio ponderado de las precipitaciones mensuales observadas en el mes  $j$  de las estaciones que integran el grupo hidrológico homogéneo (Paulhus & Kohler, 1952; Radi *et al.*, 2015). El factor de ponderación de cada estación usada en la estimación se define como el cociente entre la precipitación promedio anual de la estación objeto de estimación y la precipitación promedio anual de la estación predictiva. Este método no tiene en cuenta la dependencia espacial de la precipitación por lo cual, es un método no-geoestadístico.

En la serie de datos bajo análisis se definieron las estaciones de cada grupo hidrológico homogéneo por medio de una correlación lineal de Pearson (R) entre las estaciones pluviométricas (Tabla 1),

Tabla 1: Estaciones pluviométricas de la zona de estudio.

| N°   | Nombre [Código]           | Elevación (m.s.n.m) | Lat. (°) | Long. (°) | P* (mm) | No. de meses faltantes | Referencia | Resolución (mm) |
|------|---------------------------|---------------------|----------|-----------|---------|------------------------|------------|-----------------|
| 1    | Cascada [21210150]        | 3080                | 4.2847   | -75.5419  | 1294    | 60                     | Hellmann   | 0.1             |
| 2    | Cucuana Hda [21215130]    | 2120                | 4.3413   | -75.5185  | 1139    | 60                     | Hellmann   | 0.1             |
| 3    | Plan [21210140]           | 2050                | 4.3512   | -75.5124  | 1209    | 61                     | Hellmann   | 0.1             |
| 4**  | Palogrande Hda [21210170] | 2200                | 4.3500   | -75.4166  | 1638    | 0                      | Hellmann   | 0.1             |
| 5**  | Toche [21210180]          | 2000                | 4.5210   | -75.4187  | 1117    | 3                      | Hellmann   | 0.1             |
| 6    | Silencio [21210260]       | 2500                | 4.6333   | -75.3833  | 1620    | 89                     | Hellmann   | 0.1             |
| 7**  | Palmar [21210220]         | 2200                | 4.5801   | -75.3256  | 1322    | 2                      | Hellmann   | 0.1             |
| 8**  | Juntas [21210020]         | 1765                | 4.5563   | -75.3216  | 1446    | 0                      | Hellmann   | 0.1             |
| 9    | Pastales [21210030]       | 1505                | 4.5110   | -75.3008  | 1956    | 61                     | Hellmann   | 0.1             |
| 10   | Darien [21210160]         | 1920                | 4.4666   | -75.3042  | 2064    | 64                     | Hellmann   | 0.1             |
| 11** | Secreto [21210080]        | 1490                | 4.4945   | -75.2900  | 1832    | 0                      | Hellmann   | 0.1             |
| 12** | Placer [21210110]         | 2170                | 4.5209   | -75.2778  | 1862    | 13                     | Thies      | 0.1             |
| 13** | Esmeralda [21210120]      | 1965                | 4.4909   | -75.2394  | 1576    | 2                      | Hellmann   | 0.1             |
| 14** | Interlaken [21210240]     | 1210                | 4.4363   | -75.2338  | 2191    | 0                      | Hellmann   | 0.1             |
| 15** | Cementos Dia. [21215140]  | 780                 | 4.3313   | -75.0806  | 1355    | 16                     | Hellmann   | 0.1             |
| 16** | Aceituno [21220050]       | 680                 | 4.3534   | -75.0530  | 1355    | 0                      | Hellmann   | 0.1             |

\* Precipitación promedio anual

\*\* Estaciones incluidas en los escenarios de datos faltantes

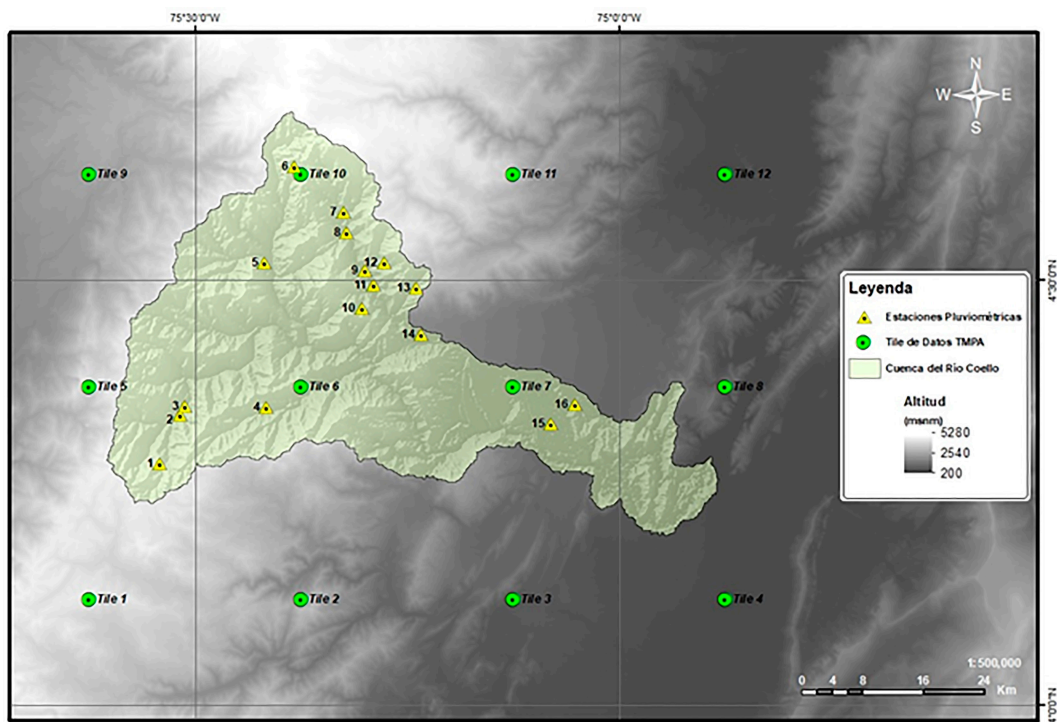


Figura 1. Distribución espacial de la red de estaciones pluviométricas en la cuenca hidrográfica del río Coello.

cuando R tiene valores iguales o mayores que 0,5, las estaciones correspondientes constituyen un grupo hidrológico homogéneo. A partir de este procedimiento se obtuvieron cuatro (4) grupos: el grupo 1 esta conformados por las estaciones 1, 2, 3, el grupo 2 por 4 y 5, el grupo 3 por la estación 6 hasta la 13, y el grupo 4 por las estaciones 14, 15 y 16.

Para cada estación se estimó la precipitación de todos los meses del periodo de análisis con base en la precipitación observada en las restantes estaciones que conforman su respectivo grupo hidrológico homogéneo.

### Métodos geoestadísticos:

El Krigeado Ordinario (OK) se define como el mejor estimador lineal insesgado de la variable no muestreada (Bayat *et al.*, 2015). En esta investigación se utilizó el OK y el Cokrigeado Ordinario (CKO). El Cokrigeado ordinario estima un valor puntual no conocido utilizando algunos valores vecinos conocidos, asociado con una o varias variables auxiliares (Bayat *et al.*, 2015; Palomares, Gil, & Serra, 2014). Para los métodos geoestadísticos de Cokrigeado la información auxiliar utilizadas para estimar datos de lluvia faltantes fueron (i) la elevación del modelo de elevación digital (DEM) de 3" de resolución espacial del proyecto Shuttle Radar Topography Mission (SRTM) disponibles para descarga por el Consortium for Spatial Information del Consultative Group on International Agricultural Research (CGIAR-CSI) y (ii) datos de lluvia satelital TMPA a escala mensual del producto 3B43V7 con 0,25° de resolución espacial, en el periodo de análisis 1998-2005 para 12 celdas, provistos por el proyecto de la NASA Goddard Earth Sciences Data y Information Services Center (Huffman, Adler, Bolvin, & Nelkin, 2010; Huffman, Stocker, Bolvin, Nelkin, & Adler, 2012). Para los métodos geoestadísticos se calcularon los variogramas experimentales y posteriormente las superficies interpoladas por los tres métodos geoestadísticos en la totalidad de las estaciones pluviométricas, calculados en la herramienta *Geostatistical Analyst Wizard* del software *ArcGIS 10.3*.

### Análisis de escenarios de datos faltantes

#### Definición de los escenarios de datos faltantes

Con el fin de indagar acerca del funcionamiento de cada método de estimación para diferentes porcentajes de datos faltantes se definieron cuatro (4) escenarios utilizando las estaciones que presentan

menos de un 20% de datos faltantes (estaciones 4, 5, 7, 8, 11, 12, 13, 14, 15 y 16).

Para definir cada escenario, se extrajo aleatoriamente un porcentaje  $x$  de datos estimados por cada método en cada estación, con la restricción de que cada valor extraído corresponda a un valor originalmente observado. Lo anterior con el fin de garantizar la constatación empírica del desempeño de cada método de estimación.

Los cuatro escenarios definidos corresponden a (i) un 10% de datos estimados (cuando faltan 10 datos de 96 registros totales), (ii) un 20% de datos estimados (19 datos faltantes), (iii) un 30% de datos estimados (29 datos faltantes) y (iv) un 40% de datos estimados (39 datos faltantes).

### Evaluación del desempeño de los métodos de estimación de datos faltantes en cada escenario

Las diferencias entre los valores observados y estimados por los cuatro métodos comparados se determinaron mediante el cálculo de cuatro índices de ajuste para cada uno de los escenarios analizados. Para ello, se utilizaron la raíz cuadrada del error cuadrático medio (RMSE), el error medio absoluto (MAE), el error de sesgo promedio (MBE) y el coeficiente de correlación de Pearson (R), representados en siguientes ecuaciones:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (z_i - \hat{z}_i)^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |z_i - \hat{z}_i|$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (z_i - \hat{z}_i)$$

$$R = \frac{\sum_{i=1}^n (z_i - \bar{z})(\hat{z}_i - \bar{\hat{z}})}{\sqrt{\sum_{i=1}^n (z_i - \bar{z})^2 \sum_{i=1}^n (\hat{z}_i - \bar{\hat{z}})^2}}$$

Donde  $n$  es la longitud del conjunto de datos del respectivo escenario,  $z_i$  es la precipitación observada en el elemento  $i$ ,  $\hat{z}_i$  es la precipitación estimada en el elemento  $i$ ,  $\bar{z}$  es el promedio de lluvia observada en el escenario correspondiente y  $\bar{\hat{z}}$  es el promedio de lluvia estimada en el escenario correspondiente.

### Resultados

Los resultados de la validación presentados en la tabla 2, muestran los valores promedios para cada



Tabla 2: Valores promedio de las pruebas para las estaciones incluidas en los escenarios de datos faltantes para los métodos de Relación Normal (RN), Krigeado Ordinario (OK), Cokrigeado y la elevación (CoKE) y Cokrigeado y datos de lluvia satelital TMPA (CoKT).

| Porcentaje datos faltantes | Método | RMSE <sup>a</sup> | MAE <sup>b</sup> | MBE <sup>c</sup> | R <sup>d</sup> |
|----------------------------|--------|-------------------|------------------|------------------|----------------|
| 10%                        | RN     | 48.37             | 41.08            | 9.39             | 0.75           |
|                            | OK     | 27.39             | 20.69            | 1.56             | 0.94           |
|                            | CoKE   | 2.12              | 1.70             | -0.11            | 1.00           |
|                            | CoKT   | 47.44             | 37.37            | -2.40            | 0.81           |
| 20%                        | RN     | 53.60             | 43.88            | 8.93             | 0.71           |
|                            | OK     | 27.01             | 18.93            | -0.02            | 0.94           |
|                            | CoKE   | 20.75             | 9.38             | 0.10             | 0.96           |
|                            | CoKT   | 46.62             | 34.17            | -2.44            | 0.83           |
| 30%                        | RN     | 53.34             | 42.60            | 6.00             | 0.70           |
|                            | OK     | 28.32             | 18.78            | -0.10            | 0.93           |
|                            | CoKE   | 18.69             | 8.55             | -0.15            | 0.97           |
|                            | CoKT   | 45.00             | 31.82            | -1.32            | 0.82           |
| 40%                        | RN     | 54.39             | 42.30            | 4.74             | 0.71           |
|                            | OK     | 36.20             | 21.23            | 2.17             | 0.89           |
|                            | CoKE   | 20.61             | 9.21             | -0.78            | 0.96           |
|                            | CoKT   | 46.16             | 32.71            | 2.69             | 0.83           |

<sup>a</sup> Raíz cuadrada del error cuadrático medio

<sup>b</sup> Error medio absoluto

<sup>c</sup> Error de sesgo promedio

<sup>d</sup> Coeficiente de correlación de Pearson

prueba de cada método en los cuatro escenarios simulados de datos faltantes de lluvia bajo análisis. Los resultados indican que el método más eficiente que estima valores con un mínimo sesgo de error de los observados por la red meteorológica del IDEAM es el Cokrigeado y la elevación como variable auxiliar (CoKE), expresado por los valores más bajos en las pruebas RMSE (2.12, 20.75, 18.69 y 20.61), MAE (1.70, 9.38, 8.55 y 9.21), MBE (-0.11, 0.10, -0.15 y -0.78) y R (1.00, 0.96, 0.97 y 0.96) de los escenarios respectivamente (ver Figura 2 y 3).

La visualización de resultados de la figura 2 demuestra que los valores de RMSE en todas las estaciones de validación, la eficiencia de CoKE corresponde al método de mejor estimación, en los cuatro escenarios comparados. Los valores de RMSE y MAE tienden a cero, expresando los milímetros estimados de lluvia de diferencia de los observados. MBE expresa que el método CoKE subestima los valores en los escenarios de 10, 30 y 40%, únicamente en el escenario de 20% sobre estima los valores. Los valores de R indican que existen asociaciones positivas (i.e. los valores de R tienden a 1.0) entre los valores observados y los estimados por el método identificado como más eficiente. En el mejor caso, es confiable utilizar el CoKE cuando falten no mayores al 10% de los datos, asegurando estimaciones cercanas a las

observadas (2.12 mm de diferencia a los observados); mientras aumentan los datos faltantes (20, 30 y 40%), el CoKE presenta una tendencia a estabilizarse como se presenta en los valores de RMSE de la figura 3 (20.75, 18.69 y 20.61 respectivamente). Se infiere que el método es versátil y aproximado para varios casos prácticos en los que se enfrentan los hidrólogos para interpolar datos faltantes de lluvia.

Las estimaciones del método no-geoestadístico, la RN indica que este método tiene los valores con mayor sesgo de error comparado con los observados como se presentan en la figura 3, en todos los escenarios (48.37, 53.60, 53.34 y 54.39 respectivamente). Los resultados de este método indican que sobrestima los valores de lluvia en un grado significativo y en general éste es superado por los métodos geoestadísticos. Para el caso del Cokrigeado y datos de lluvia satelital TMPA (CoKT), el método no es adecuado por la resolución espacial de los datos producto; en la figura 3 los valores se encuentran cercanos a los del RN (47.44, 46.62, 45.00 y 46.16) con unas pocas diferencias de los valores de RMSE y MAE entre los escenarios; posiblemente el efecto orográfico del área de estudio genere poca eficiencia para estimar datos faltantes de lluvia para la cuenca del río Coello, más si supera la eficiencia respecto al método no-geoestadístico de la RN.

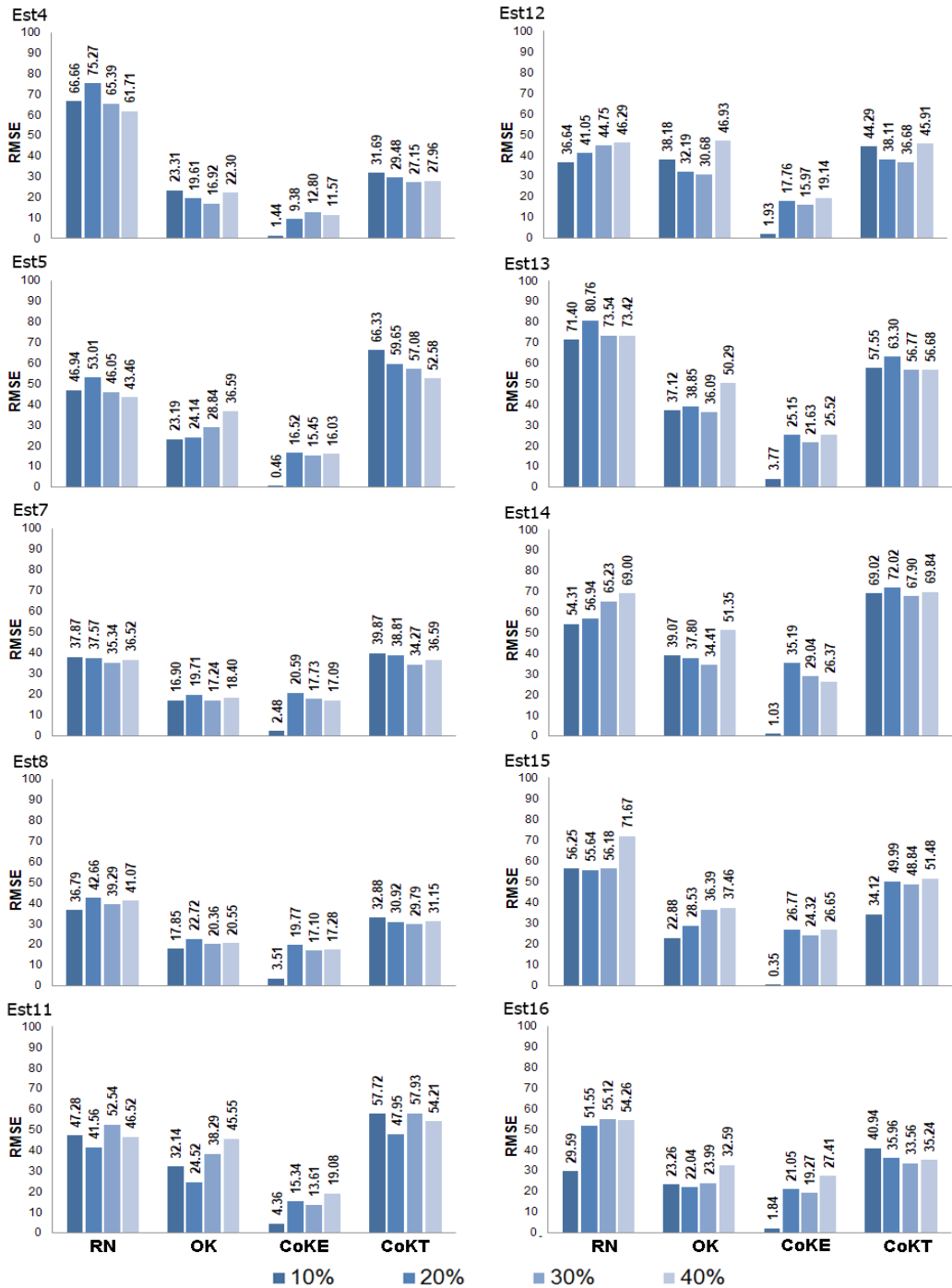


Figura 2. Valores de RMSE de los métodos comparados para las estaciones incluidas en los escenarios de datos faltantes.

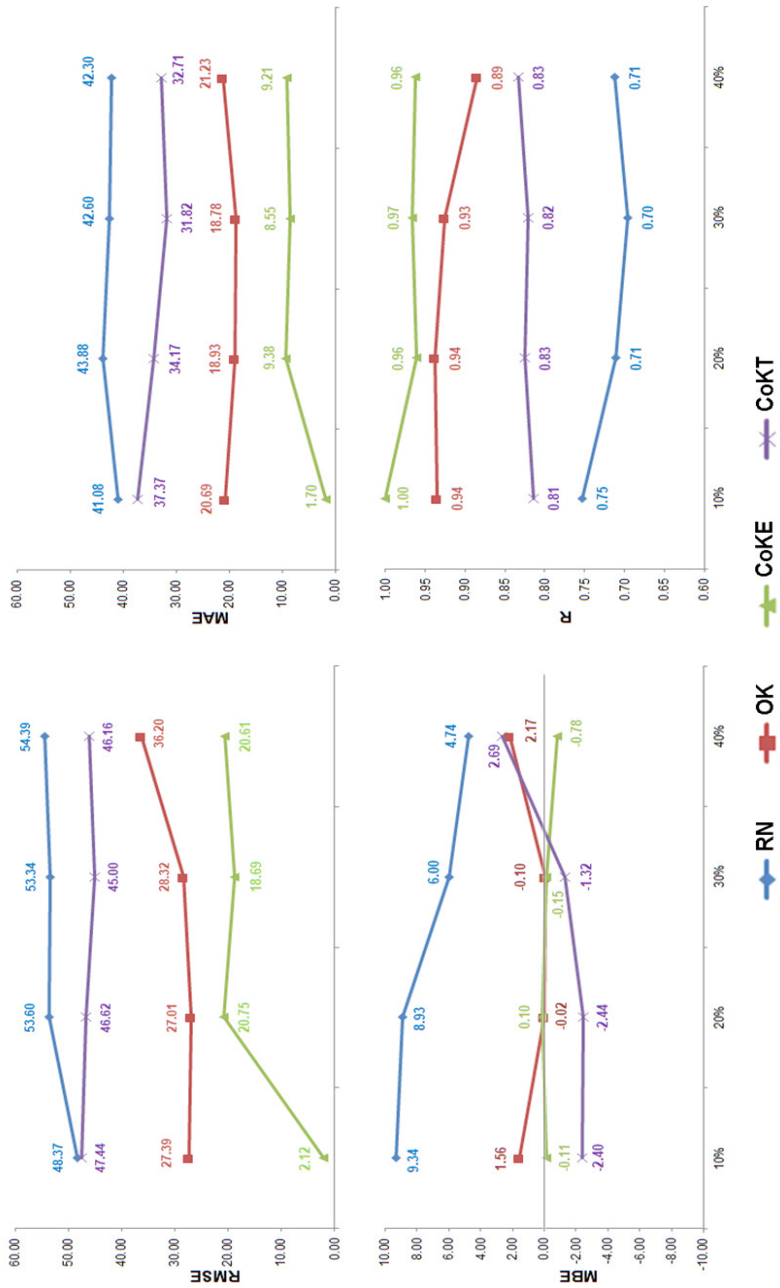


Figura 3. Comportamientos de los índices de ajuste promediados en la Raíz cuadrada del error cuadrático medio (RMSE), Error medio absoluto (MAE), Error de sesgo promedio (MBE) y Coeficiente de correlación de Pearson (R) para los métodos de Relación Normal (RN), Krigado Ordinario (OK), Cokrigado y la elevación (CoKE) y Cokrigado y datos de lluvia satelital TMPA (CoKT) en los cuatro escenarios evaluados.



## Discusión de los resultados

La cuenca del río Coello se encuentre en la cordillera central de los Andes, este relieve proporciona un escenario complejo de análisis en cuanto a la variabilidad espacial y temporal de la lluvia. Las mediciones de la lluvia satelital por TMPA reflejan gran precisión a escala anual (i.e. temporal) de la precipitación, pero presentan dificultades para espacializar la variabilidad de la lluvia en cuencas pequeñas (Álvarez-Villa *et al.*, 2011); en el caso práctico son la fuente alternativa que proporciona información satelital en donde no hay información del monitoreo meteorológico en terreno, pero los resultados de las comparaciones entre métodos de estimación de datos de lluvia (ver Tabla 2 – CoKT) sugieren en conjunto con Conti, Hsu, Noto, y Sorooshian (2014) que los estudios hidrólogos deberían utilizar estos productos en cuencas hidrográficas a mesoescala y orográficamente planas, ya que las estimaciones satelitales no son las adecuadas para representar el clima en áreas orográficas complejas (Serrat-Capdevila, Valdes, & Stakhiv, 2014).

Para estimar los datos faltantes de lluvia se dividieron en cuatro posibles escenarios de comparación, aumentando uno más crítico (i.e. escenario del 40%) que la investigación de Radi *et al.* (2015) pero con la diferencia metodológica de comparaciones entre métodos geoestadísticos. Los resultados de los tres índices de ajuste indican que el RN es superado en gran proporción por los métodos geoestadísticos en los cuatro escenarios evaluados, posiblemente porque el método de RN no contempla la relación entre la lluvia y los patrones de dependencia espacial y requiere una alta densidad de estaciones pluviométricas (Adhikary *et al.*, 2017; Di Piazza *et al.*, 2011; Mair & Fares, 2010), obteniendo resultados opuesto a lo recomendación de Radi *et al.* (2015) en la utilización del RN en la estimación de datos faltantes de lluvia. En la figura 3 se muestra la eficiencia y versatilidad de los métodos OK y CoKE para estimar datos de lluvia en cuencas andinas; nuestros resultados confirman los estudios previos (Adhikary *et al.*, 2017; Di Piazza *et al.*, 2011; Goovaerts, 2000; Mair & Fares, 2010) en donde se destaca que los mejores resultados para estimar datos faltantes de lluvia son más aproximado a la realidad cuando se utilizan los métodos geoestadísticos y la elevación como información auxiliar, demostrando los mejores rendimientos en términos de RMSE, MAE, MBE y R.

En los escenarios estudiados (ver Figura 3) es posible identificar que las eficiencias del método geoestadístico CoKE tiene un comportamiento óptimo y con poco error a los registros observados en el escenario del

10% en los valores de RMSE (2,12), MAE (1,70), MBE (-0,11) y un R con asociación perfecta (1,00), pero consecuente al aumento de datos faltantes los comportamientos muestran un aumento del sesgo de error pero se estabilizan en los escenarios de 20, 30 y 40% de datos faltantes. En general los resultados pueden ser aplicados en diferentes casos críticos en la práctica de los estudios hidrológicos. El método del OK es el segundo mejor comportamiento de los resultados, estos se comportan según los índices de ajuste de una forma estable en los cuatro escenarios estudiados y su sesgo se mantiene entre los rangos de 27,01 a 36,20 milímetros de precipitación en la prueba RMSE. La RN y CoKT tienen los mayores sesgos de error evaluados, el primero puede sobre o subestimar el valor entre 48,37 y 54,39 mm de precipitación estimada, este comportamiento posiblemente es el resultado de la baja densidad de estaciones y la complejidad orográfica de la cuenca de estudio, en el mismo sentido el comportamiento del CoKT presenta valores de RMSE entre 45,00 a 47,44 mm de precipitación estimada en referencia con las mediciones observadas, posiblemente esta poca eficiencia se deba a la escala del área de estudio y a la resolución espacial de los productos de TMPA a escala mensual del producto 3B43V7 con 0,25° de resolución espacial; es necesario corroborar este método geoestadístico con la información auxiliar satelital en un área de estudio de mayor escala y densidad espacial de puntos de medición.

## Conclusiones

Los resultados obtenidos confirman que entre los métodos geoestadísticos comparados, el cokrigado entre los datos históricos de lluvia disponibles y la elevación como variable auxiliar, estima con mejores aproximaciones los valores faltantes de lluvia mensual en cuencas hidrográficas andinas. El CoKE tiene el mejor comportamiento en los cuatro índices de ajuste (RMSE, MAE, MBE y R) cuando solo faltan el 10% de los datos; los resultados indican que en los escenarios de 20, 30 y 40% de datos faltantes, el CoKE estabiliza el sesgo de error de estimación, pero sigue siendo más eficiente que los demás métodos evaluados.

La investigación demuestra que el uso de métodos geoestadísticos empleando variables auxiliares, tiende a mejorar la estimación de datos faltantes de lluvia mensual, por lo que el desarrollo y selección de nuevos algoritmos para la asimilación de información satelital con datos observados en campo, se constituye en una línea de investigación promisoriosa para la estimación de datos faltantes de lluvia.

## Agradecimientos

Los autores agradecen al Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia

(IDEAM) por proveer la información pluviométrica de su red de estaciones y a la Universidad del Tolima (Colombia) por la financiación de esta investigación a través del proyecto 410112.

## Literatura citada

- Adhikary, S. K., Muttill, N., & Yilmaz, A. G.  
2017. Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments. *Hydrological Processes*, 31 (12): 2143-2161.
- Álvarez-Villa, O. D., Vélez, J. I., & Poveda, G.  
2011. Improved long-term mean annual rainfall fields for Colombia. *International Journal of Climatology*, 31 (14): 2194-2212.
- Bayat, B., Nasser, M., & Zahraie, B.  
2015. Identification of long-term annual pattern of meteorological drought based on spatiotemporal methods: evaluation of different geostatistical approaches. *Natural Hazards*, 76 (1): 515-541.
- Conti, F. L., Hsu, K.-L., Noto, L. V., & Sorooshian, S.  
2014. Evaluation and comparison of satellite precipitation estimates with reference to a local area in the Mediterranean Sea. *Atmospheric Research*, 138: 189-204.
- Cruz-Roa, A., Olaya-Marín, E., & Barrios, M.  
2017. Ground and satellite based assessment of meteorological droughts: The Coello river basin case study. *International Journal of Applied Earth Observation and Geoinformation*, 62: 114-121.
- de Amorim Borges, P., Franke, J., da Anunciação, Y. M. T., Weiss, H.; Bernhofer, C.  
2016. Comparison of spatial interpolation methods for the estimation of precipitation distribution in Distrito Federal, Brazil. *Theoretical and Applied Climatology*, 123 (1-2): 335-348.
- De Silva, R., Dayawansa, N., & Ratnasiri, M.  
2007. A comparison of methods used in estimating missing rainfall data. *Journal of agricultural sciences*, 3 (2).
- Di Piazza, A., Conti, F. L., Noto, L., 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.
- Goovaerts, P.  
2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, 228 (1): 113-129.
- Huffman, G., Adler, R., Bolvin, D.; Nelkin, E.  
2010. The TRMM multi-satellite precipitation analysis (TMPA). En: Hossain, F.; Gebremichael, M. (Eds.). *Satellite rainfall applications for surface hydrology*. Springer. Netherlands. pp.3-22.
- Huffman, G., Stocker, E., Bolvin, D., Nelkin, E.; Adler, R.  
2012. TRMM Version 7 3B42 and 3B43 Data Sets. *National Aeronautics and Space Administration, Goddard Space Flight Center*. Greenbelt, Maryland US.
- Mair, A., & Fares, A.  
2010. Comparison of rainfall interpolation methods in a mountainous region of a tropical island. *Journal of Hydrologic Engineering*, 16(4): 371-383.
- Manz, B., Buytaert, W., Zulkafli, Z., Lavado, W., Willems, B., Robles, L. A.; Rodríguez-Sánchez, J. P.  
2016. High-resolution satellite-gauge merged precipitation climatologies of the Tropical Andes. *Journal of Geophysical Research: Atmospheres*, 121 (3): 1190-1207.
- Mejía, F., Mesa, O., Poveda, G., Vélez, J., Hoyos, C., Mantilla, R.; Cuartas, A.  
1999. Distribución espacial y ciclos anual y semianual de la precipitación en Colombia. *Dyna*, 127:7.
- Palomares, F. G., Gil, J. Q. C., & Serra, J. A. M.  
2014. Un laboratorio virtual para la enseñanza del Krigeado. *Modelling in Science Education and Learning*, 7: 73-82.
- Paulhus, J. L., & Kohler, M. A.  
1952. Interpolation of missing precipitation records. *Monthly Weather Review*, 80 (8): 129-133.
- Poveda, G.  
2004. La hidroclimatología de Colombia: una síntesis desde la escala inter-decadal hasta la escala diaria. *Rev. Acad. Colomb. Cienc*, 28(107): 201-222.
- Poveda, G., Álvarez, D. M., & Rueda, Ó. A.  
2011. Hydro-climatic variability over the Andes of Colombia associated with ENSO: a review of climatic processes and their impact on one of the Earth's most important biodiversity hotspots. *Climate Dynamics*, 36 (11-12): 2233-2249.
- Poveda, G., Vélez, J. I., Mesa, O. J., Cuartas, A., Barco, J., Mantilla, R. I., Ceballos, L. I.  
2007. Linking long-term water balances and statistical scaling to estimate river flows along the drainage network of Colombia. *Journal of Hydrologic Engineering*, 12 (1): 4-13.
- Radi, N. F. A., Zakaria, R., & Azman, M. A.-z.  
2015. *Estimation of missing rainfall data using spatial interpolation and imputation methods*. Paper presented at the AIP Conference Proceedings.
- Rogelis, M., & Werner, M.  
2013. Spatial interpolation for real-time rainfall field estimation in areas with complex topography. *Journal of Hydrometeorology*, 14 (1): 85-104.
- Serrat-Capdevila, A., Valdes, J. B., & Stakhiv, E. Z.  
2014. Water management applications for satellite precipitation products: Synthesis and recommendations. *JAWRA Journal of the American Water Resources Association*, 50 (2): 509-525.
- 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.
- Teegavarapu, R. S., & Chandramouli, V.  
2005. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology*, 312 (1): 191-206.

Vélez, J. I., Poveda, G., & Mesa, O.

2000. Balances hidrológicos de Colombia. *Aprovechamiento de Recursos Hidráulicos*. Universidad Nacional de Colombia. Medellín, Colombia. 150p.

