

Imputación de Datos en Series de Precipitación Diaria

Caso de Estudio Cuenca del Río Quindío*

Imputation of Missing Daily Rainfall Data Case Study of Quindio River Basin

Pedro León García Reinoso**

RESUMEN

En este artículo se presentan los resultados obtenidos por la aplicación de cinco técnicas de imputación de datos en series de precipitación diaria para ocho estaciones que tienen afinidad sobre la cuenca del río Quindío, localizada en la zona central colombiana. Con el propósito de preservar la generación de valores de precipitación igual a cero, se consideró el cálculo de probabilidades empíricas a partir de una cadena de Markov de primer orden. Las técnicas fueron implementadas en un algoritmo iterativo, en el que los valores faltantes fueron iniciados con la precipitación promedio diaria. El algoritmo se ejecuta recursivamente sustituyendo los valores estimados en la corrida anterior, finaliza cuando la diferencia máxima entre dos iteraciones sucesivas es menor que un valor fijado previamente. Los datos imputados por la técnica de distancia estadística ponderada conservan adecuadamente las medidas de tendencia central de la serie temporal de precipitación diaria con datos faltantes.

Palabras clave: Precipitación pluvial, Interpolación espacial, Métodos ponderados.

ABSTRACT

This paper shows the results obtained when five methods for imputing missing daily rainfall were applied to records of eight hydrological stations located in the Quindío river basin, on the west-center part of Colombia. With the purpose of preserving the presence of no rainfall data, were considered calculate the empirical probabilities of first-order Markov chains. The five methods were implemented with a recursive algorithm which initializes missing data with the average daily rainfall. After this, the algorithm runs iteratively, replacing the previous run missing data imputations, it runs until the maximum difference between two successive imputations is smaller than a threshold value. Data imputed by the Statistical Measure Weighting Method conserves the measures of central tendency from each station daily rainfall record when it includes missing data.

Key words: Rainfall, Spatial interpolation, Weighting methods.

* Proyecto de Investigación: Estimación de datos faltantes de precipitación diaria en la cuenca del río Quindío. Grupo de Investigación, Desarrollo y Estudio del Recurso Hídrico y el Ambiente –CIDERA–. Línea de Investigación: Calidad del agua en corrientes hídricas. Investigador principal: Pedro León García Reinoso.

** Ingeniero Civil de la Universidad del Quindío (1998). Magíster en Ingeniería Civil con énfasis en Recursos Hídricos de la Universidad de Los Andes (2000). Profesor Asistente del Programa de Ingeniería Civil de la Universidad del Quindío. Investigador Asociado del Grupo de Investigación, Desarrollo y Estudio del Recurso Hídrico y el Ambiente –CIDERA– en la línea de investigación de Calidad del agua en corrientes hídricas. Carrera 15 Calle 12N, Armenia, Quindío. pedrogarcia@uniquindio.edu.co

1. INTRODUCCIÓN

La precipitación pluvial, en adelante precipitación, es una variable climatológica de gran importancia en la hidrología, ya que es insumo fundamental para el cálculo de balances hídricos; la generación de alertas tempranas por riesgo de inundación o sequía; el estudio del régimen ambiental de caudales; y el diseño de sistemas de acueducto y alcantarillado, por mencionar las más importantes. La medición de la precipitación tiene como propósito conocer su distribución en el espacio y en el tiempo. En la escala temporal, el registro de precipitación se estudia como una serie de tiempo horaria, diaria, mensual o anual, y su utilidad dependerá, en gran medida, de su completitud [1].

El análisis hidrológico requiere de series de tiempo continuas que cubran el mayor intervalo de tiempo posible. En las series de tiempo de precipitación es normal localizar vacíos de información que se pueden explicar por la ausencia de lectura, falla del instrumento de registro, error de transcripción, etc. Al iniciar el estudio de una serie de tiempo se debe considerar el tratamiento que se dará a los vacíos de información. Este tratamiento se identifica en la literatura como relleno, completado, llenado, estimación o imputación de datos faltantes. En este artículo se considera adecuado el uso del término imputación, el cual se define desde la estadística como el proceso de sustitución de un único valor razonable para una observación faltante [2]. Aunque existen diferentes técnicas para la imputación de datos, la ejecución de este procedimiento puede poner en tela de juicio la utilidad de la serie de tiempo para ciertos fines [1].

Existen métodos de imputación que consideran el uso de satélites, programas estadísticos y modelos hidrológicos, pero el alto costo ha impedido su uso masivo [3]. Alternativamente, se han desarrollado métodos de imputación más sencillos que ofrecen buenos resultados. Dentro de estos métodos se pueden citar aquellos basados en la estadística, como el análisis de regresión múltiple; otros que emplean técnicas de interpolación espacial, como el inverso de la distancia al cuadrado; y unos que se basan en fórmulas empíricas que se expresan como proporciones respecto de la precipitación registrada, como lo hace el método de la razón normal. La interpolación espacial se considera un enfoque adecuado para la imputación de datos faltantes en series de precipitación diaria [4-8], siendo los métodos ponderados los de mayor uso y aceptación [7].

El objetivo de esta investigación es determinar la confiabilidad de cinco técnicas de imputación para datos diarios de precipitación, estando dichas técnicas inscritas dentro de los métodos ponderados. Estas técnicas corresponden a Distancia inversa ponderada [4-8], Coeficiente de correlación ponderado [7], Exponencial inverso ponderado [7], Medida estadística ponderada [6] y Radio normal ponderado [8]. Dado que la imputación de datos puede poner en tela de juicio su utilidad para ciertos fines, la operación debería ser constatable de manera que la presencia de datos imputados sea discernible y el proceso pueda ser revertido [1]. Con este propósito se determinaron las diferencias entre estadísticos de la serie con datos faltantes y la serie con datos imputados. En consideración de estos elementos, se discute el desempeño de cada técnica para completar datos de precipitación diaria en ocho estaciones que cubren un periodo de 34 años y tienen aferencia sobre la cuenca del río Quindío.

2. METODOLOGÍA

2.1. Área de estudio y datos

En esta investigación, la cuenca del río Quindío fue elegida como el área de estudio. Su centroide está localizado en $4^{\circ} 31' 49''$ N- $75^{\circ} 35' 35''$ W, región centro occidental de Colombia conocida como el Triángulo del Café. Con un área de $718,87 \text{ km}^2$, la mayor parte de la cuenca drena la vertiente occidental de la Cordillera Central y es atravesada por el río Quindío, el cual conduce las aguas recolectadas desde el nevado El Quindío (3780 msnm) hasta la confluencia con el río Barragán (1040 msnm) en el Valle de Maravélez, en un recorrido de 65.35 km [9]. La cuenca del río Quindío está completamente contenida en el departamento del Quindío (Figura 1).



Figura 1. Localización de las estaciones de precipitación en la cuenca del río Quindío

Fuente: Elaboración propia

Los datos de precipitación fueron obtenidos del Boletín Meteorológico CRQ 2005 [10] y se procesaron en dos etapas. En la primera etapa se realizó la construcción de un archivo único con los datos de precipitación diaria de la red de estaciones administrada por la Corporación Autónoma Regional del Quindío. En la segunda etapa se ejecutó la depuración y control de calidad de los datos, siguiendo procedimientos avalados por el Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM). Luego de considerar estaciones activas, área de aferencia y longitud del registro, fueron elegidas las ocho estaciones que se ilustran en la Figura 1 para el periodo comprendido entre el 01 de abril de 1971

y el 31 de diciembre de 2005. En la Tabla 1 se relacionan elementos que describen la localización física de las estaciones y propiedades estadísticas del registro existente. Es de resaltar la presencia de largos periodos con ausencia de datos que se explican tanto por la falla del instrumento de registro como por la ausencia de lectura explicada por la suspensión temporal de la estación, situaciones que son comunes en el contexto colombiano y dan origen al desarrollo de este tipo de investigación.

Tabla 1. Características de las estaciones seleccionadas

Nombre de la Estación	Número de la Estación	Latitud	Longitud	Altitud	Aferencia sobre la cuenca	Precipitación promedio anual	Desviación estándar anual	Precipitación promedio diaria	Desviación estándar diaria	Porcentaje de datos válidos	Porcentajes de días con precipitación
		°	°	msnm	(%)	mm	mm	mm	mm	(%) ^a	(%)
Bremen	1	04° 40' 00"	75° 36' 00"	2040	18,20	2935	721	8,12	13,70	94,56 (304)	61,97
La Picota	2	04° 39' 00"	75° 28' 00"	2680	20,69	1409	307	3,87	7,80	86,67 (972)	53,47
CRQ	3	04° 33' 20"	75° 39' 50"	1550	8,62	2433	513	6,70	12,90	90,81 (424)	53,27
Planadas	4	04° 29' 00"	75° 38' 00"	2350	19,36	2016	658	5,39	12,78	92,21 (549)	36,96
Navarco	5	04° 29' 02"	75° 33' 18"	2860	14,70	2000	414	5,52	10,98	91,99 (396)	54,90
Vivero Tebaida	6	04° 26' 57"	75° 47' 21"	1220	0,56	1905	423	5,33	11,21	88,95 (488)	41,38
Centro Guadua	7	04° 24' 07"	75° 42' 57"	1212	17,62	2104	639	5,84	12,37	89,07 (1096)	48,32
Barragán	8	04° 20' 05"	75° 47' 30"	1180	0,25	1354	708	3,86	10,23	51,79 (2628)	30,55

a: El número dentro del paréntesis indica el intervalo más largo de datos faltantes
Fuente: Elaboración propia

Se considera apropiado completar series temporales de precipitación diaria que tengan máximo un 20 % de datos faltantes, aunque un estudio realizado en los Andes ecuatorianos obtuvo buenos resultados al considerar series de tiempo que incluyeron hasta un 26 % de datos faltantes [11]. Utilizando métodos basados en inteligencia artificial y experimentos controlados, han sido imputadas series de precipitación hasta con un 80 % de datos faltantes [12]. Aunque la Estación Barragán supera estos límites, ha sido elegida para imputar los datos faltantes con el propósito de observar el comportamiento de las técnicas implementadas en este artículo y construir las bases que soporten el desarrollo de trabajos futuros que involucren otras metodologías de imputación.

2.2. Métodos de imputación

Es posible clasificar los métodos de imputación en tres categorías, a saber: i) los métodos determinísticos, aquellos que consideran un modelo matemático que produce una respuesta única para ciertas condiciones iniciales; ii) los métodos estocásticos, que ofrecen una estimación probabilística para el dato imputado, y iii) los métodos de inteligencia artificial, basados en modelos matemáticos complejos, con un inherente alto costo computacional. Por ser una aproximación inicial al problema de imputación en el área de estudio, por las características de los datos y las ventajas que ofrece el método, en esta investigación se implementaron métodos de interpolación ponderados, los cuales se clasifican dentro de los métodos determinísticos. Además, el método elegido es aplicable dado que los datos disponibles satisfacen el supuesto de que exista una correlación espacial entre la cantidad y ocurrencia de precipitación entre estaciones vecinas, situación que es soportada por lo ilustrado en la Figura 2, la cual indica la relación negativa que existe entre la correlación de las cantidades de precipitación y la distancia entre las estaciones.

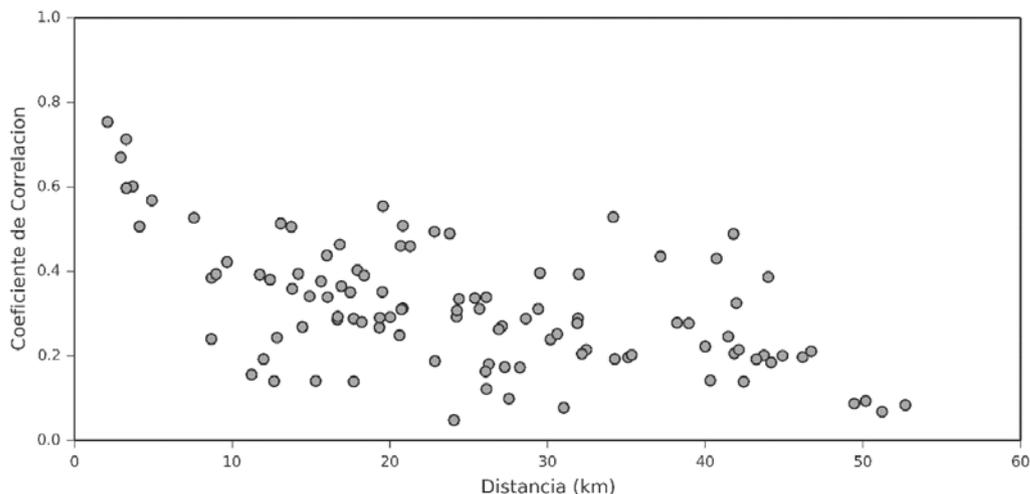


Figura 2. Relación coeficiente de correlación contra distancia (km)

Fuente: Elaboración propia

Teniendo en cuenta la precipitación acumulada sobre un intervalo de tiempo dado, la precipitación local que ha caído en un punto es estimada usando un promedio ponderado de los valores de precipitación puntual que ha sido registrada en estaciones localizadas cerca al punto considerado [13]:

$$P_M^*(t) = \frac{\sum_{i=1}^n \lambda_{M,i} P_i(t)}{\sum_{i=1}^n \lambda_{M,i}} \quad (1)$$

Donde $\lambda_{M,i}$ es el peso asociado con la estación. El método más simple de interpolación ponderada es asignar el mismo peso a todas las estaciones vecinas, lo que es equivalente al cálculo del promedio aritmético. Este método podría usarse en casos donde la precipitación y las estaciones de registro están uniformemente distribuidas sobre el área de estudio [13]. En esta investigación se eligieron la distancia inversa ponderada, el coeficiente de correlación ponderado, el exponencial inverso ponderado, la medida estadística ponderada y el radio normal ponderado, como técnicas de interpolación para la imputación de datos faltantes. La aplicación de estos métodos requiere, además de la información de precipitación, la localización geográfica de las estaciones (Tabla 1) y la distancia que existe entre ellas (Tabla 2).

Tabla 2. Matriz de distancias entre estaciones, en kilómetros

	Planadas	Bremen	CRQ	Vivero Tebaida	Barragán	Centro Guadua	Navarco	La Picota
Planadas	---	20.615	8.679	17.708	24.067	12.843	8.697	26.119
Bremen	20.615	---	14.190	31.933	42.437	31.981	20.827	14.914
CRQ	8.679	14.190	---	18.218	28.247	17.943	14.456	24.257
Vivero Tebaida	17.708	31.933	18.218	---	12.662	9.673	26.279	42.132
Barragán	24.067	42.437	28.247	12.662	---	11.233	31.027	50.180
Centro Guadua	12.843	31.981	17.943	9.673	11.233	---	20.025	38.962
Navarco	8.697	20.827	14.456	26.279	31.027	20.025	---	20.828
La Picota	26.119	14.914	24.257	42.132	50.180	38.962	20.828	---

Fuente: Elaboración propia

Uno de los problemas que enfrenta la imputación de datos mediante técnicas de interpolación es la sobrestimación de datos con precipitación igual a cero [14]. Con el propósito de preservar esta característica, se realizó el cálculo de la probabilidad empírica de ocurrencia de días lluviosos (R) y secos (D) en una cadena de Markov de primer orden. En la cadena de primer orden se presentan cuatro eventos: RR, RD, DD, DR. A partir de los datos de precipitación se obtiene la frecuencia de ocurrencia de los eventos recién descritos, permitiendo el cálculo de la probabilidad de ocurrencia empírica como el cociente de la frecuencia de ocurrencia entre el número de casos. De esta forma, para cada día del año se obtiene una probabilidad de ocurrencia empírica que debe ser consultada antes de efectuar cualquier cálculo, si la probabilidad de ocurrencia de un día lluvioso (RR, DR) es superior a cierto umbral se realiza la imputación del dato utilizando una técnica de interpolación ponderada, en caso contrario el valor imputado es cero.

Para implementar las técnicas de imputación seleccionadas se requirió el desarrollo del algoritmo iterativo que se ilustra en el diagrama de flujo presentado en la Figura 3. El proceso iterativo tiene el propósito de establecer valores razonables para la probabilidad de ocurrencia de un día lluvioso y los datos imputados. El algoritmo se inicia reemplazando los datos faltantes con la precipitación promedio diaria de la respectiva estación. El algoritmo es ejecutado recursivamente, sustituyendo los valores estimados en la corrida anterior hasta alcanzar la diferencia máxima de la función objetivo entre dos iteraciones sucesivas. Para cada estación, la función objetivo es la diferencia del valor promedio diario entre dos iteraciones sucesivas.

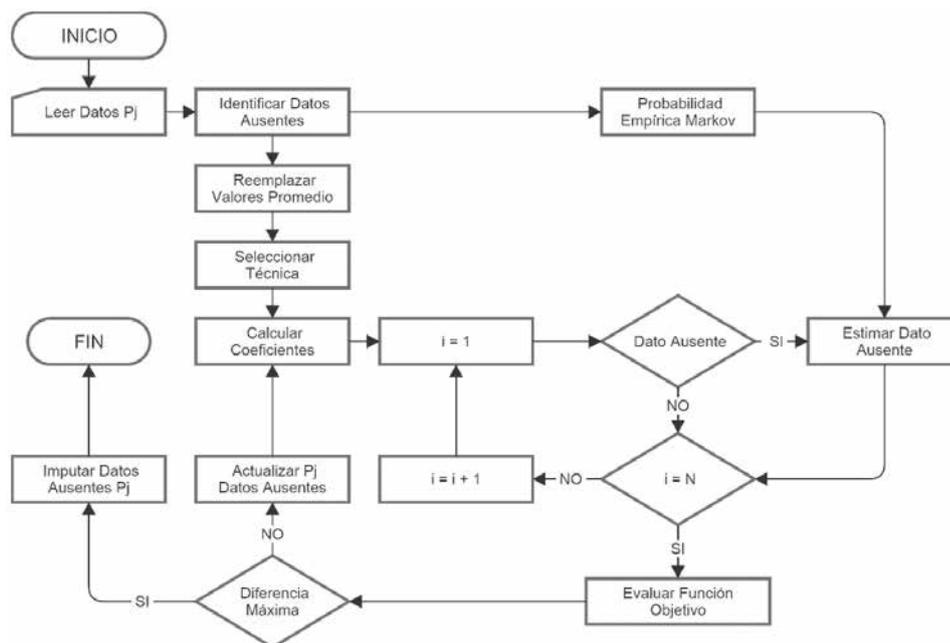


Figura 3. Diagrama de flujo para la imputación de datos de precipitación

Fuente: Elaboración propia

2.2.1. Distancia inversa ponderada

El método de la distancia inversa ponderada por su simplicidad es el más ampliamente usado ya que se obtienen buenos resultados cuando existe una fuerte autocorrelación espacial [7]. Los pesos para cada estación se calculan según,

$$\lambda_{M,i} = d_{M,i}^{-k} \quad (2)$$

Donde $d_{M,i}$ es la distancia desde la estación M hasta la estación i y es referenciada como una distancia de fricción que se ubica en un intervalo que va de 1,0 a 6,0 [15]. El valor comúnmente usado para k es 2 [7]. Una condición importante para la aplicación de la técnica es la selección del radio de influencia, el cual se recomienda sea menor a 100 km [8]. Dado que este valor supera en cerca de dos veces el mayor valor presentado en la Tabla 2, no constituye limitante para la aplicación del método.

2.2.2. Coeficiente de correlación ponderado

Al igual que el método de la distancia inversa ponderada, los resultados dependen de la existencia de autocorrelación espacial. La existencia de esta autocorrelación es calculada mediante el coeficiente de correlación calculado para dos conjuntos de datos registrados en dos localizaciones diferentes. El método asume el peso como [7],

$$\lambda_{M,i} = r_{Mi} \quad (3)$$

Donde r_{Mi} es el coeficiente de correlación calculado usando los datos de la estación M y de la estación i . La aplicación de este método exige la disponibilidad de registros históricos de precipitación como los relacionados en la Tabla 1.

2.2.3. Exponencial inverso ponderado

Es un método empleado en geografía para la generación de superficies, en el cual una función exponencial negativa es considerada en el cálculo del peso [7],

$$\lambda_{M,i} = e^{-kd} \quad (4)$$

Por ser un método estrechamente vinculado con el método de la distancia inversa ponderada, d representa la distancia entre las estaciones localizadas en los puntos M e i , además se mantiene el supuesto de $k = 2$.

2.2.4. Medida estadística ponderada

Al igual que el método del coeficiente de correlación ponderado, el método de la medida estadística ponderada requiere de registros históricos de precipitación en los puntos de observación e imputación.

El método reemplaza la distancia geográfica que existe entre las estaciones por una distancia estadística calculada entre los datos de precipitación de cada una de las estaciones. La formulación matemática para el peso es [6],

$$\gamma_{Mi} = \frac{1}{2T} \sum_{t=1}^T (P_{Mt} - P_{it})^2 \quad (5)$$

Donde T es la longitud de la serie de tiempo y $P_{Mt} - P_{it}$ son los valores de precipitación en los puntos M e i , respectivamente. En el cálculo del peso solamente se consideran días con precipitación en los puntos de observación e imputación.

2.2.5. Radio normal ponderado

Es un método de interpolación espacial formulado en los años cincuenta del siglo pasado que ha sufrido mejoras con el tiempo [8]. El peso es calculado según,

$$\gamma_{Mi} = r_{Mi}^2 [(T - 2)/(1 - r_{Mi}^2)] \quad (6)$$

2.3. Evaluación de las técnicas de interpolación

En una etapa posterior, los resultados de esta investigación serán utilizados en la implementación de un modelo lluvia-escorrentía con paso temporal diario, por lo que resulta necesario aprovechar racionalmente la información disponible. Por esta razón, principalmente, la investigación no consideró el enfoque tradicional en el que se ejecuta una generación aleatoria de vacíos que posteriormente son considerados como faltantes, permitiendo establecer las comparaciones entre la precipitación real y la imputada. Puesto que los vacíos de información son reales y se cumplen las condiciones mínimas para la implementación de métodos de interpolación ponderados, se evaluará el desempeño del método a través de la similitud entre la serie de tiempo original y la serie de tiempo imputada por la comparación de estadísticos de tendencia central (media y mediana), estadísticos de dispersión (desviación estándar y coeficiente de variación), estadísticos de posición (deciles) y estadísticos de forma (coeficiente de asimetría, coeficiente de Curtosis). Adicionalmente, la consistencia de los registros imputados fue verificada con una Curva de Doble Masa que utilizó el registro de la Estación Aeropuerto El Edén, administrada por el IDEAM y con las siguientes características:

- Fecha de instalación: 01 de abril de 1949
- Fecha de suspensión: Operando en la actualidad
- Intervalo mayor de datos faltantes: 273 (01/01/1955-30/09/1955)
- Porcentaje de datos faltantes en el periodo de registro: 2,63 %
- Porcentaje de datos faltantes entre 01/04/1971 y 31/12/2005: 0,00 %

3. RESULTADOS Y DISCUSIÓN

La aplicación y evaluación de los cinco métodos de imputación presentados en esta investigación fue

realizada sobre los vacíos de información que se relacionan en la Tabla 1. El algoritmo ilustrado en la Figura 3 fue desarrollado en Python 2.7 [16], incluyendo la lectura de la base de datos del Boletín Meteorológico CRQ 2005 [10] y estableciendo una diferencia máxima admisible de 0,00001 mm para la función objetivo. La definición de una probabilidad empírica para la ocurrencia de un día lluvioso demandó la construcción de un intervalo para la búsqueda del valor óptimo. En la Figura 4 se puede observar el comportamiento de los eventos RR vs. DD (izquierda) y RD vs. DR (derecha). En la imagen izquierda se observa un comportamiento reflejado, que se explica por la alternancia presente a lo largo del año entre temporadas seca y húmeda. En la imagen derecha las probabilidades se superponen manteniendo valores dentro de límites que permanecen constantes durante el año. Considerando que los cuatro eventos presentan una probabilidad de ocurrencia promedio del 25 % y que para el propósito de la investigación interesan los eventos RR-DR, se definió el intervalo 15 %-30 % para orientar la búsqueda del valor óptimo, el cual resultó ser del 20 %.

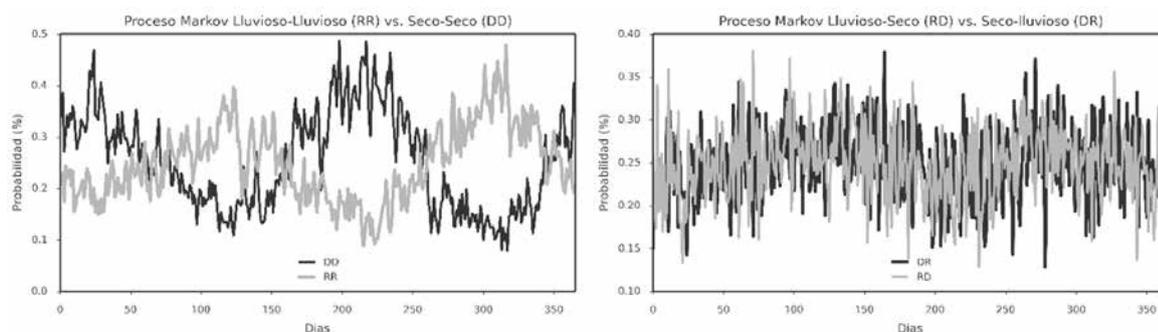


Figura 4. Probabilidades empíricas del proceso de Markov de primer orden

Fuente: Elaboración propia

El desempeño de cada uno de los métodos de imputación fue evaluado considerando la capacidad que tiene el método para preservar el valor medio diario de la serie original con un intervalo de confianza del 95 %. En la Tabla 3 se presentan los resultados para el valor medio diario obtenido en cada estación, los valores rechazados se encuentran en subrayados en negrita. En la Estación Barragán, con el mayor número de datos faltantes, son rechazados cuatro de los cinco métodos, aceptando tan solo el valor generado por el método de la Medida Estadística Ponderada. Los métodos de la Distancia Inversa Ponderada y el Exponencial Inverso Ponderado son rechazados en las Estaciones Barragán y La Picota, en las estaciones restantes –con datos faltantes menores al 11 %– no se rechaza el valor obtenido por los cinco métodos de imputación. Dado que el valor medio diario de precipitación resulta muy similar entre los diferentes métodos, se elige el método de la Medida Estadística Ponderada para realizar un análisis más detallado de los resultados.

Tabla 3. Comparación de los diferentes métodos de imputación usando la precipitación media diaria

	Estaciones en el área de estudio							
	Planadas (7,79 %)	Bremen (5,44 %)	CRQ (9,19 %)	Tebaida (11,05 %)	Barragán (48,21 %)	Guadua (10,93 %)	Navarco (8,01 %)	Picota (13,33 %)
Datos originales	5,39	8,12	6,70	5,33	3,86	5,84	5,52	3,87
Distancia inversa ponderada	5,39	7,93	6,53	5,32	4,15	5,65	5,46	4,08
Coefficiente correlación ponderado	5,38	7,92	6,51	5,25	4,17	5,70	5,46	4,01
Exponencial inverso ponderado	5,41	7,98	6,51	5,39	4,17	5,61	5,42	4,31
Medida estadística ponderada	5,38	7,91	6,50	5,22	4,03	5,67	5,45	3,97
Radio normal ponderado	5,39	7,93	6,54	5,29	4,22	5,72	5,46	4,01

a: El número dentro del paréntesis indica el porcentaje de datos faltantes

Fuente: Elaboración propia

Tabla 4. Indicadores de desempeño para la serie original (Ori.) e Imputada (Imp.)

Estaciones Estadísticas	Bremen		La Picota		CRQ		Planadas		Navarco		Vivero Tebaida		Centro Guadua		Barragán	
	Ori.	Imp.	Ori.	Imp.	Ori.	Imp.	Ori.	Imp.	Ori.	Imp.	Ori.	Imp.	Ori.	Imp.	Ori.	Imp.
Porcentaje Días No Lluvia	38,0 %	37,3 %	46,5 %	43,6 %	46,7 %	44,7 %	63,0 %	59,8 %	45,1 %	43,3 %	58,6 %	55,2 %	51,7 %	48,8 %	69,4 %	54,2 %
Media	8,12	7,91	3,87	3,97	6,70	6,50	5,39	5,38	5,52	5,45	5,33	5,22	5,84	5,67	3,86	4,03
Mediana	1,50	1,50	0,30	0,50	0,30	0,40	0,00	0,00	0,60	0,60	0,00	0,00	0,00	0,00	0,00	0,00
Desviación estándar	13,70	13,45	7,80	7,77	12,90	12,52	12,78	12,46	10,98	10,73	11,21	10,82	12,37	11,95	10,23	9,17
Coefficiente de Variación	1,69	1,70	2,02	1,95	1,93	1,93	2,37	2,32	1,99	1,97	2,10	2,07	2,12	2,11	2,65	2,27
Coefficiente de asimetría	2,67	2,73	4,14	3,95	2,94	3,01	3,85	3,89	4,23	4,25	3,13	3,19	3,31	3,39	4,07	4,05
Curtosis	9,35	9,82	26,61	24,45	10,86	11,56	19,11	19,79	29,27	30,00	11,88	12,62	13,69	14,56	21,41	23,02
Nivel de confianza (95 %)	0,24	-	0,14	-	0,23	-	0,23	-	0,20	-	0,21	-	0,23	-	0,22	-

Fuente: Elaboración propia

En la Tabla 4 se encuentra relacionado el resumen de los resultados obtenidos con el método de la Medida Estadística Ponderada. Con una probabilidad empírica del 20 % para la ocurrencia de un día lluvioso, la serie imputada reproduce adecuadamente el porcentaje de días de no lluvia en las estaciones Bremen, La Picota, CRQ, Planadas, Navarco, Vivero Tebaida y Centro Guadua, alcanzando una diferencia máxima del 3,44 %. En la Estación Barragán se obtiene una diferencia del 15,25 %. En todas las estaciones, los estadísticos de tendencia central resultan similares entre la serie original y la imputada. Sin embargo, los estadísticos de dispersión permiten diferenciar dos grupos de estaciones, a saber: i) conformado por las estaciones localizadas por encima de los 1500 msnm, se caracteriza por presentar pequeñas diferencias en la desviación estándar; ii) congrega las estaciones Vivero Tebaida, Centro Guadua y Barragán, todas ubicadas por debajo de los 1500 msnm y mayor diferencia en la desviación estándar.

Para estudiar el desempeño con respecto a los estadísticos de posición se eligieron los deciles, su comportamiento se encuentra ilustrado en la Figura 5, la cual permite inferir que la serie imputada no presenta diferencias significativas con respecto a la serie original en siete de las ocho estaciones. En

la estación Barragán, al presentarse un valor para D_6 y exhibir diferencias notables en D_7 y D_8 , la serie imputada emula un comportamiento similar al de las estaciones Vivero Tebaida y Centro Guadua. Los estadísticos de forma indican que la serie imputada mantiene una distribución asimétrica positiva con mayor concentración de valores alrededor de la precipitación media. Considerando el desempeño de los estadísticos, se colige que es adecuada la representación que hace el método de la Medida Estadística Ponderada aunque el porcentaje de datos faltantes sea mayor al 20 %.

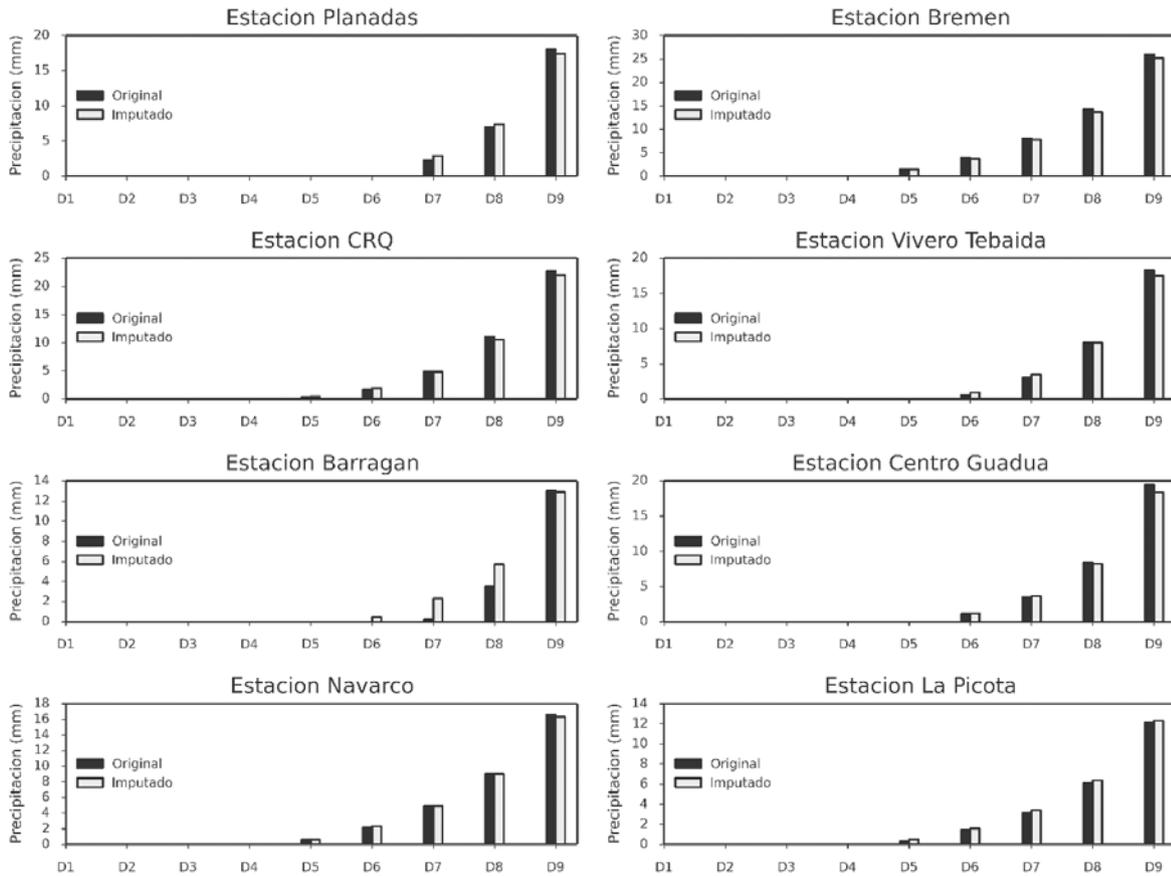


Figura 5. Deciles para las estaciones en estudio

Fuente: Elaboración propia

De acuerdo con las Curvas de Doble Masa, presentadas en la Figura 6, los datos imputados por el método de la Medida Estadística Ponderada resultan consistentes en las estaciones Planadas, Bremen, CRQ, Vivero Tebaida, Centro Guadua, Navarco y La Picota. Aunque la Curva de Doble Masa de la Estación Barragán no describe el comportamiento lineal que exhiben registros consistentes, el cambio de pendiente que se observa en la parte final se da entre 01/06/1994-01/02/2005, para un total de 2008 datos, de los cuales 1482 representan días de no lluvia y 64 son datos faltantes. En ese mismo periodo la Estación Aeropuerto Edén presenta 906 días de no lluvia. Por lo tanto, el cambio de pendiente no es

atribuible exclusivamente a los valores imputados ya que el 94,1 % de estos se ubica entre 01/04/1971-01/06/1994, que corresponde con el tramo inicial y luce una pendiente similar a la del tramo final. La revisión detallada, en el periodo indicado, al registro de la estación Barragán resulta entonces necesaria.

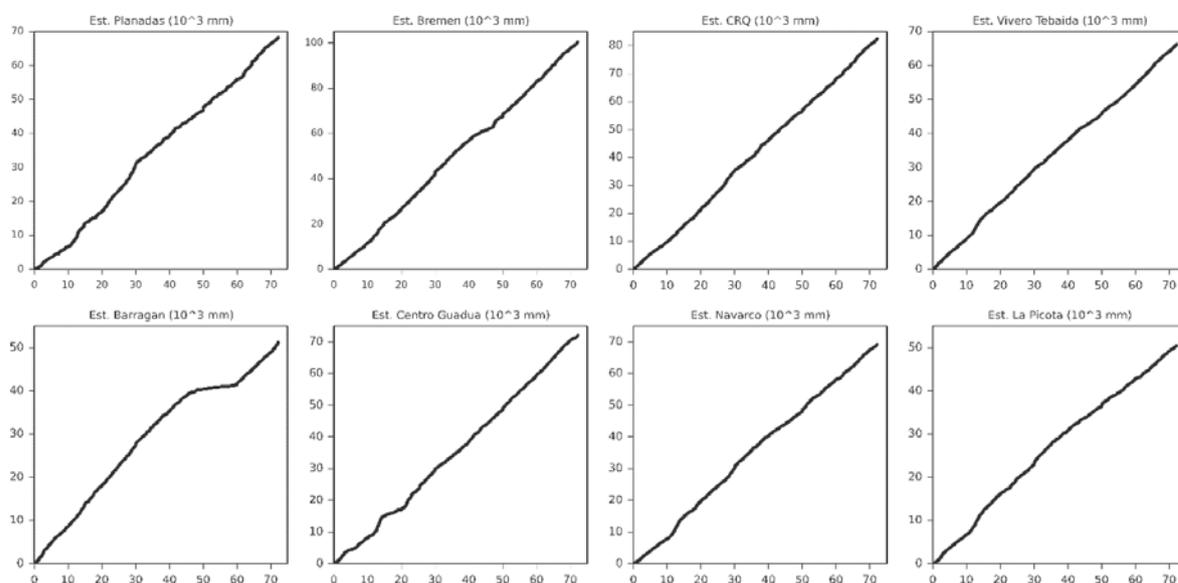


Figura 6. Curvas de doble masa para las estaciones en estudio

Fuente: Elaboración propia

4. CONCLUSIONES

La imputación de datos diarios de precipitación con cinco métodos de interpolación ponderados en ocho estaciones con aferencia en la cuenca del río Quindío ha sido investigada. Luego de comparar los estadísticos de la serie original con la serie imputada, se ha demostrado el buen desempeño del método de la Medida Estadística Ponderada, aunque resultados similares se obtienen de los métodos de la Distancia Inversa Ponderada y el Radio Normal Ponderado (no presentados). Por su desempeño, se ratifica la ventaja que ofrece el Método de la Medida Estadística Ponderada al aplicarse en áreas montañosas donde es importante considerar el efecto orográfico sobre la precipitación, condición implícitamente definida en el método [6].

Las series imputadas representan adecuadamente los estadísticos para registros con datos faltantes inferiores al 20 %, aunque el comportamiento de la serie imputada en la Estación Barragán (48,21 % de datos faltantes) plantea la posibilidad de considerar su uso, es necesario revisar la calidad del registro original y explorar otros métodos de imputación que incluyan la ventaja que ofrece la aparente agrupación que surge entre las estaciones Barragán, Vivero Tebaida y Centro Guadua, todas ellas localizadas en altitudes inferiores a los 1500 msnm. En estas condiciones, el trabajo futuro se podría orientar hacia

el estudio de métodos de inteligencia artificial como las redes neuronales y los árboles de regresión [5].

Se recomienda, siempre que sea necesario imputar datos faltantes, involucrar estaciones cercanas con registros debidamente procesados, debido a que el mejor resultado obtenido está asociado con estaciones procesadas conforme con los protocolos del IDEAM, con datos faltantes inferiores al 20 % y localizadas en la misma subzona hidrológica [9]. Dado que la metodología presentada podría ser empleada por investigadores, docentes, profesionales y estudiantes cuyo accionar se da en el área de los recursos hídricos, se debe enfatizar que la decisión de conservar o no los datos imputados dependerá del juicio experto del usuario, el cual se complementa con el análisis de los resultados desde un punto de vista estadístico, el estudio de factores geográficos y la densidad de la red de monitoreo asociada con el área de estudio bajo consideración. Estos elementos influyen directamente en la exactitud de los resultados y dificultan su comparación directa con trabajos similares.

Agradecimientos

El autor agradece a la Corporación Autónoma Regional del Quindío y al Instituto de Hidrología, Meteorología y Estudios Ambientales por suministrar la información necesaria para llevar a buen término esta investigación.

5. REFERENCIAS BIBLIOGRÁFICAS

- [1] Organización Meteorológica Mundial, *Guía de prácticas hidrológicas*. Volumen I. *Hidrología – De la medición a la información hidrológica*. OMN-Nº 168, 6ta. ed. Ginebra: Organización Meteorológica Mundial, 2011.
- [2] D. L. Fairclough, *Design and Analysis of Quality of Life Studies in Clinical Trials*, 2da. ed. Boca Ratón: CRC Press, 2010.
- [3] R.P. Tapia, P.A. Tapia, D.A. Garrido and C.S. Pool, *Evaluación de métodos hidrológicos para la completación de datos faltantes de precipitación en estaciones pluviográficas de la VII región del Maule, Chile*. 2008. [Online]. Disponible: http://ctha.utralca.cl/Docs/pdf/Publicaciones/articulos_cientificos/pizarro_ausensi_aravena_sanguesa.pdf
- [4] F.-W. Chen and C.-W. Liu, "Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan", *Paddy and Water Environment*, vol.10, n°. 3, pp. 209-222, 2012.
- [5] J.-W. Kim and Y.A. Pachepsky, "Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation", *Journal of Hydrology*, n°. 394, pp. 305-314, 2010.
- [6] B. Ahrens. "Distance in spatial interpolation of daily rain gauge data", *Hydrol. Earth Syst. Sci*, n°. 10, pp. 197-208, 2006.
- [7] R.S.V. Teegavarapu and V. Chandramouli, "Improved weighting methods, deterministic and sto-