

Relleno de datos diarios faltantes en registros de series climatológicas temporales

Jesica Natalia Valencia-González^{1§}
Ramón Arteaga-Ramírez¹
Mario Alberto Vásquez-Peña¹
Abel Quevedo-Nolasco²

¹Posgrado en Ingeniería Agrícola y Uso Integral del Agua-Universidad Autónoma Chapingo. Carretera México-Texcoco km 38.5, Texcoco, Estado de México, México. CP. 56230. (nvg.912@gmail.com, mvazquezd@correo.chapingo.mx). ²Colegio de Posgraduados. Carrera México-Texcoco km 36.5, Montecillo, Texcoco, Estado de México, México. CP. 56230. (anolasco@colpos.mx).

§Autor para correspondencia: rarteagar@taurus.chapingo.mx.

Resumen

La falta de datos diarios en las estaciones meteorológicas es frecuente y esto no permite que se utilicen las series en estudios agroclimáticos. Con lo anterior no se conoce la variación temporal y espacial de las variables que integran el agroclima de una región. El objetivo de este trabajo fue estimar y verificar mediante los métodos: razón normal, serie de Fourier y cuadrado del inverso de la distancia, el método con menor error para el relleno de datos diarios faltantes de las variables precipitación, brillo solar, evaporación, temperatura máxima, temperatura mínima y humedad relativa de las estaciones climatológicas circundantes a la zona de producción de arroz en el departamento del Valle del Cauca, Colombia. Se analizaron nueve estaciones, las cuales no presentan distancias mayores a 50 km, ni diferencias altitudinales de más de 750 m. Estas fueron empleadas con diferentes periodos de estudio de acuerdo con la variable en curso y se evaluaron los métodos con los índices estadísticos raíz cuadrada del cuadrado medio del error y el coeficiente de determinación, el primero permitió conocer el valor máximo admisible de error y el segundo, el nivel de ajuste entre los valores observados y los estimados. Por lo tanto, estos permitieron inferir que la variable brillo solar y evaporación obtuvieron los mejores resultados con la razón normal; la temperatura mínima y humedad relativa con la serie de Fourier y el cuadrado del inverso de la distancia para la precipitación y temperatura máxima.

Palabras clave: cuadrado del inverso de la distancia, datos faltantes, razón normal, serie de Fourier.

Recibido: marzo de 2022

Aceptado: mayo de 2022

Introducción

El análisis climático requiere de series temporales con datos meteorológicos continuos, homogéneos y que contengan el máximo intervalo temporal posible (Nauman y Vargas, 2012). La ausencia de datos en los registros meteorológicos es común debido a varias situaciones, como lo son: la sustitución del operador, fallos de los dispositivos de registro o a negligencias operativas, lo que limita la realización de estudios agroclimáticos (Gallegos *et al.*, 2016). Por consiguiente, Medina *et al.* (2008), proponen que se debe intentar completar las bases de datos, con datos generados por alguna metodología estadística que garantice la aleatoriedad de estos, y que se ajusten a la estructura del proceso estocástico subyacente.

Desde el punto de vista de Nauman y Vargas (2012) considera como estación o serie de referencia a aquella que dispone de un amplio rango de registros (en lo posible debe abarcar el período instrumental), además éstas deben representar regiones climáticas distintas o particularizadas. Es necesario efectuar el diagnóstico sobre estaciones climatológicas de referencia, ya que la estabilidad de los resultados está condicionada a la mayor disponibilidad de información. Dicho por Mintegui y López (1990), la estimación de datos faltantes ha generado métodos que permiten completar las bases de datos climatológicas, donde las variables en curso independientemente de su base teórica responden a la estructura de un estudio estadístico.

La razón normal ha sido un método útil para complementar datos faltantes de una serie a partir de los datos de las series de n estaciones vecinas y contemporáneas, que presenten un alto grado de correlación con la serie a completar (Paulhus y Kohler, 1952). De igual manera, el análisis armónico mediante las series de Fourier ha sido empleado para estimar series diarias de temperatura máxima y mínima, presión, evaporación y humedad relativa, ya que estos elementos meteorológicos, se comportan como una función periódica representándose como la suma de senos y cosenos asignándole a cada uno un coeficiente de ponderación, cuyo tratamiento ha tolerado adecuar procesos de transmisión, compresión y reconstrucción de la información requerida (Cortés *et al.*, 2007).

El cuadrado del inverso de la distancia (IDW) es de los métodos más utilizados en la hidrología, climatología y meteorología, puesto que, la superficie que se interpola debe ser la de una variable dependiente de la ubicación y así estimar valores en puntos desconocidos, con la distancia entre las estaciones (Eischeid *et al.*, 2000). El objetivo de este trabajo fue estimar y verificar mediante los métodos: razón normal (RN), serie de Fourier (SF) y cuadrado del inverso de la distancia (IDW), el método con menor error para el relleno de los datos diarios ausentes de las variables precipitación, brillo solar, evaporación, temperatura máxima, temperatura mínima y humedad relativa de las estaciones climatológicas circundantes a la zona de producción de arroz en el sur del departamento del Valle del Cauca, Colombia.

Materiales y métodos

Se realizó una revisión bibliométrica de los artículos científicos en la base de datos Scopus sobre los métodos de relleno de datos faltantes: ‘razón normal, serie de Fourier y el cuadrado del inverso de la distancia’. Esta biblioteca fue elegida por tener una amplia variedad de revistas científicas de calidad, así como un metabuscador que facilita la evaluación de la producción científica por diversos filtros (Biblioteca Scopus, 2019).

Localización del área de estudio

La producción de arroz en el Departamento del Valle del Cauca (Colombia) acoge los municipios de Jamundí, Quinamayo, Timba, Potrerito y Villa Paz que limitan con el departamento del Cauca (Figura 1).

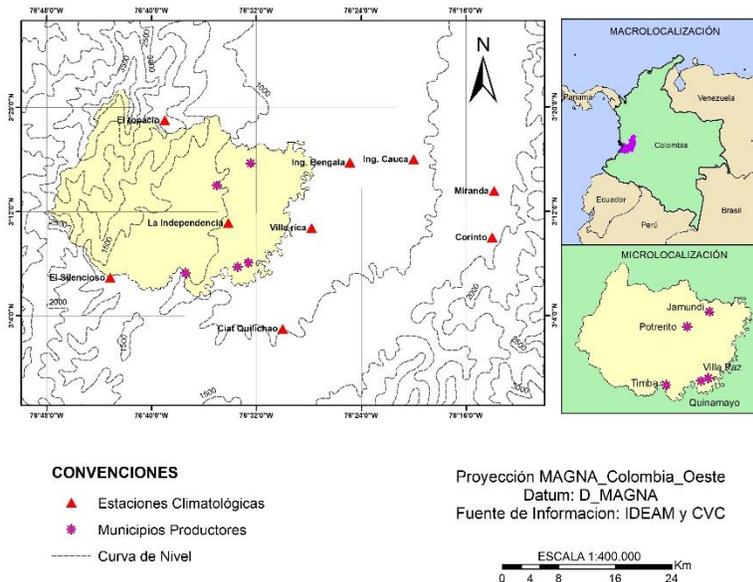


Figura 1. Ubicación de la zona de estudio.

Estaciones meteorológicas utilizadas

Se eligieron nueve estaciones alrededor de la zona productora de arroz (no superan los 50 km de distancia), las cuales proporcionaron información de registros diarios de las variables: Precipitación (mm), temperatura máxima y mínima (°C), humedad relativa (%), brillo solar (horas sol) y evaporación (mm) con más de 20 años de información. Las bases de datos se obtuvieron del Instituto de Estudios Ambientales de Colombia (IDEAM) y de los boletines hidro climatológicos de la Corporación Autónoma del Valle del Cauca (CVC) (Cuadro 1).

Cuadro 1. Descripción de las estaciones climáticas.

COD	Estación	Altitud (msnm)	Latitud (N)	Longitud (W)
1	El topacio	1 676	03° 19'	76° 39'
2	El silencioso	1 151	03° 06'	76° 43'
3	Ingenio Miranda	1 128	03° 16'	76° 20'
4	Corinto	1 100	03° 10'	76° 14'
5	Villa Rica	1 000	03° 10'	76° 28'
6	Ingenio Cauca	1 000	03° 14'	76° 13'
7	Ingenio Bengala	1 000	03° 15'	76° 24'
8	Ciat Quilichao	977	03° 03'	76° 30'
9	La Independencia	963	03° 11' 8"	76° 34' 9"

Caracterización de la información de las estaciones meteorológicas.

Las estaciones seleccionadas no contienen el mismo periodo de información de las variables climatológicas del estudio en curso, se eligieron aquellas estaciones cuyos registros cumplieran con la mayor información registrada de datos diarios, los cuales se observan a las 8:00 am, así se establecen los periodos de estudio para cada variable y el porcentaje total de datos ausentes (Cuadro 2).

Cuadro 2. Datos ausentes para cada variable de acuerdo con su periodo de estudio.

Variable	NE	PE	NA	TD	TDA	DA (%)
Precipitación	9	1969-2018	50	164 358	26 171	16
Brillo solar	5	1985-2018	34	62 090	11 774	19
Evaporación	4	1983-2017	35	51 136	5 064	10
Temperatura máxima	5	1989-2016	28	51 135	14 761	29
Temperatura mínima	5	1989-2016	28	51 135	12 252	24
Humedad relativa	4	1982-2016	35	51 136	7 843	15

NE= número de estaciones; PE= periodo de estudio; NA= número de años; TD= total de datos de cada variable; TDA= total de datos ausentes; DA= datos ausentes.

Estimación de datos faltantes diarios

Se realizó el relleno de las series históricas diarias de la zona de estudio mediante los siguientes métodos.

Razón normal (RN)

Con el paquete climatol programado en R, afín a la metodología propuesta por Guijarro (2004), cuyo proceso consiste en: el cálculo de cada serie, como si no existiera, por interpolación de datos estandarizados de series vecinas, el análisis de las diferencias entre las series calculadas y las observadas para detectar datos erróneos, saltos en la media y tendencias y por último el relleno automático de los datos faltantes y la corrección de los errores aislados. A continuación, se presenta la ecuación del método: $P_x = \frac{1}{n} \left[\frac{N_x}{N_a} P_a + \frac{N_x}{N_b} P_b + \frac{N_x}{N_n} P_n \right]$. Donde: P_x = dato faltante de precipitación en la estación x, N_x , N_a , N_b ... N_n = valor de la precipitación media diaria de la estación del dato faltante (x) y las estaciones auxiliares a, b, n (medias de todas las series históricas) y P_a , P_b , P_n = valor de la precipitación registrada en las estaciones auxiliares del día donde falta el dato en la estación x.

Análisis armónico de las series de Fourier (SF)

Esta es una función utilizada para variables periódicas como: la temperatura máxima y mínima, humedad relativa y evaporación. Es conocida como la función periódica $s(t)$, donde se calculan los coeficientes a_i y b_i y como resultado se tiene que la señal $s(t)$ se puede descomponer en la suma de senos y cosenos cuyas frecuencias son múltiplos armónicos de la frecuencia fundamental (Nauman y Vargas, 2012) Se estima de la siguiente manera: $s(t) = \frac{a_0}{2} + \sum_{n=1}^{N/2} (a_n \cos(j\omega t) + b_n \text{Sen}(j\omega t))$. $\frac{a_0}{2} =$

$\frac{2}{p} \sum_{-p/2}^{p/2} s(t) dt$, $a_1 = \frac{2}{p} \sum_{-p/2}^{p/2} s(t) \cos(iwt) dt$, $b_1 = \frac{2}{p} \sum_{-p/2}^{p/2} s(t) \sin(iwt) dt$. Donde: P= es el período $P = 2\pi/\omega$; ω = función de frecuencia angular; N= es la cantidad de datos a usar; cos= la función coseno; sen= la función seno. En este estudio, este método fue procesado en el programa Excel versión 2018. Para determinar la homogeneidad de los datos de las variables meteorológicas, se realizó una cuantificación de datos ausentes y posteriormente, se eliminó la información considerada discrepante o inconsistente, con el fin de obtener agrupaciones de datos más representativos; y como se indicó previamente, los datos perdidos permitieron estimar los índices estadísticos propuestos para evaluar el comportamiento de cada método respecto a cada variable.

Cuadrado del inverso de la distancia (IDW)

El procedimiento del cuadrado del inverso de la distancia es explicado por Campos (1998), donde la influencia del valor de la variable en una estación, para el cálculo de esta en cualquier punto es

inversamente proporcional a la distancia entre los dos puntos. $P_x = \frac{\sum_{i=1}^N \frac{1}{d_i^2} P_i}{\sum_{i=1}^N \frac{1}{d_i^2}}$. Donde: P_x = el dato

estimado de la estación incompleta; P_i = valores de lluvia de los pluviómetros utilizados para la estimación; d_i = distancia de cada lugar a punto que se está estimando; N = número de estaciones vecinas.

De acuerdo con Campos (1998), se recomienda utilizar cuatro estaciones auxiliares (las más cercanas), de manera que cada una quede localizada en uno de los cuadrantes que definen los ejes coordenados que pasan por la estación incompleta, generalmente norte-sur y oriente-poniente. Se realizó el procedimiento en el programa Excel 2018. Para emplear estos métodos, se hizo una pérdida de datos observados que se calcularon y posteriormente se compararon, para poder definir la bondad de cada uno de estos.

Se denominó ‘datos perdidos’ a aquella información perteneciente al volumen de datos real por variable (TD) y que luego fue asumida como inexistente durante la aplicación de los métodos de relleno. La cantidad de datos perdidos se fijó de acuerdo con el porcentaje de datos ausentes del TD (Cuadro 2 y 3). Así pues, se consideró un porcentaje de datos al azar para efectuar los métodos de prueba, y utilizar dicha información para calcular los índices estadísticos.

Cuadro 3. Datos perdidos para cada variable de acuerdo con su periodo de estudio.

Variable	TD	TDP	%DP
Precipitación	164 358	16 439	10
Brillo solar	62 090	2 924	5
Evaporación	51 136	7 309	14
Temperatura Máxima	51 135	9 497	19
Temperatura Mínima	51 135	9 497	19
Humedad Relativa	51 136	7 308	14

TD= total de datos de cada variable; TDP= total de datos perdidos; %DP= porcentaje de datos perdidos.

Estimación del error

A fin de estimar la exactitud de los métodos mencionados para reconstruir los datos ausentes y seleccionar el mejor, se empleó. La raíz cuadrada del cuadrado medio del error (RSME) (Teegavarapu y Chandramouli, 2005) y el coeficiente de determinación (R^2) (Cervantes *et al.*, 2013). El modelo ideal es cuando $R^2 \approx 1$ o con buen ajuste ≥ 0.5 y el $RMSE \approx 0$ (Pereira, 2004).

$RMSE = \sqrt{\frac{\sum_{i=1}^n |V_{obs_i} - V_{est_i}|^2}{n}}$. Donde: vobs= dato observado; Vest= dato estimado; n= número de observaciones o estimaciones. $R^2 = \frac{[\sum_{i=1}^n (a_i - \bar{a})(t_i - \bar{t})]^2}{[\sum_{i=1}^n (a_i - \bar{a})^2][\sum_{i=1}^n (t_i - \bar{t})^2]}$. Donde: a_i = dato estimado por el método; t_i = dato observado; \bar{a} = promedio de los datos estimados por el método; \bar{t} = promedio de los datos observados.

Resultados y discusión

Una búsqueda en Scopus se ha encontrado 17 documentos relacionados con el uso de las series o análisis de fourier; desde el año 1986 hasta la actualidad. Seguidamente, el método de la distancia inversa ponderada (IDW), se reconocieron 46 archivos a nivel de estudio diario entre 2001 y 2019. Scopus registra 33 documentos con el empleo del método de la razón normal (RN), los cuales corresponden a las áreas de estudio de: ingeniería, agricultura, ciencias de la tierra y el ambiente en la estimación de datos faltantes a nivel diario usado desde 1963 hasta la fecha actual. Dicha búsqueda exhibe la evidencia reciente sobre el tema en curso, el cual sugiere que las divulgaciones presentes pueden ser llamativas y apuntan a que la investigación agrícola logre, a largo plazo, llevar a que la agricultura alcance otros nichos de investigación prometedores en el futuro.

Precipitación

Para esta variable se emplearon nueve estaciones para completar sus registros y se obtuvo intervalos de los errores (RMSE) de 7.6 mm a 13 mm y de 6.8 mm a 13.9 mm con el método RN e IDW, respectivamente, estos errores son aceptados dado que para la zona de estudio la precipitación anual corresponde a 1 706 mm año⁻¹ y la precipitación más baja se presenta en el mes de julio con alrededor de 62 a 80 mm mes⁻¹. Los valores obtenidos son superiores a los reportados por: Kanda *et al* (2017), quienes completaron esta variable para 12 estaciones de 31 años de registros diarios; y estimaron errores (RMSE) para toda la serie histórica, para siete, quince y treinta días faltantes acumulados y el error promedio de la serie histórica; a través del método IDW obtuvieron 2.531, 2.497, 2.524 y 2.432 mm respectivamente y en el método RN reportan 2.375, 2.382, 2.387 y 2.372 mm correspondientemente.

Asimismo, Xia *et al.* (1999), reportan que para el IDW y RN un error promedio de 1.5 y 1 mm día⁻¹; en un estudio realizado para 44 estaciones con un periodo de 5 años de registros diarios. No obstante, se obtuvo un R^2 máximo entre las estaciones de 0.45 y 0.48 en los métodos de RN e IDW, estos son superiores según Gallegos *et al.* (2016), indican un valor de R^2 de 0.376 y 0.363 en los métodos RN e IDW para siete estaciones auxiliares usadas en San Luis Potosí, México. Por implicación, la distancia es un criterio lógico que se debe considerar en la estimación de relleno de series climáticas y las estaciones vecinas deberían estar cerca del punto a analizar, sin obstáculos como montañas y barreras que modifiquen su comportamiento (Vincent y Gullet, 1999).

Por ende, un aspecto relacionado con la distancia es la altura, ya que para el área de estudio las estaciones no exceden los 50 km en la horizontal y tampoco los 750 m en la vertical, esto permite tener climas parecidos como en la estación Ing. Miranda e Ing. Cauca que están a 13 km de distancia y 128 m de altura una de la otra, y aun así muestran ajustes similares de R^2 en dos métodos de relleno diferentes (Figuras 2).

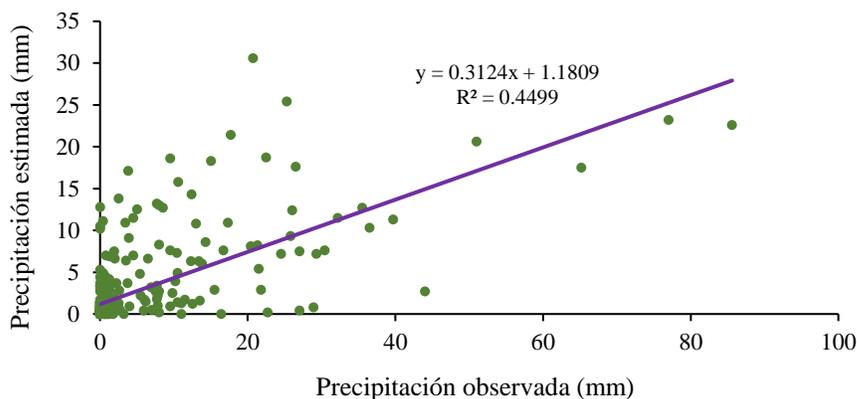


Figura 2. Relación de la precipitación observada y estimada de la estación Ing. miranda a través del método RN.

Expresado lo anterior, con el propósito de obtener rellenos más adecuados en grandes cantidades de información diaria (mayor a 10 años de registros diarios), algunos autores sugieren emplear estaciones no mayores a 50 km de distancia (De Luis *et al.*, 2007; Toro *et al.*, 2015). Hasta ahora, ha quedado expuesto que una serie de referencia se calcula mediante una media ponderada entre estaciones vecinas, al utilizar como factor de ponderación el coeficiente de correlación o la distancia; no obstante, el método IDW es el que se divulga como el que mejor se ajusta al relleno de esta serie climatológica y concuerda con lo obtenido en este trabajo (Teegavarapu y Chandramouli, 2005; Toro *et al.*, 2015).

Brillo solar

En cuanto a la variable brillo solar (horas sol o heliofanía) se obtuvo un rango de errores (RMSE) de 1.6 h a 2.4 h y de 1.4 a 2.5 h con coeficientes de determinación de 0.71 y 0.78 por los métodos RN e IDW para las cinco estaciones usadas (Figura 3). Estos resultados son aceptables e inferiores respecto a los de Grossi *et al.* (2010), quienes lograron promedios diarios similares de 2.1, 2.9, 2.5 y 2.2 h sol para los meses de enero, octubre, noviembre y diciembre respectivamente; al completar series temporales mensuales de 20 años de heliofanía en 11 estaciones de Uruguay mediante el método Kriging.

En América Latina, muy pocos países reconocen la importancia de la radiación solar y carecen de esta; además, de datos históricos de ensayos de campo sobre este elemento climático (Degiovanni *et al.*, 2010) de modo que, dadas las semejanzas que presentan en el índice RMSE y el R^2 de las estimaciones, se pudieran emplear cualquiera de los dos métodos; ya que, el coeficiente de determinación admite distinguir que existe dependencia entre los parámetros de ponderación o pesos, y la distancia de las estaciones (Toro *et al.*, 2015).

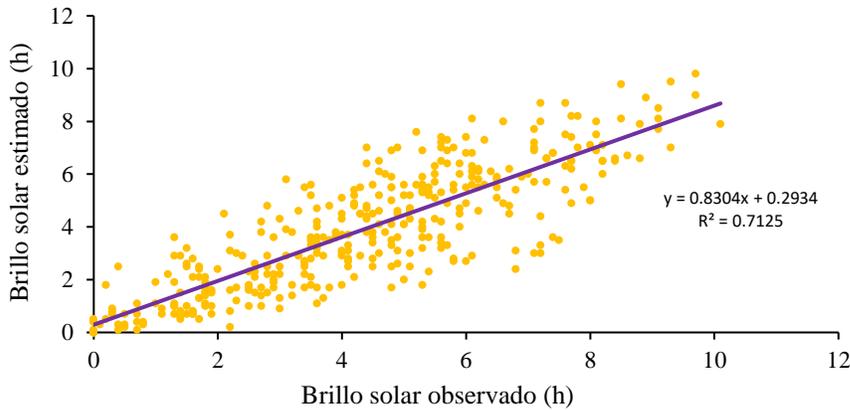


Figura 3. Relación del brillo solar observado y estimado de la estación Ing. Cauca a través del método RN.

Temperatura mínima

En este caso se presentan valores máximos de RMSE de hasta 1.5 °C, 1.03 °C, 1.64 °C, y R^2 de 0.21, 0.25, 0.1 en los métodos RN, SF e IDW usados para las cinco estaciones empleadas, se observó que tienen una tendencia entre sí, aunque los RMSE más altos se obtienen con el método IDW. Al relacionar los valores estimados con los observados a través de cada método, se encontró bajos coeficientes de determinación; por lo tanto, no presentan una relación significativa (Figura 4). Una característica de este argumento es que para esta variable se presenta poca variabilidad espacial y temporal, además por las características altitudinales del lugar y su ubicación ecuatorial.

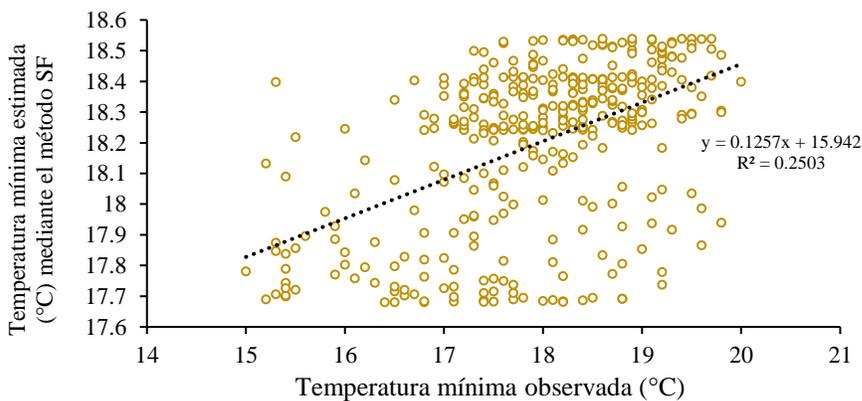


Figura 4. Relación de la temperatura mínima observada y estimada de la estación Ciat Quilichao.

Estudios previos destacan que el método RN es viable dado que los resultados obtenidos de RMSE se encuentran por debajo de 3.5 °C reportado por Kanda *et al.* (2017), que completaron la cantidad de datos ausentes en 12 estaciones de Karakorum Himalaya. La fiabilidad de los métodos RN e IDW es corroborada por Gallegos *et al.* (2016), quienes revelan valores de RMSE de 4.1 y 3.7 °C respectivamente, para un periodo de estudio de 16 años de registros diarios con 25.4% de información faltante, es interesante observar que para este estudio se tiene 24% de información faltante en series de 28 años de registros diarios y los índices estadísticos de RMSE fueron mucho menores; lo que indica que ambos métodos serán más adecuados para la completar información faltante en periodos de estudio de más de 25 años.

También, se muestra un trabajo empleado por Cervantes *et al.* (2019), el cual presenta resultados de RMSE no mayores a 0.16 °C, para las estimaciones realizadas a cuatro estaciones; por un periodo de estudio de cinco años de registros diarios. La temperatura mínima se presenta de forma más regular en el tiempo; puesto que, las variaciones temporales de los elementos meteorológicos están relacionadas directa o indirectamente con los movimientos de la tierra (Riehl, 1965). Siendo así, se elige el mejor método de la temperatura mínima, la serie de Fourier tras el empleo de su segunda armónica. Este método permite calcular el valor ausente, más probable de la temperatura en cualquier día del año sobre un lugar determinado.

Temperatura máxima

En lo que se refiere al relleno de las series históricas de la temperatura máxima se obtuvieron como resultados de los métodos RN, SF e IDW mediante el RMSE 1.9 °C, 2 °C, 1.8 °C, respectivamente. Se observó que los valores de RMSE son similares entres si, comparando los valores reportados por Kanda *et al.* (2017), en los métodos RN y IDW, encontraron 2.6 y 2.5 °C, de error máximo, siendo los valores superiores a los obtenidos en este estudio.

Una de las estimaciones representativas hechas por los métodos RN e IDW para la temperatura máxima, fue realizada para la estación Ing. Bengala, la cual presentó el valor de $R^2 = 0.5$ (Figura 5); esto indica, que estos métodos tienen una bondad moderada en el ajuste respecto a esta variable. Caso diferente acontece con la estación Ing. Cauca que no obtuvo un coeficiente de determinación aceptable. Gallegos *et al.* (2016), estimaron registros faltantes para esta variable y adquirieron ajustes máximos de R^2 de 0.641 y 0.493 para la RN e IDW para 16 años de registros diarios y ajustes mínimos de hasta 0.25 de R^2 en ambos métodos, lo que indica que los obtenidos en este trabajo son admisibles, Cervantes *et al.* (2019), estimaron 229 datos faltantes a través del método SF en una serie histórica de 5 años y reportaron RMSE de hasta 0.149 °C en cuatro estaciones, lo que indica que este método presenta mejores resultados a menor cantidad de datos faltantes.

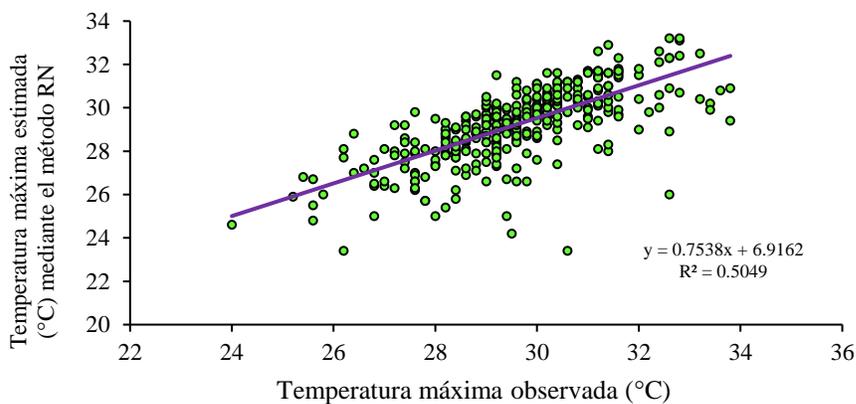


Figura 5. Relación de la temperatura máxima observada y estimada de la estación Ing. Bengala.

Humedad relativa

Diagnosticada la variable humedad relativa a lo largo de todo el conjunto de datos se ha identificado un amplio intervalo en sus registros diarios que van de 45.7% a 99.4% en las cuatro estaciones disponibles; para los cuales, en el relleno de esta serie permite admitir hasta 10.4%, 9.02%, 8% de

RMSE para los métodos RN, SF e IDW, teniendo en cuenta que para esta variable, la manera de completar los datos ausentes en los registros históricos no se ha explorado con gran detalle a nivel diario, se han elegido estos métodos para este estudio.

Por otra parte, Coutinho *et al.* (2018), completaron esta variable a nivel mensual para un estudio de 13 años, en el cual, adquirieron valores inferiores a 2.9% de error con cuatro métodos diferentes. Estos resultados son admisibles, ya que, esta variable periódica tiende a representarse con valores mayores al 70% en lugares cercanos al ecuador (como lo es la ubicación de este estudio) y adicionalmente, la diferencia altitudinal entre las estaciones usadas para dicha estimación no supera los 150 msnm (Araya, 2011).

Respecto a los resultados adquiridos por el R^2 fueron: 0.07, 0.32, 0.18 de los métodos RN, SF e IDW, cuyos valores no son un estadístico adecuado que por sí solo pueda predecir el comportamiento de dicha variable (Teegavarapu y Chandramouli, 2005), dado que, los datos difieren entre sí debido al proceso de relleno, puede notarse, entre estas representaciones la ventaja de mostrar no solo los datos que no fueron aleatoriamente reemplazados por dato faltante, sino que también da una idea de cuánto se desvían los datos con respecto al dato real (Araya, 2014).

De las estaciones usadas, se observa que Ciat Quilichao es la que tiende a mostrar un mejor ajuste en la estimación mediante el método de la SF. Esto es posible gracias al elevado grado de humedad que presenta el área de estudio, ya que, para los valores observados y los estimados es evidente un gran número de datos concentrados (Figura 6). En cambio, la relación de la humedad observada y la estimada para la estación Ing. Miranda con el método RN un valor observado obtuvo un conjunto de datos estimados para el periodo de evaluación en curso.

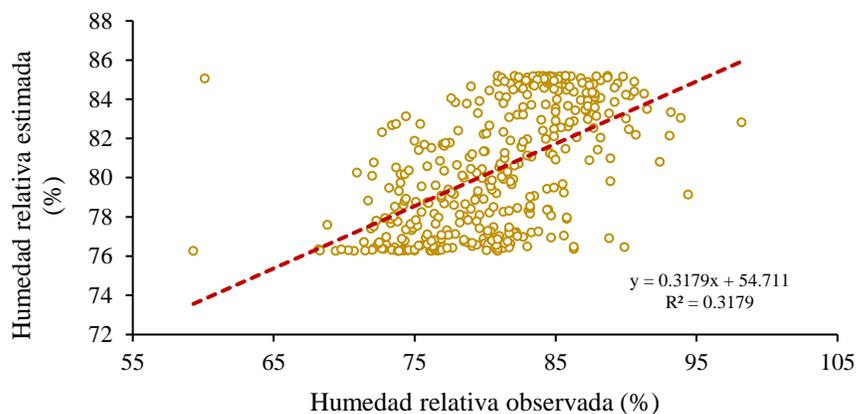


Figura 6. Relación de la humedad relativa observada y estimada de la estación Ciat Quilichao mediante el método SF.

Evaporación

La evaporación presentó valores de RMSE de 1.8 mm, 2.1 mm, 1.61 mm, en los métodos RN, SF e IDW, cuyo error promedio obtenido es 1.84 mm en las cuatro estaciones usadas, considerándose mínimo. Una interpretación de los resultados expuestos es que el relleno de las series históricas de esta variable es más conveniente mediante la aproximación de segundo orden de las series de Fourier, porque, para la zona de estudio elegida se presenta una amplia variabilidad en la humedad relativa y esta influye directamente sobre la manifestación de la evaporación e implícitamente en la reconstrucción de su serie temporal.

Los métodos RN e IDW para la estación Ciat Quilichao se muestran de forma conveniente, dado que en sus regresiones presentan ligeras tendencias o acumulación en la pendiente entre los valores de 3 y 5.5 mm de la evaporación observada. En comparación a la estación Topacio esta carece de una tendencia clara, y se observó una distribución amplia en la parte inferior derecha de la regresión, lo que permite inferir que esta variable para esta estación presenta algunos valores extremos en ciertas secciones de la serie histórica presente (Figura 7).

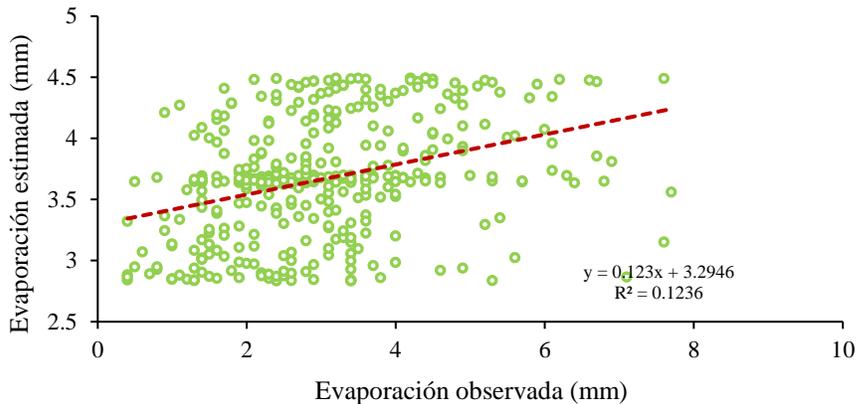


Figura 7. Relación de la evaporación observada y estimada de la estación topacio mediante el método SF.

La literatura científica desarrollada hasta el momento exterioriza que el relleno de datos faltantes de series climatológicas comprende un análisis detallado y cauteloso por variable, del cual, se podría mencionar que es concluyente la inexistencia de un método explícito para cada caso de estudio, si no, de un método que se ajusta al comportamiento de la variable.

Conclusiones

Se determina que el método del cuadrado del inverso de la distancia fue el que mejor se comportó para el relleno de las diferentes variables utilizadas, le siguió el de la razón normal, por lo que ambos se recomiendan para el relleno de series de datos climatológicos diarios. Un método que no se debe aplicar a todas las variables, es el de las series de Fourier, ya que para utilizarlo debe ser en series periódicas en el tiempo. Con los datos observados de las variables y los estimados con los diferentes métodos, para todas las variables se obtuvieron errores admisibles, no así para el coeficiente de determinación, esto último se debe por las características: geográficas, latitudinales y altitudinales de la zona de estudio.

Literatura citada

- Araya, J. L. 2011. Resultados de un control de calidad de datos de temperatura superficial. Costa Rica. Tecnología en Marcha. 1(24):33-49.
- Araya, J. L. 2014. Experiencias en la aplicación operativa de un método multivariado de imputación de datos meteorológicos. Costa Rica. Tecnología en Marcha. 3(27):70-79.
- Biblioteca Scopus. 2019. Universidad Autónoma Chapingo. <https://www.scopus.com>.

- Campos, A. D. F. 1998. Procesos del ciclo hidrológico. Editorial Universitaria Potosina. 3^a (Ed.). Facultad de Ingeniería. Universidad Autónoma San Luis Potosí, México. 500 p.
- Cervantes, R.; Arteaga, R.; Vázquez, M. A.; Ojeda, B. W. y Quevedo, N. A. 2013. Modelos Hargreaves Priestley-Taylor y redes neuronales artificiales en la estimación de la evapotranspiración de referencia. *Ingeniería Investigación y Tecnología*. 16(2):163-176.
- Cervantes, R.; Arteaga, R.; Vázquez, M. A.; Ojeda, B. W. y Quevedo, N. A. 2019. Red neuronal artificial y series de Fourier para pronóstico de temperaturas en el Distrito de Riego 075. Sinaloa, México. *Tecnología y Ciencias del Agua*. 1(10):253-268.
- Cortés, J. A.; Medina, A. F. A. y Chaves, O. J. A. 2007. Del análisis de Fourier a las wavelets análisis de Fourier. Colombia. *Scientia Technica*. 34(13):151-156.
- Coutinho, E. R.; Silva, R. M.; Madeira, J. G. F.; Coutinho, P. R. Boly, R. A. M. y Delgado, A. R. S. 2018. Aplicación de redes neuronales artificiales (ANN) en el relleno de brechas de series meteorológicas de tiempo. *Rev. Bras. Meteorol.* 33(2):317-328. Doi: <https://doi.org/10.1590/0102-7786332013>.
- Degiovanni, B. V.; Martínez, R. C. P. y Motta, O. F. 2010. Producción ecoeficiente del arroz en América Latina, Tomo I. Conferencia internacional de arroz para América Latina y el Caribe. (Ed.). Cali-Colombia. 513 p.
- De-Luis, M.; Longares, L. A.; Stepanek, P. y González, J.C. 2007. Tendencias estacionales de la precipitación en la cuenca del Ebro 1951-2000. España. *Geographica*. 52:53-78.
- Eischeid, J. K.; Pasteris, P. A.; Díaz, H. F.; Plantico, M. S. and Lott, N. J. 2000. Creating a serially complete, national daily time series of temperature and precipitation for the western united states. United states. *J. Appl. Meteorol.* 39(9):1580-1591. Doi:10.1175/15200450(2000)039<1580:CASCND>2.0.CO;2.
- Gallegos, J.; Arteaga, R.; Vázquez, M. A. y Juárez, J. 2016. Estimación de la falta de precipitación diaria y registros de temperatura máxima y mínima en San Luis Potosí. México. *Ingeniería Agrícola y Biosistemas*. 8(1):3-16. Doi: 10.5154/r.inagbi.2015.11.008.
- Grossi, G. H.; Raichijk, C. y Righini, R. 2010. Algunos aspectos de la climatología solar del Uruguay. Brasil. *Rev. Bras. Meteorol.* 4(25):479-486.
- Guijarro, J. A. 2004. Climatol: software libre para la depuración y homogeneización de datos climatológicos. *In: el clima, entre el mar y la montaña*. (Ed.). A-4. Asociación Española de Climatología. España. 493-502 pp.
- Kanda, N.; Negi, H. S.; Rishi, M. S. and Shekhar, M. S. 2017. Performance of various techniques in estimating missing climatological data over snowbound mountainous areas of Karakorum Himalaya. United States. *Meteorological Applications*. 25(3):337-349. Doi:10.1002/met.1699.
- Medina, R. D., Montoya, E. C. y Jaramillo, R. A. 2008. Estimación de valores faltantes en series históricas de lluvia. Colombia. *Cenicafé*. 59(3):260-273.
- Mintegui, J. y López, F. 1990. La ordenación agrohidrológica en la planificación. Servicio central de publicaciones del gobierno Vasco. 1 (Ed.). Vasco. 306 p.
- Nauman, G. y Vargas, W. M. 2012. Estabilidad de la estimación de la onda anual en escala diaria de la temperatura. Brasil. *Rev. Bras. Meteorol.* 4(27):401-412.
- Paulhus, J. L. H. and Kohler, M. A. 1952. Interpolation of missing precipitation records. United States. *Monthly Weather Review*. 80(8):129-133.
- Pereira, A. R. 2004. The Priestley Taylor parameter and the decoupling factor for reference evapotranspiration. Canada. *Agric. Forest Meteorol.* 125(3):305-313.
- Riehl, H. 1965. Tropical meteorology. McGraw Hill. New York, United states of America. 392 p.

- Teegavarapu, R. S. and Chandramouli, V. 2005. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. Países Bajos. *J. Hydrol.* 312(1-4):191-20
- Toro, A. M.; Arteaga, R.; Vázquez, M. A. e Ibáñez, L. A. 2015. Relleno de series diarias de precipitación, temperatura mínima, máxima de la región norte del urabá antioqueño. *Rev. Mex. Cienc. Agríc.* 6(3):577-588.
- Xia, Y. Fabian, P.; Stohl, A. and Winterhalter, M. 1999. Forest climatology: estimation of missing values for Bavaria, Germany. *Canada. Agric. Forest Meteorol.* 96(1-3):131-14. Doi: 10.1016/S0168-1923(99)00056-8.