

Estimation of missing daily precipitation and maximum and minimum temperature records in San Luis Potosí

Estimación de datos diarios faltantes en registros de precipitación y temperatura máxima y mínima en San Luis Potosí

Jesús Gallegos-Cedillo^{1*}; Ramón Arteaga-Ramírez¹; Mario Alberto Vázquez-Peña¹; Juan Juárez-Méndez²

Universidad Autónoma Chapingo, Ingeniería Agrícola y Uso Integral del Agua¹, Departamento de Suelos², Carretera México-Texcoco km 38.5, Chapingo, Estado de México, C. P. 56230, México. gallegosjc_1@hotmail.com, (595) 95 21 650 (*Corresponding author)

Abstract

Weather stations often have incomplete records for a certain period, due to the absence or replacement of the operator, recording device failures or operational negligence, thereby limiting the carrying out of agro-climatic and hydrological studies. Therefore, the aim of this study was to compare the deductive rational (DR), normal ratio (NR) and US National Weather Service's inverse square distance (ISD) methods, and then select the best for estimating missing daily precipitation and maximum and minimum temperature records of the weather stations located in San Luis Potosi, Mexico. Six stations were analyzed, of which 15 % of the precipitation (1,489 of 9,862) and 25 % of the maximum and minimum temperature (1,489 of 5,844) information was eliminated. Missing data were generated with the ISD, DR and NR methods, and their evaluation was performed using the RMSE and Willmott index of agreement, d , statistical indices, which indicated that the ISD method has the Willmott index close to one and the mean standard error close to zero; consequently, it was the one used to estimate the missing data from 108 weather stations in San Luis Potosí.

Keywords: statistical indices, normal ratio, deductive rational, inverse square distance.

Resumen

En las estaciones meteorológicas es común que los registros de información de un cierto periodo estén incompletos, por la ausencia o sustitución del operador, fallas del aparato registrador o negligencia operacional, lo que limita la realización de estudios agroclimáticos e hidrológicos. Por lo anterior, el objetivo del presente trabajo fue comparar los métodos: racional deductivo (RD), relación normalizada (RN) y cuadrado del inverso de la distancia o U.S. National Weather Service (WS), y seleccionar el mejor para calcular los registros de datos diarios faltantes de precipitación y temperatura máxima y mínima de las estaciones climatológicas ubicadas en San Luis Potosí, México. Se analizaron seis estaciones, de las cuales se eliminó 15 % de información de precipitación (1,489 de 9,862) y 25 % de temperatura máxima y mínima (1,489 de 5,844). Los datos faltantes se generaron con los métodos WS, RD y RN, y la evaluación de éstos se realizó con los índices estadísticos RMSE e índice de Willmott (d), los cuales indicaron que el WS presenta el índice de Willmott cercano a uno y el error estándar promedio cercano a cero; por consecuencia, fue el que se empleó para estimar los datos faltantes de las 108 estaciones climatológicas de San Luis Potosí.

Palabras clave: índices estadísticos, relación normalizada, racional deductivo, cuadrado del inverso de la distancia.

Please cite this article as follows (APA 6): Gallegos-Cedillo, J., Arteaga-Ramírez, R., Vázquez-Peña, M. A., & Juárez-Méndez, J. (2016). Estimation of missing daily precipitation and maximum and minimum temperature records in San Luis Potosí. *Ingeniería Agrícola y Biosistemas*, 8(1), 3-16. doi: 10.5154/r.inagbi.2015.11.008

Introduction

Several tools and techniques have been developed for measuring climatic elements in order to know the current weather (Gutiérrez, García, Magaña, & Escalante, 2007) and provide information for short-term forecasting, statistical analysis, farm support, air and maritime navigation safety, and characterization and classification of the weather and atmospheric state at a particular time and place. A meteorological instrument is the combination of a sensitive element and a system able to convert measurements of this element into a numerical value, which represents the meteorological variable being evaluated. The representative area of a region having meteorological equipment, at which climatic elements can be observed, measured, recorded, concentrated and processed, is commonly referred to as a weather station (Nájera & Arteaga, 1998).

Often weather station records for a certain period are incomplete due to the absence or replacement of the operator, recording device failures or operational negligence (Aparicio, 2007; Campos-Aranda, 1998), which has limited the conducting of agro-climatic studies whose results enable increasing productivity, optimizing resources, reducing the risk of crop loss, planning irrigation and drainage infrastructure in an integrated manner and making weather forecasts (Toro-Trujillo, Arteaga-Ramírez, Vázquez-Peña, & Ibáñez-Castillo, 2015). According to the World Meteorological Organization (OMM, 2011), the planning and implementation of user projects cannot be deferred until there are sufficient meteorological or climatological observations; therefore, estimation is used to expand, fill in or complete the information. It also recommends that the study period be at least 30 years.

Among the most common methods for estimating missing data series are: normal ratio (NR), which is used when the mean annual precipitation of any of the neighboring stations differs by more than 10 % (Aparicio, 2007), and when it differs by less than 10 % it is estimated with the arithmetic average of the records of the neighboring stations (Alfaro & Pacheco, 2000; McCuen, 1998). Aparicio (2007) recommends using at least three auxiliary stations in both cases.

The WMO (2011), Campos-Aranda (2015), Allen, Pereira, Raes, and Smith (2006) and DeGaetano, Eggleston, and Knapp (1995) propose using linear regression (LR) and multiple regression (MR). The latter consists of obtaining a mathematical equation that expresses the relationship between the dependent variable (Y) and the independent one (X), or explanatory variables (X_1, X_2, \dots, X_n), although rarely is a perfect linear relationship observed because the phenomena studied in climatology are usually nonlinear, so both records

Introducción

En la actualidad se han desarrollado diversas herramientas y técnicas para la medición de los elementos del clima con la finalidad de conocer el estado del tiempo (Gutiérrez, García, Magaña, & Escalante, 2007) y brindar información para previsión a corto plazo, análisis estadísticos, apoyo en la agricultura, seguridad en la navegación aérea y marítima, caracterización y clasificación del clima y del estado atmosférico en un momento y en un lugar determinado. Un instrumento meteorológico es la combinación de un elemento sensible y un sistema capaz de convertir las mediciones de este elemento en un valor numérico, el cual representa la variable meteorológica que se está evaluando. Al área representativa de una región dotada de equipo meteorológico, en la que se observa, miden, registran, concentran y procesan las mediciones del clima, comúnmente se le denomina estación meteorológica (Nájera & Arteaga, 1998).

Es frecuente que en las estaciones meteorológicas los registros de información de un cierto periodo estén incompletos, debido a la ausencia o sustitución del operador, fallas del aparato registrador o negligencia operacional (Aparicio, 2007; Campos-Aranda, 1998); lo que ha limitado la realización de estudios agroclimáticos cuyos resultados permitan aumentar la productividad, optimizar los recursos, reducir los riesgos de pérdida de cosecha, planificar integralmente la infraestructura de riego y drenaje, así como realizar predicción climática (Toro-Trujillo, Arteaga-Ramírez, Vázquez-Peña, & Ibáñez-Castillo, 2015). De acuerdo con la Organización Meteorológica Mundial (OMM, 2011), la planificación y ejecución de los proyectos de usuarios no pueden aplazarse hasta que se cuente con suficientes observaciones meteorológicas o climatológicas; por lo tanto, se utiliza la estimación para ampliar, rellenar o completar la información. Asimismo, recomienda que el periodo de estudio sea mínimo de 30 años.

Entre los métodos más comunes para la estimación de series de datos faltantes se encuentran: relación normalizada (RN), el cual se utiliza cuando la precipitación media anual de cualquiera de las estaciones circundantes difiere en más de 10 % (Aparicio, 2007), y cuando difiere en menos de 10 % se estima con el promedio aritmético de los registros de las estaciones circundantes (Alfaro & Pacheco, 2000; McCuen, 1998). Aparicio (2007) recomienda emplear tres estaciones auxiliares como mínimo en ambos casos.

La OMM (2011), Campos-Aranda (2015), Allen, Pereira, Raes, y Smith (2006) y DeGaetano, Eggleston, y Knapp (1995) proponen utilizar la regresión lineal (RL) y múltiple (RM). Esta última consiste en obtener una ecuación matemática que exprese la relación entre la variable dependiente (Y) y la independiente (X), o

must be homogeneous; that is, they need to represent the same conditions. When there are no stations surrounding the station with missing information, the missing values are deduced with the deductive rational (DR) method proposed by Campos-Aranda (1998), which allows estimating them with the information provided by the complete years of the same series (Puertas-Orozco, Carvajal-Escobar, & Quintero-Angel, 2011).

The method most used in hydrological and geographical studies is the U.S. National Weather Service's inverse square distance (ISD) method (Ramírez-Cruz, López-Velasco, & Ibáñez-Castillo, 2015). In this case, the influence of rain at a station for the calculation thereof at any point is inversely proportional to the distance between the station and the auxiliary stations (OMM, 2011). The most significant advantage of the ISD method is that it uses daily data, grouped into periods of five or ten days, monthly or yearly (Teegavarapu & Chandramouli, 2005; Toro-Trujillo et al., 2015).

Other procedures include moving averages (Campos-Aranda, 1998) and autoregressive, logarithmic, and triangulation methods, all based on normality assumptions. According to the WMO (2011), there are also estimation methods known as geostatistics based on theoretical foundations, which are statistical assumptions. The most important are the Kriging and optimal interpolation methods. The former is a spatial interpolation method based on showing the proportion at which the variance changes between points in space, and is expressed in a variogram. These procedures are limited by requiring a certain amount of data to produce a reliable and appropriate variogram (Toro-Trujillo et al., 2015). In addition, Wagner, Fiener, Wilken, Kumar, and Schneider (2012) suggest that these procedures do not represent an improvement over the ISD method.

To evaluate and select the method that best fits the variable analyzed, there are various tests and procedures known as statistical indices, which consist of estimating the error or bias; some of them are: the mean standard error or root mean square error (RMSE), the Willmott index of agreement (d) (Willmott, 1981), the coefficient of determination (R^2) and the relative error (RE) (Cervantes-Osornio, Arteaga-Ramírez, Vázquez-Peña, Ojeda-Bustamante, & Quevedo-Nolasco, 2013; Kelso-Bucio, Bâ, Sánchez-Morales, & Reyes-López, 2012; Wright, 2001; Cai, Liu, Lei, & Pereira, 2007).

The aim of this study was to compare the deductive rational (DR), normal ratio (NR) and United States Weather Services' inverse square distance (ISD) methods, and then select the best one to estimate the missing daily precipitation and maximum and minimum temperature data at weather stations located in San Luis Potosí, Mexico.

variables explicativas (X_1, X_2, \dots, X_n), aunque raras veces se observa una relación lineal perfecta debido a que los fenómenos que se estudian en climatología usualmente no son lineales, por lo que ambos registros deben ser homogéneos; es decir, que necesitan representar las mismas condiciones. Cuando no se dispone de estaciones circundantes a la estación incompleta, o bien, las existentes no cuentan con observaciones en el periodo analizado, los valores faltantes se deducen con el método racional deductivo (RD) propuesto por Campos-Aranda (1998), el cual permite estimarlos con la información que brindan los años completos de la misma serie (Puertas-Orozco, Carvajal-Escobar, & Quintero-Angel, 2011).

El método más utilizado en estudios hidrológicos y geográficos es el cuadrado del inverso de la distancia, también conocido como U.S. National Weather Service (WS) (Ramírez-Cruz, López-Velasco, & Ibáñez-Castillo, 2015). En este caso, la influencia de la lluvia en una estación para el cálculo de la misma en cualquier punto es inversamente proporcional a la distancia comprendida entre la estación y las estaciones auxiliares (OMM, 2011). La ventaja más importante del WS es que utiliza datos diarios, agrupados en periodos de cinco o diez días, mensuales o anuales (Teegavarapu & Chandramouli, 2005; Toro-Trujillo et al., 2015).

Otros procedimientos son: medias móviles (Campos-Aranda, 1998), autorregresivo, logarítmico, y de triangulación; todos basados en postulados de normalidad. Según la OMM (2011) también existen métodos de estimación conocidos como geoestadísticos basados en fundamentos teóricos, los cuales son supuestos estadísticos. Los más importantes son el Kriging y la interpolación óptima. El primero es un método de interpolación espacial que se basa en mostrar la proporción en la que cambia la varianza entre los puntos en el espacio, y se expresa en un variograma. Estos procedimientos se ven limitados al requerir una cantidad determinada de datos para producir un variograma confiable y adecuado (Toro-Trujillo et al., 2015). Además, Wagner, Fiener, Wilken, Kumar, y Schneider (2012) sugieren que estos no presentan mejoría con respecto al método del WS.

Para evaluar y seleccionar el método que mejor se ajuste a la variable analizada, existen diversas pruebas y procedimientos conocidos como índices estadísticos, que consisten en estimar el error o sesgo; algunos de ellos son: el error estándar promedio o raíz cuadrada del cuadrado medio del error (RMSE), el índice de concordancia de Willmott (d) (Willmott, 1981), el coeficiente de determinación (R^2) y el error relativo (RE) (Cervantes-Osornio, Arteaga-Ramírez, Vázquez-Peña, Ojeda-Bustamante, & Quevedo-Nolasco, 2013; Kelso-Bucio, Bâ, Sánchez-Morales, & Reyes-López, 2012; Wright, 2001; Cai, Liu, Lei, & Pereira, 2007).

Materials and methods

Description of the study area

The state of San Luis Potosí is located in the Central Mexican Plateau. Its extreme geographical coordinates are: 24° 29' and 21° 10' north latitude, 98° 20' and 102° 18' west longitude. This area accounts for 3.2 % of the country's total area (Instituto Nacional para el Federalismo y Desarrollo Municipal [INAFED], 2010), and because of its climatic and morphological diversity, the state is divided into four geographical regions: *Altiplano*, *Centro*, *Media* and *Huasteca* (Instituto Nacional de Estadística y Geografía [INEGI], 2007).

Climatological databases

The databases used were: ERIC III 2.0 (rapid climatological information extractor) administered by the Mexican Institute of Water Technology (IMTA, 2009) and the network of weather stations operated by the National Water Commission (CONAGUA, 2015); daily precipitation (1984-2010 period) and maximum and minimum temperature (1995-2010 period) records were obtained from both.

Prior analysis of information from 108 weather stations

Since not all the stations contain records of the climatological variables under study and many lack information in different periods, those stations whose precipitation and maximum and minimum temperature records were more than 85 % complete were selected. Years with no data or few records were removed. In addition, the data series were updated to 2010, with the values obtained from the National Meteorological Service (SMN) database, called KMZ, an extension that allows downloading up-to-date weather records. Finally, 27 years were considered as the study period for precipitation and 16 years for maximum and minimum temperature.

Of the total of 190 weather stations, 81 located in San Luis Potosí were examined, plus some neighboring ones in Tamaulipas, Zacatecas, Nuevo León, Querétaro, Guanajuato, Hidalgo and Veracruz, of which 27 were selected for analysis.

Figure 1 shows the percentage of daily information obtained from the precipitation and maximum and minimum temperature records of the 108 weather stations. In the case of precipitation, two neighboring stations with a lesser amount of information were taken to correctly delimit San Luis Potosí. Similarly, for maximum temperature, four stations with less than 85 % information were taken, while the

El objetivo de este trabajo fue comparar los métodos: racional deductivo (RD), relación normalizada (RN) y cuadrado del inverso de la distancia o U.S. National Weather Service (WS), y seleccionar el mejor para estimar los datos diarios faltantes de precipitación, temperatura máxima y mínima, en las estaciones climatológicas ubicadas en San Luis Potosí, México.

Materiales y métodos

Descripción del área de estudio

El estado de San Luis Potosí se localiza en la altiplanicie central mexicana. Sus coordenadas geográficas extremas son: 24° 29' y 21° 10' de latitud norte, 98° 20' y 102° 18' de longitud oeste. Esta superficie constituye 3.2 % del área total del país (Instituto Nacional para el Federalismo y Desarrollo Municipal [INAFED], 2010), y debido a la diversidad climática y morfológica, el estado se divide en cuatro regiones geográficas: *Altiplano*, *Centro*, *Media* y *Huasteca* (Instituto Nacional de Estadística y Geografía [INEGI], 2007).

Bases de datos climatológicas

Las bases de datos utilizadas fueron: ERIC III 2.0 (extractor rápido de información climatológica) del Instituto Mexicano de Tecnología del Agua (IMTA, 2009) y la red de estaciones climatológicas de la Comisión Nacional del Agua (CONAGUA, 2015); de ambas se obtuvieron los registros diarios de precipitación (periodo 1984 a 2010), y temperatura máxima y mínima (periodo 1995 a 2010).

Análisis previo de la información de las 108 estaciones climatológicas

Dado que no todas las estaciones contienen registros de las variables climatológicas bajo estudio y muchas carecen de información en diferentes periodos, se seleccionaron aquellas estaciones cuyos registros de precipitación, y temperatura máxima y mínima contaban con más de 85% de información. Se eliminaron años sin datos o con pocos registros. Además, se actualizaron las series de datos hasta el 2010, con los valores obtenidos de la base de datos del Servicio Meteorológico Nacional (SMN) denominado KMZ, extensión que permite bajar información actualizada de los registros de los elementos meteorológicos. Finalmente, se consideraron 27 años como periodo de estudio para precipitación y 16 años para temperatura máxima y mínima.

Del total de 190 estaciones climatológicas, se examinaron 81 pertenecientes a San Luis Potosí, y se incluyeron algunas estaciones climatológicas circundantes, correspondientes a Tamaulipas, Zacatecas,

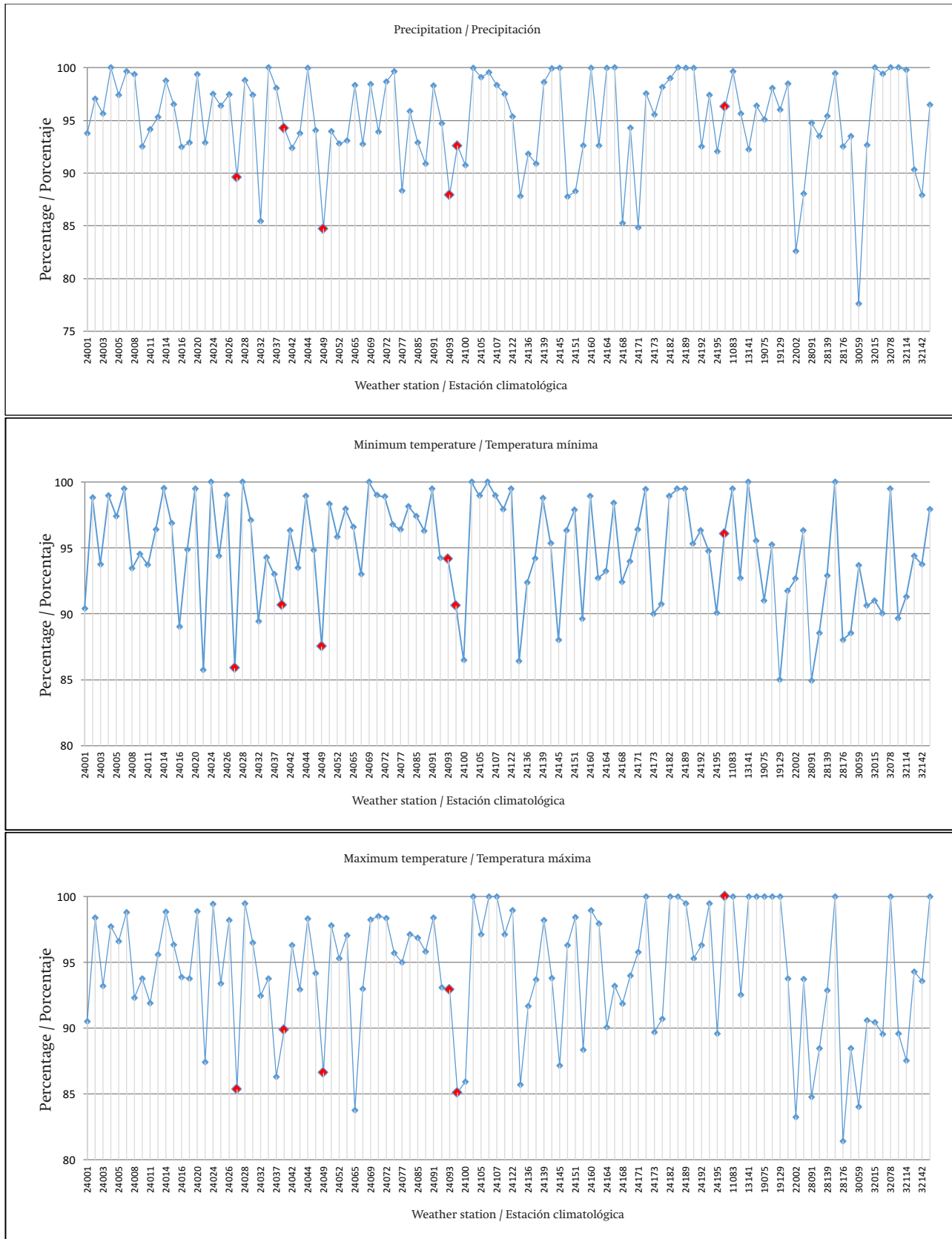


Figure 1. Percentage of historical information by weather station.

Figura 1. Porcentaje de información histórica por estación climatológica.

minimum temperature records have the required information. For all three cases the same selected stations were taken.

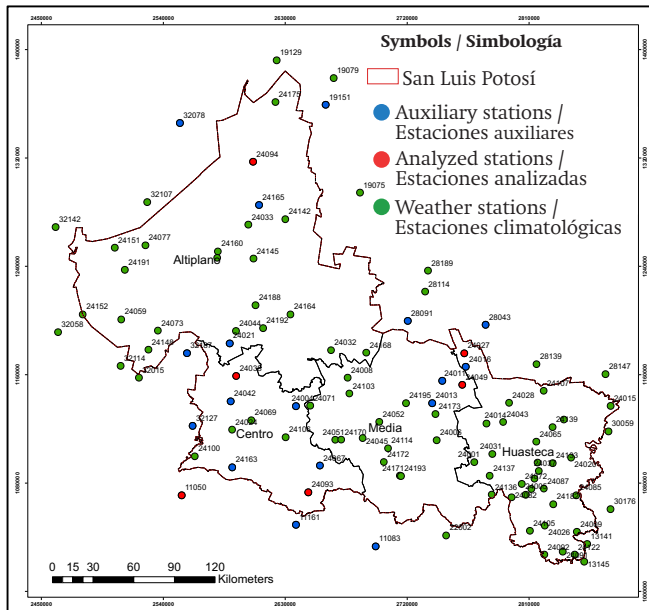


Figure 2. Location of the study area and the network of weather stations.

Figura 2. Ubicación de la zona de estudio y de la red de estaciones climatológicas.

Nuevo León, Querétaro, Guanajuato, Hidalgo y Veracruz, de las cuales se seleccionaron 27 para su análisis.

En la Figura 1 se observa el porcentaje de la información diaria obtenida de los registros de precipitación, y temperatura máxima y mínima de las 108 estaciones climatológicas. En el caso de precipitación, se tomaron dos estaciones vecinas con menor cantidad de información, para delimitar correctamente a San Luis Potosí. De igual manera, para la temperatura máxima se tomaron cuatro estaciones con información inferior al 85 %. Mientras que los registros de la temperatura mínima presentan la información requerida. Para los tres casos se tomaron las mismas estaciones seleccionadas.

En la Figura 2, se presenta la ubicación y distribución de las 108 estaciones climatológicas. También, se muestran las estaciones analizadas y auxiliares, las cuales sirvieron para definir el mejor método para la deducción de datos faltantes.

Selección de estaciones climatológicas

Se seleccionaron seis estaciones climatológicas (Cuadro 1), dos en la región Centro, una de cada región (Altiplano, Media y Huasteca) y una de las estaciones circundantes al estado, con la finalidad de delimitar y representar el área de estudio. De cada una de estas se eliminó 15 % de información de precipitación (1,489 de 9,862 de 27 años) y 25 % de

Table 1. Distance between the analyzed and auxiliary stations (km).

Cuadro 1. Distancia entre las estaciones analizadas y auxiliares (km).

Station / Estación	Auxiliary stations / Estaciones auxiliares			
Ocampo, Ocampo, Gto.: 11050 Distancia (km)	Pino Suarez, Zac.: 32127 52.34	Armadillo de los Infante, S. L. P.: 24004 107.5	El Vergel, San Luis de la Paz, Gto.: 11161 87.67	
El Salto, Cd. del Maíz, S. L. P.: 24027 Distancia (km)	Abritas, El Naranjo, S. L. P.: 24016 10.05	San Pablo, Tula, Tamps.: 28091 48.34	La Boquilla, Ocampo, Tamps.: 28043 26.33	
Los Pilares, S. L. P.: 24038 Distancia (km)	Armadillo de los Infante, SLP: 24004 49.64	Mexquitic de Carmona, SLP: 24042 19.57	El Grito, Moctezuma, S.L.P.: 24021 24.67	Espíritu Santo, Pinos, Zac.: 32187 40.86
Papagayos, Cd. del Maíz, S. L. P.: 24049 Distancia (km)	Cd. Del Maíz, S. L. P.: 24011 15.68	Abritas, El Naranjo, S. L. P.: 24016 13.72	Álvaro Obregón, Cd. del Maíz, S. L. P.: 24013 26.17	
Tierranueva, S. L. P.: 24093 Distancia (km)	Bledos, Villa de Reyes, S. L. P.: 24163 59.5	San José Alburquerque, Santa María del Río, S. L. P.: 24067 21.56	El Vergel, San Luis de la Paz, Gto.: 11161 26.02	Xichu, Gto.: 11083 64.02
Vanegas, S. L. P.: 24094 Distancia (km)	San Tiburcio, Mazapil, Zac.: 32078 61.78	Las Margaritas, Dr. Arroyo, N. L.: 19151 68.55	San José de Coronado, Real de Catorce, S. L. P.: 24165 32.58	

Figure 2 shows the location and distribution of the 108 weather stations. It also shows the analyzed and auxiliary stations, which served to define the best method for deducing the missing data.

Selection of weather stations

Six weather stations (Table 1), two in the *Centro* region, one from each of the other regions (*Altiplano*, *Medio* and *Huasteca*) and one of the stations bordering the state, were selected in order to delimit and represent the study area. In each of these 15 % of the precipitation (1,489 of 9,862 from 27 years) and 25 % of the maximum and minimum temperature (1,489 to 5,844 from 16 years) information was removed; later, in order to generate the missing data, records from the respective auxiliary stations corresponding to each analyzed station were used.

Methods for estimating missing data

The methods used were: deductive rational (DR), proposed by Campos-Aranda (1998), which is determined from the values of the record itself, and normal ratio (NR) (Aparicio, 2007), represented in Equation 1.

$$P_x = \frac{1}{n} \left[\frac{N_x}{N_a} P_a + \frac{N_x}{N_b} P_b + \dots + \frac{N_x}{N_n} P_n \right] \quad (1)$$

where P_x = missing precipitation at station x , N_x , N_a , N_b, \dots, N_n = average daily precipitation of the missing station (x) and the auxiliary stations a, b, \dots, n (averages of all historical series) and P_a, P_b, \dots, P_n = precipitation recorded at the auxiliary stations of the day where the datum is missing at station x .

The U.S. National Weather Service's inverse square distance (ISD) method (Chow, Maidment, & Mays, 1996), represented in the following equation, was also used:

$$P_x = \frac{\sum (P_i W_i)}{\sum W_i} \quad (2)$$

where: P_x = lost datum at station x and P_i = existing datum at auxiliary station i , for which $i = 1, 2, \dots, n$ for the same day.

$$W_i = \frac{1}{D_i^2} \quad (3)$$

In Equation 3, D_i = distance between each neighboring auxiliary station and station x where the lost datum is presented. Campos-Aranda (1998) recommend using four auxiliary stations (the nearest), so that each remains located in one of the quadrants that define the coordinate axes that pass through the incomplete station, generally north-south and east-west.

temperatura máxima y mínima (1,489 de 5,844 de 16 años); para posteriormente generar los datos faltantes se utilizaron los registros de las estaciones auxiliares respectivas a cada estación analizada.

Métodos para estimación de datos faltantes

Se emplearon los métodos: racional deductivo (RD), propuesto por Campos-Aranda (1998), el cual se determina a partir de los valores del propio registro, y relación normalizada (RN) (Aparicio, 2007), representado en la Ecuación 1.

$$P_x = \frac{1}{n} \left[\frac{N_x}{N_a} P_a + \frac{N_x}{N_b} P_b + \dots + \frac{N_x}{N_n} P_n \right] \quad (1)$$

donde P_x = precipitación faltante en la estación x , N_x , N_a , N_b, \dots, N_n = precipitación media diaria de la estación faltante (x) y las estaciones auxiliares a, b, \dots, n (medias de todas las series históricas) y P_a, P_b, \dots, P_n = precipitación registrada en las estaciones auxiliares del día donde falta el dato en la estación x .

También se usó el cuadrado del inverso de la distancia a U.S. National Weather Service (WS) (Chow, Maidment, & Mays, 1996), representado en la siguiente ecuación:

$$P_x = \frac{\sum (P_i W_i)}{\sum W_i} \quad (2)$$

donde: P_x = dato perdido en la estación x y P_i = dato existente en la estación auxiliar i , para lo cual $i = 1, 2, \dots, n$ para el mismo día.

$$W_i = \frac{1}{D_i^2} \quad (3)$$

En la Ecuación 3, D_i = distancia entre cada estación circundante auxiliar y la estación x donde se presenta el dato perdido. De acuerdo con Campos-Aranda (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.

Índices estadísticos para evaluar la estimación del error

Se utilizaron tres índices estadísticos: el índice de Willmott (d) (Willmott, 1981), el error estándar promedio o raíz cuadrada del cuadrado medio del error (RMSE) (Rivas & Carmona, 2010) y el coeficiente de determinación (R^2) (Cervantes-Osornio et al., 2013; Kelso-Bucio et al., 2012; Wright, 2001); los cuales se definen como:

$$d = 1 - \frac{\sum_{i=1}^N (a_i - t_i)^2}{\sum_{i=1}^N (|a_i - \bar{t}| + |t_i - \bar{t}|)^2} \quad (4)$$

Statistical indices to evaluate the error estimate

Three statistical indices were used: the Willmott index of agreement (d) (Willmott, 1981), mean standard error or root mean square error (RMSE) (Rivas & Carmona, 2010) and the coefficient of determination (R^2) (Cervantes-Osornio et al., 2013; Kelso-Bucio et al., 2012; Wright, 2001), which are defined as:

$$d = 1 - \frac{\sum_{i=1}^N (a_i - t_i)^2}{\sum_{i=1}^N (|a_i - \bar{a}| + |t_i - \bar{t}|)^2} \quad (4)$$

$$RMSE = \left(\frac{\sum_{i=1}^N (a_i - t_i)^2}{N} \right)^{1/2} \quad (5)$$

$$R^2 = \frac{[\sum_{i=1}^N (a_i - \bar{a})(t_i - \bar{t})]^2}{[\sum_{i=1}^N (a_i - \bar{a})^2][\sum_{i=1}^N (t_i - \bar{t})^2]} \quad (6)$$

where R^2 = coefficient of determination, a_i = datum estimated by the method, t_i = observed datum, N = number of observations or estimates, \bar{a} = average of the data estimated by the method and \bar{t} = average of the observed data.

If $RMSE = 0$ and $R^2 = d = 1$, then there is a perfect fit. The results are excellent when $d \geq 0.95$ and $R^2 > 0.80$ (Cai et al., 2007; Ruiz-Álvarez, Arteaga-Ramírez, Vázquez-Peña, Ontiveros-Capurata, & López-López, 2012).

Results and discussion

Table 2 shows a summary of the statistical precipitation indices, and a better fit with the ISD method with an RMSE average of 38.7 mm is observed. The d values on average were 0.85, 0.68 and 0.68 for ISD, DR and NR, respectively, which have no values greater than 0.95 as recommended by Cai et al. (2007) and Ruiz-Álvarez et al. (2012); however, the ISD method shows a better fit.

Table 2. Statistical fit indices of the methods for estimating missing precipitation data records.

Cuadro 2. Índices estadísticos del ajuste de los métodos que estiman los registros faltantes con datos de precipitación.

Station/Estación	ISD/WS			DR/RD		NR/RN		
	RMSE	d	R^2	RMSE	d	RMSE	d	R^2
Ocampo, Ocampo: 11050	19.503	0.917	0.144	41.672	0.556	32.254	0.865	0.166
El Salto, Cd. del Maíz: 24027	73.682	0.911	0.216	84.242	0.861	80.477	0.926	0.376
Los Pilares, San Luis Potosí: 24038	21.004	0.890	0.363	29.509	0.672	71.649	0.587	0.271
Papagayos, Cd. del Maíz: 24049	61.666	0.918	0.298	97.865	0.856	117.633	0.869	0.173
Tierranueva: 24093	28.728	0.845	0.335	42.390	0.638	41.627	0.728	0.184
Vanegas: 24094	27.870	0.615	0.030	24.206	0.485	78.453	0.129	0.012
Average/Promedio	38.742	0.849	0.231	53.314	0.678	70.349	0.684	0.197

$$RMSE = \left(\frac{\sum_{i=1}^N (a_i - t_i)^2}{N} \right)^{1/2} \quad (5)$$

$$R^2 = \frac{[\sum_{i=1}^N (a_i - \bar{a})(t_i - \bar{t})]^2}{[\sum_{i=1}^N (a_i - \bar{a})^2][\sum_{i=1}^N (t_i - \bar{t})^2]} \quad (6)$$

donde R^2 = coeficiente de determinación, a_i = dato estimado por el método, t_i = dato observado, N = número de observaciones o estimaciones, \bar{a} = promedio de los datos estimados por el método y \bar{t} = promedio de los datos observados.

Si $RMSE = 0$ y $R^2 = d = 1$, entonces se tiene un ajuste perfecto. Los resultados son excelentes cuando $d \geq 0.95$ y $R^2 > 0.80$ (Cai et al., 2007; Ruiz-Álvarez, Arteaga-Ramírez, Vázquez-Peña, Ontiveros-Capurata, & López-López, 2012).

Resultados y discusión

En el Cuadro 2 se muestra un resumen de los índices estadísticos de precipitación, y se observa un mejor ajuste con el método WS, con RMSE promedio de 38.7 mm. Los valores del d en promedio fueron 0.85, 0.68 y 0.68 para WS, RD y RN, respectivamente; los cuales no presentan valores mayores a 0.95 como lo recomiendan Cai et al. (2007) y Ruiz-Álvarez et al. (2012), sin embargo, el primero indica un mejor ajuste. Por otro lado, las estaciones 11050, 24027, 24038, 24049 y 24093 muestran un valor de RMSE menor con WS, en comparación con el obtenido con RD y RN; mientras que los datos analizados en la estación 24094, según el estadístico RMSE, indican un menor error con el método RD.

En un estudio realizado en Colombia, Toro-Trujillo et al. (2015) reportan valores similares entre los índices

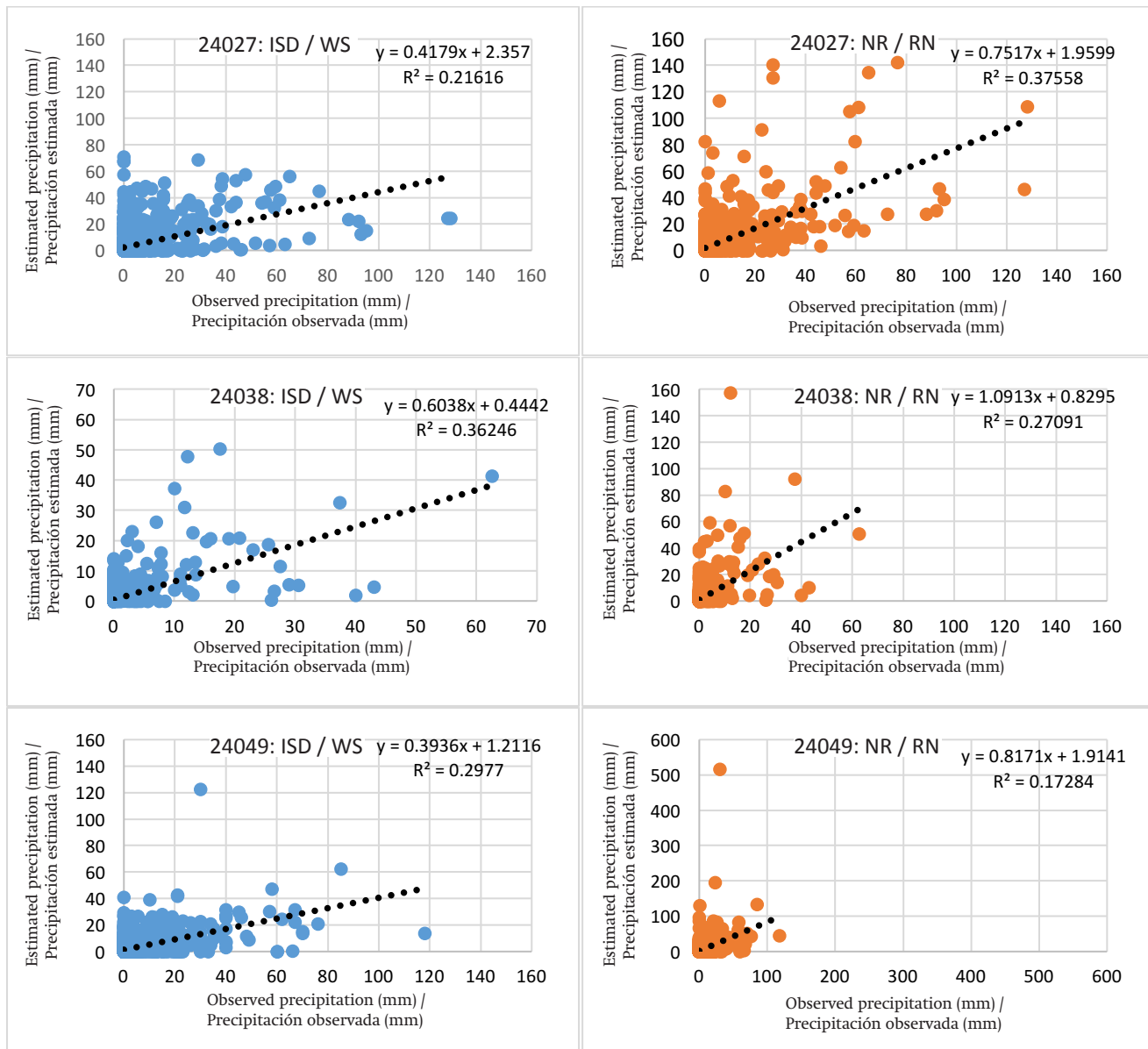


Figure 3. Comparison between the estimated and observed precipitation value.

Figura 3. Comparación entre el valor estimado y observado de precipitación.

On the other hand, stations 11050, 24027, 24038, 24049 and 24093 show a lower RMSE value with ISD, compared with that obtained with DR and NR, while the data analyzed at station 24094, according to the RMSE statistic, indicate a lower error with the DR method.

In a study conducted in Colombia, Toro-Trujillo et al. (2015) report similar values among the statistical indices when analyzing the precipitation variable, but a better Willmott index of agreement (d) result with the ISD method is highlighted; therefore, they propose it as the best method for estimating missing data, a situation also presented in this study, since the best estimate was obtained with the ISD method. Moreover, considering that R^2 averaged 0.231 and 0.197 for ISD and NR, respectively, it can be deduced that there is little strength in the linear relationship between the estimated and observed value, mainly

estadísticos al analizar la variable precipitación, aunque resalta un mejor resultado del índice de Willmott (d) con el método WS, por ello lo proponen como el mejor método para la estimación de datos faltantes; situación que se presenta en este estudio, ya que se observa que la mejor estimación se obtuvo con WS. Además, considerando que en promedio R^2 fue de 0.231 y 0.197 para WS y RN, respectivamente, se puede deducir que existe poca fuerza de la relación lineal entre el valor observado y estimado. Debido, principalmente, a que la precipitación es una variable discreta con alta frecuencia de valores cercanos a cero.

En la Figura 3 se ilustra la relación entre los valores de precipitación observados y estimados, con los métodos WS y RN en las estaciones climatológicas 24027 El Salto, 24038 Los Pilares y 24049 Papagayos. El método WS presenta, en promedio, un mejor ajuste de los valores

because precipitation is a discrete variable with a high frequency of values close to zero.

Figure 3 illustrates the relationship between observed and estimated precipitation values, with the ISD and NR methods at weather stations 24027 El Salto, 24038 Los Pilares and 24049 Papagayos. The ISD method has, on average, a better fit of the estimated values with respect to the observed ones, which is consistent with the results obtained by Hubbard and You (2005), Eischeid, Pasteris, Diaz, Plantico, and Lott (2000) and You, Hubbard, and Goddard (2008). Moreover, Teegavarapu and Chandramouli (2005) mention that the ISD method better estimates the observed data, plus it can be used in any time period, i.e., in daily, monthly and annual data.

Figure 4 shows the relationship between the observed and estimated maximum temperature with the ISD

estimados con respecto de los observados. Datos que concuerdan con los resultados obtenidos por Hubbard y You (2005), Eischeid, Pasteris, Diaz, Plantico, y Lott (2000) y You, Hubbard, y Goddard (2008). Por otro lado, Teegavarapu y Chandramouli (2005) mencionan que el método WS estima mejor el dato observado, además de que se utiliza en cualquier periodo de tiempo, es decir, en datos diarios, mensuales y anuales.

En la Figura 4 se presenta la relación entre la temperatura máxima observada y estimada con los métodos WS y RN; en ambos se puede observar un comportamiento similar, posiblemente porque la temperatura es una variable más localizada (Toro-Trujillo et al., 2015; You et al., 2008).

En el Cuadro 3 se muestran los índices estadísticos calculados con los tres métodos descritos para determinar cuál permite estimar de manera más precisa los datos faltantes de los registros de

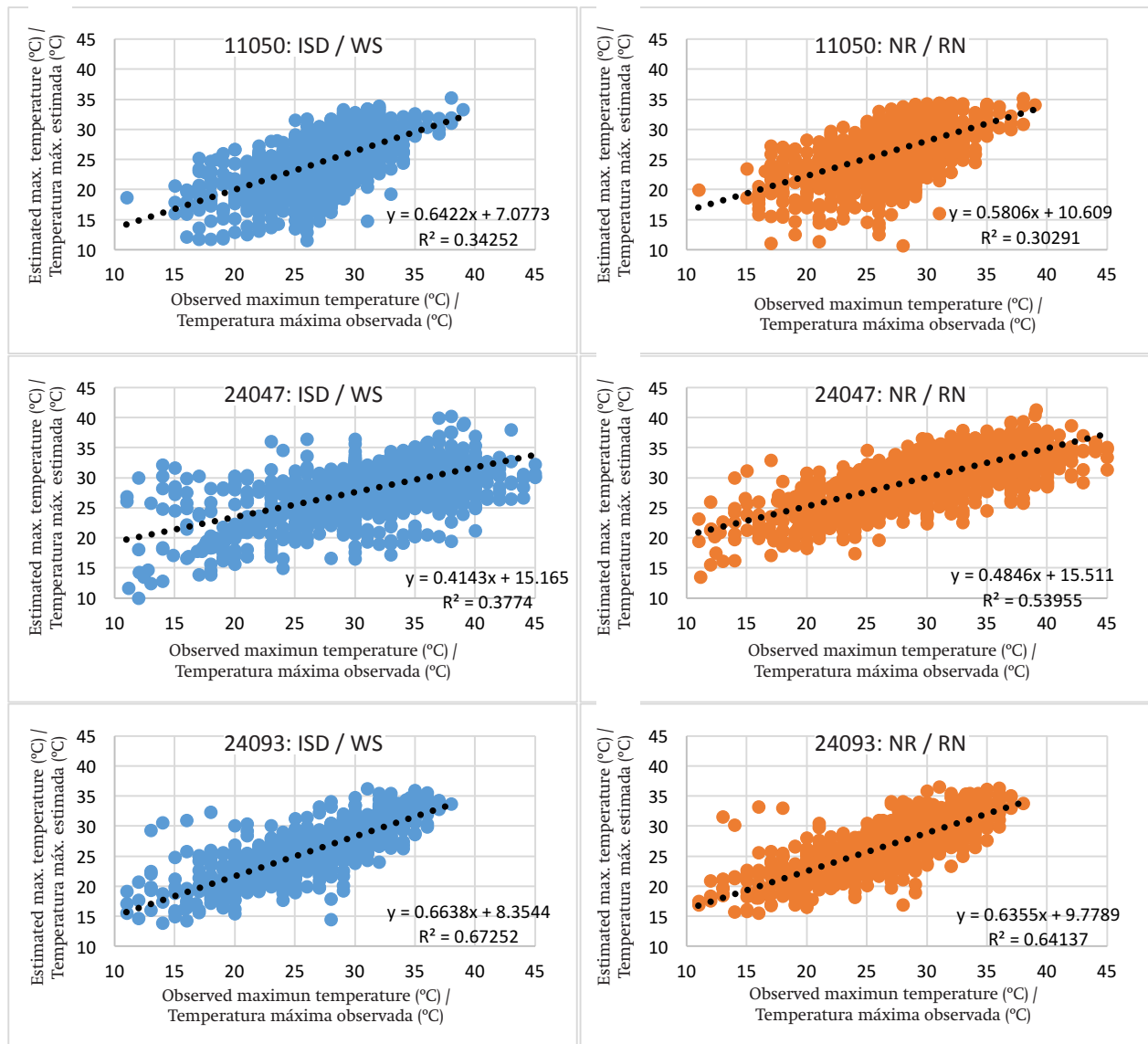


Figure 4. Comparison between the estimated and observed maximum temperature value.

Figura 4. Comparación entre el valor estimado y observado de temperatura máxima.

Table 3. Statistical fit indices of the methods for estimating missing maximum temperature data records.

Cuadro 3. Índices estadísticos del ajuste de los métodos que estiman los registros faltantes con datos de temperatura máxima.

Station/Estación	ISD/WS			DR/RD		NR/RN		
	RMSE	d	R ²	RMSE	d	RMSE	d	R ²
Ocampo, Ocampo: 11050	3.264	0.685	0.343	3.620	0.582	1.761	0.843	0.303
El Salto, Cd. del Maíz: 24027	3.744	0.769	0.377	2.692	0.803	2.670	0.880	0.540
Los Pilares, San Luis Potosí: 24038	3.938	0.707	0.493	2.282	0.887	3.244	0.769	0.501
Papagayos, Cd. del Maíz: 24049	2.823	0.751	0.251	2.177	0.845	1.816	0.864	0.285
Tierranueva: 24093	1.277	0.937	0.673	1.933	0.888	1.133	0.952	0.641
Vanegas: 24094	2.905	0.801	0.467	2.194	0.891	1.938	0.897	0.463
Average/Promedio	2.992	0.780	0.434	2.150	0.866	2.094	0.867	0.455

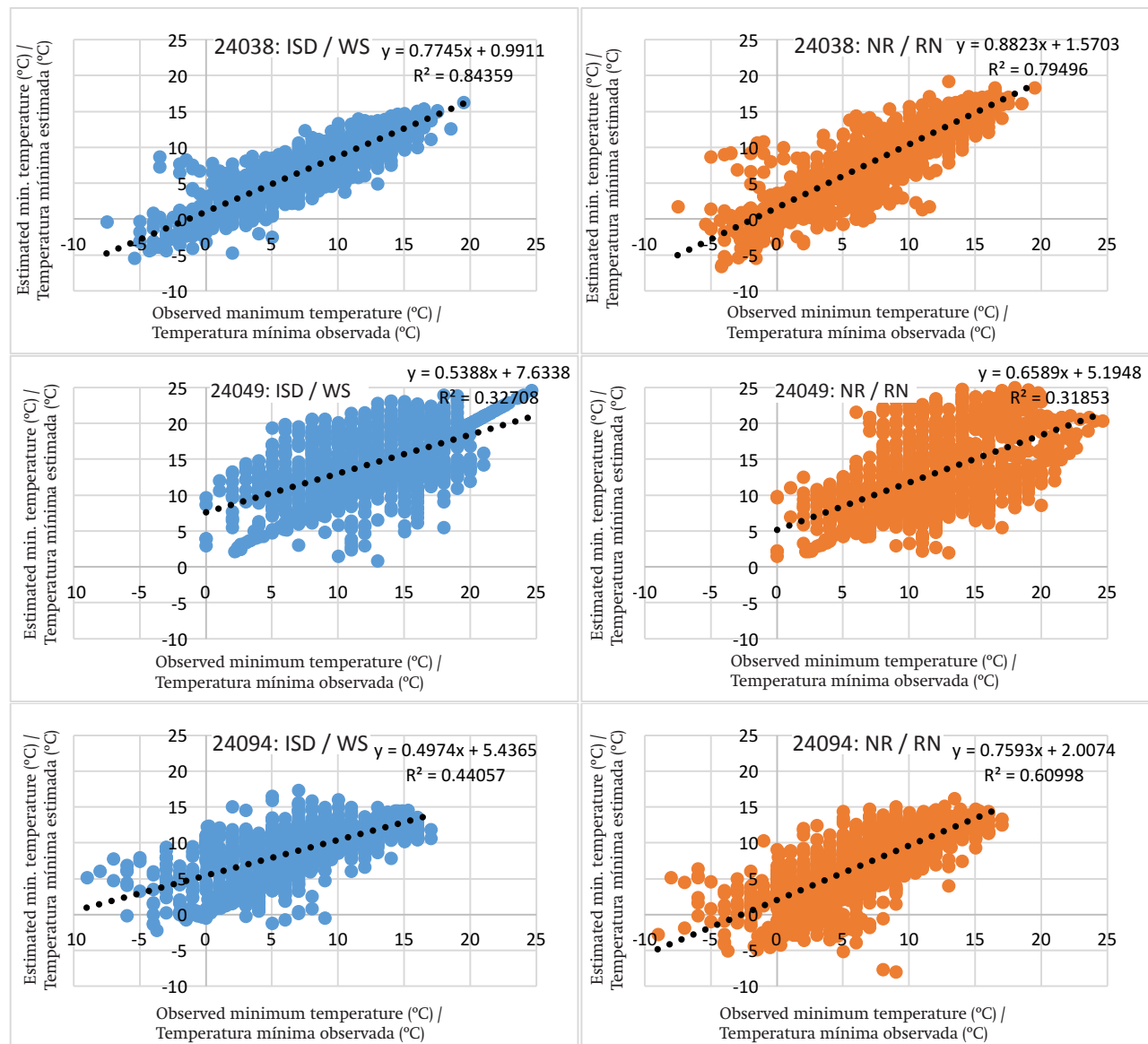


Figure 5. Comparison between the estimated and observed minimum temperature.

Figura 5. Comparación entre el valor estimado y observado de temperatura mínima.

Table 4. Statistical fit indices of the methods for estimating missing minimum temperature data records.**Cuadro 4. Índices estadísticos del ajuste de los métodos que estiman los registros faltantes con datos de temperatura mínima.**

Station/Estación	ISD/WS			DR/RD		NR/RN		
	RMSE	<i>d</i>	R ²	RMSE	<i>d</i>	RMSE	<i>d</i>	R ²
Ocampo, Ocampo: 11050	1.019	0.975	0.794	1.031	0.968	0.944	0.978	0.764
El Salto, Cd. del Maíz: 24027	3.097	0.849	0.447	1.921	0.948	1.979	0.944	0.706
Los Pílares, San Luis Potosí: 24038	1.450	0.967	0.844	1.578	0.962	1.494	0.970	0.795
Papagayos, Cd. del Maíz: 24049	3.710	0.761	0.327	2.661	0.854	4.110	0.755	0.319
Tierranueva: 24093	1.261	0.969	0.810	1.833	0.942	2.275	0.925	0.711
Vanegas: 24094	3.307	0.801	0.441	1.569	0.960	1.829	0.947	0.610
Average / Promedio	2.318	0.887	0.611	1.766	0.939	2.105	0.920	0.651

and NR methods; similar behavior can be observed in the two methods, possibly because temperature is a more localized variable (Toro-Trujillo et al., 2015; You et al., 2008).

Table 3 shows the statistical indices calculated with the three described methods to determine which one estimates more accurately the missing maximum temperature data records. It also shows that the values are similar to one other, although the RMSE results are higher than those with the ISD method, followed by the DR and finally the NR methods. According to the Willmott index, the NR method has a better fit because, in general, it has a higher value at the stations, followed by the DR and ISD methods, respectively.

At stations 11050, 24038, 24049 and 24093, the coefficient of determination R^2 is greater with the ISD method, and with the others it is greater with NR, although the R^2 are very similar. Campos-Aranda (2005) recommends against using low linear relationship methods if the coefficients of determination do not have values greater than 0.8; therefore, because the present study presents a low linear relationship, it is advisable to use either the ISD or NR methods, since they have similar statistics.

Figure 5 compares the observed and estimated minimum temperature with the ISD and NR methods, at the stations studied.

Table 4 shows the statistics that allow identifying the method that has the best fit to the minimum temperature data. The Willmott index is higher with DR at stations 24027, 24049 and 24094, with NR at 11050 and 24038, and with ISD at 24093. It also shows that the d value varies in a very similar way in the three methods. Stations 24027, 24049 and 24094 reflected lower RMSE values with the DR

temperatura máxima. Asimismo, se observa que los valores son similares entre sí, aunque los resultado de RMSE más grades se tienen con el método WS, seguido del RD y finalmente el RN. De acuerdo con el índice de Willmott, el método RN presenta un mejor ajuste dado que, en general, se tiene un valor mayor en las estaciones, seguido del RD y WS, respectivamente.

En las estaciones 11050, 24038, 24049 y 24093, el coeficiente de determinación R^2 es mayor con el método WS, y con los otros es mayor con RN. Aunque los R^2 son muy parecidos. Campos-Aranda (2005) recomienda no utilizar métodos de relación lineal baja si los coeficientes de determinación no tienen valores mayores a 0.8; por lo que, debido a que en este estudio se presenta baja relación lineal, se recomienda utilizar el WS o el RN, ya que presentan estadísticos similares.

La Figura 5 compara la temperatura mínima observada y estimada con los métodos WS y RN, en las estaciones estudiadas.

En el Cuadro 4, se muestran los estadísticos que permiten identificar el método que presenta un mejor ajuste a los datos de temperatura mínima. El índice de Willmott es más alto con el RD en las estaciones 24027, 24049 y 24094, con RN en 11050 y 24038, y en la estación 24093 con el WS. Asimismo, se identifica que el valor de d varía de manera muy semejante en los tres métodos. Las estaciones 24027, 24049 y 24094 plasman valores de RMSE menores con el método RD, mientras que 24038 y 24093 con el método WS, y la estación 11050 con RN.

Conclusiones

Los métodos utilizados para estimar los datos faltantes de los registros de precipitación permitieron

method, whereas 24038 and 24093 did so with the ISD method, and 11050 with the NR method.

Conclusions

The methods used to estimate the missing precipitation data records allowed identifying the one which best estimates the results, and according to the statistical indicators it is the ISD method; therefore, for precipitation, this is the best alternative when data are available for the same period at auxiliary stations. The results obtained by analyzing the maximum and minimum temperature data indicate little variability among methods, as the calculated data are similar to those observed; therefore, both the behavior and errors are very similar with all three methods. Accordingly, the U.S. National Weather Service's inverse square distance (ISF) method is an efficient alternative for completing missing data records in a continuous series; therefore, it was the method used to estimate missing data at 108 weather stations in the state of San Luis Potosí.

End of English version

References / Referencias

- Alfaro, R., & Pacheco, R. (2000). Aplicación de algunos métodos de relleno a series anuales de lluvia de diferentes regiones de Costa Rica. *Top. Meteor. Oceanogr*, 7(1), 1-20.
- Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (2006). *Evapotranspiración del cultivo: Guías para la determinación de los requerimientos de agua de los cultivos* (pp. 323). Roma: Ed. FAO.
- Aparicio, M. F. J. (2007). *Fundamentos de hidrología de superficie* (pp. 304). México, D.F.: Editorial Limusa.
- Cai, J., Liu, Y., Lei, T., & Pereira, S. L. (2007). Estimating reference evapotranspiration with the FAO Penman-Monteith equation using daily weather forecast messages. *Agricultural and Forest Meteorology*, 145(1), 22-35. doi: 10.1016/j.agrformet.2007.04.012
- Campos-Aranda, D. F. (2015). Estimación simultánea de datos hidrológicos anuales faltantes en múltiples sitios. *Revista Ingeniería Investigación y Tecnología*, 16(2), 295-306. doi: 10.1016/j.riit.2015.03.013
- Campos-Aranda, D. F. (2005). *Agroclimatología cuantitativa de cultivos* (pp. 320). México, D.F.: Trillas.
- Campos-Aranda, D. F. (1998). *Procesos del ciclo hidrológico* (pp. 529). México, San Luis Potosí: Editorial Universitaria Potosina. Universidad Autónoma de San Luis Potosí.
- Cervantes-Osornio, R., Arteaga-Ramírez, R., Vázquez-Peña, M. A., Ojeda-Bustamante, W., & Quevedo-Nolasco, A. (2013). Modelos Hargreaves Priestley-Taylor y redes neuronales artificiales en la estimación de la evapotranspiración de referencia. *Revista Ingeniería Investigación y Tecnología*, 16(2), 163-176. Retrieved from <http://pruebaredalyc.redalyc.org/articulo.oa?id=40426155002>
- Chow, V. T., Maidment, D. R., & Mays, L. W. (1996). *Hidrología aplicada* (pp. 583). Colombia, Bogotá: McGraw Hill
- Comisión Nacional del Agua (CONAGUA). (2015). *Red de estaciones climatológicas: Estaciones climatológicas.KML*. Servicio Meteorológico Nacional (SMN)-Comisión Nacional del Agua (CONAGUA). México. D.F.: Author.
- DeGaetano, A. T., Eggleson, K. L., & Knapp, W. W. (1995). A method to estimate missing daily maximum and minimum temperature observations. *Journal of Applied Meteorology*, 34(2), 371-380. doi: 10.1175/1520-0450-34.2.371
- Eischeid, J. K., Pasteris, P. A., Diaz, H. F., Plantico, M. S., & Lott, N. J. (2000). Creating a serially complete, national daily time series of temperature and precipitation for the Western United States. *Journal of Applied Meteorology*, 39(9), 1580-1591. doi: 10.1175/1520-0450(2000)039<1580:CASCND>2.0.CO;2
- Gutiérrez, L. W., García, E. M., Magaña, R. V. O., & Escalante, G. J. A. (2007). Diseño y construcción de un globo meteorológico cautivo instrumentado. *Ingeniería y Ciencia*, 3(5), 29-42. Retrieved from <http://www.redalyc.org/articulo.oa?id=83530502>
- Hubbard, K. G., & You, J. (2005). Sensitivity Analysis of Quality Assurance Using the Spatial Regression Approach. A case study of the Maximum/Minimum air temperature. *Journal of Atmospheric and Oceanic Technology*, 22(10), 1520-1530. doi: 10.1175/JTECH1790.1
- Instituto Mexicano de Tecnología del Agua (IMTA). (2009). *ERIC III: Extractor rápido de información climatológica (versión 2.0. Disco)*. Comisión Nacional del Agua-Secretaría de Medio Ambiente y Recursos Naturales-IMTA. Jiutepec, Morelos, México: Author.
- Instituto Nacional de Estadística y Geografía (INEGI). (2007). *Regiones agropecuarias de San Luis Potosí* (pp. 69). México: Author.
- Instituto Nacional para el Federalismo y Desarrollo Municipal (INAFED). (2010). *Enciclopedia de los*

Fin de la versión en español

- Municipios y Delegaciones de México; San Luis Potosí. 09-04-2015. Retrieved from <http://www.inafed.gob.mx/work/enciclopedia/EMM24sanluispotosi/index.html>
- Kelso-Bucio, H. A., Bâ, K. M., Sánchez-Morales, S., & Reyes-López, D. (2012). Calibración del exponente de la ecuación Hargreaves-ETo en los estados de Chiapas, Oaxaca, Puebla y Veracruz, México. *Agrociencia*, 46(3), 221-229. Retrieved from <http://www.scielo.org.mx/pdf/agro/v46n3/v46n3a2.pdf>
- McCuen, R. H. (1998). *Hydrologic Analysis and Design* (pp. 833). Department of Civil Engineering University of Maryland, New Jersey: Pearson Prentice-Hall.USA.
- Nájera, M. N. L., & Arteaga, R. R. (1998). *Antología Agroclimatología* (pp. 230). México: Dirección General de Educación Tecnológica Agropecuaria (DGETA).
- Organización Meteorológica Mundial (OMM). (2011). *Guía de prácticas climatológicas* (pp. 128). Ginebra: Author.
- Puertas-Orozco, O. L., Carvajal-Escobar, Y., & Quintero-Angel, M. (2011). Estudio de tendencias de precipitación mensual en la cuenca alta-media del río Cauca, Colombia. *Dyna*, 78(169), 112-120. Retrieved from <http://www.redalyc.org/articulo.oa?id=49622390013>
- Ramírez-Cruz, H., López-Velasco, O., & Ibáñez-Castillo, L. A. (2015). Estimación mensual de intensidad de la lluvia en 30 minutos a partir de datos pluviométricos. *Terra Latinoamericana*, 33(2), 151-159. Retrieved from <https://www.researchgate.net/publication/279450618>
- Rivas, R., & Carmona, F. (2010). La ecuación de Priestley-Taylor aplicada a nivel de píxel: una alternativa para estudios detallados de cuencas. *Boletín Geológico y Minero*, 121(4), 401-412. Retrieved from http://www.igme.es/Boletin/2010/121_4/8-ARTICULO%205.pdf
- Ruiz-Álvarez, O., Arteaga-Ramírez, R., Vázquez-Peña, M. A., Ontiveros-Capurata, R. E., & López-López, R. (2012). Balance hídrico y clasificación climática del estado de Tabasco, México. *Universidad y Ciencia, Trópico Húmedo*, 28(1), 1-14. Retrieved from <https://www.researchgate.net/publication/262548360>
- 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(1-4), 191-206. doi: 10.1016/j.jhydrol.2005.02.015
- Toro-Trujillo, A. M., Arteaga-Ramírez, R., Vázquez-Peña, M. A., & Ibáñez-Castillo, L. A. (2015). Relleno de series de precipitación, temperatura mínima, máxima de la región norte del Urabá Antioqueño. *Revista Mexicana de Ciencias Agrícolas*, 6(2), 577-588. Retrieved from <https://www.researchgate.net/publication/277248839>
- Wagner, D. P., Fiener, P., Wilken, F., Kumar, S., & Schneider, K. (2012). Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. *Journal of Hydrology*, 464, 388-400. doi: 10.1016/j.jhydrol.2012.07.026
- Willmott, C. J. (1981). On the validation of models. *Physical geography*, 2(2), 184-194. doi: 10.1080/02723646.1981.10642213
- Wright, J. (2001). Correlaciones de la fracción difusa. *Top. Meteor. Oceanog*, 8(1), 15-18. Retrieved from <http://www.imn.ac.cr/publicaciones/revista/2001/Julio/3-CWrith-Julio01.pdf>
- You, J. S., Hubbard, K. G., & Goddard, S. (2008). Comparison of methods for spatially estimating station temperatures in a quality control system. *International Journal of Climatology*, 28(6), 777-787. doi: 10.1002/joc.1571