

## Relleno de series diarias de precipitación, temperatura mínima, máxima de la región norte del Urabá Antioqueño\*

### Fill of daily series of precipitation, minimum and maximum temperature from the northern region of Urabá Antioquia

Ana María Toro Trujillo<sup>1</sup>, Ramón Arteaga Ramírez<sup>1§</sup>, Mario Alberto Vázquez Peña<sup>1</sup> y Laura Alicia Ibáñez Castillo<sup>1</sup>

<sup>1</sup>Ingeniería Agrícola y Uso Integral del Agua (IAUIA)-Universidad Autónoma Chapingo (UACH). Carretera México-Texcoco, Estado de México. C. P. 56230. (namariatoro@hotmail.com; mvazquezd@correo.chapingo.mx; libacasa@gmail.com). <sup>§</sup>Autor para correspondencia: arteagar@correo.chapingo.mx.

#### Resumen

Los estudios agroclimáticos requieren la utilización de series cronológicas de las variables meteorológicas, que generalmente presentan datos faltantes limitando su uso, por lo que los métodos de relleno son utilizados. La finalidad de este trabajo fue determinar la confiabilidad de los métodos de relleno: U.S National Weather Service (WS), deductivo racional (RD), la regresión múltiple (RM) y regresión lineal (RL) y de ellos utilizar el mejor para rellenar los datos de las series de precipitación, temperatura máxima y temperatura mínima en el periodo 2006-2009 de las estaciones ubicadas en la zona Norte del eje bananero del Urabá Antioqueño. La estación base fue la estación Pista Indira, y las estaciones Uniban, Turbo, Aeropuerto y Prado Mar se utilizaron como vecinas. La estimación del error se hizo con la: raíz cuadrada del cuadrado medio del error (RCCME), coeficiente de determinación ( $R^2$ ), error medio absoluto (MAE), error relativo (RE) e Índice de concordancia de Willmott (d). Los métodos RM y RL presentaron RCCME y MAE similares a los del WS, lo que con llevaría a tener errores similares pero debido a que la  $R^2$  de los primeros fueron menores de 0.8 y el índice (d) para WS fue mayor o similar a lo demás método, se utilizó este último para el relleno de las series.

**Palabras claves:** deductivo, distancia inversa, regresión.

#### Abstract

Agroclimatic studies require the use of time series of meteorological variables, which generally present missing data limiting its use, so filling methods are used. The purpose of this study was to determine the reliability of filling methods: US National Weather Service (WS), rational deductive (RD), multiple regression (MR) and linear regression (LR) and from them use the best to fill the data series of precipitation, maximum and minimum temperature during the period 2006-2009 of the stations located in the northern area on banana axis from Urabá Antioquia. The base station was Pista Indira, and Uniban, Turbo, Aeropuerto and Prado Mar stations were used as neighbors. Standard error was made with: square root of the mean squared error (RCCME), coefficient of determination ( $R^2$ ), mean absolute error (MAE), relative error (RE) and Willmott's concordance index (d). RM and RL methods presented similar RCCME and MAE to those of WS, which would lead to have similar errors but due to  $R^2$  of the first were 0.8 and the index (d) for WS was higher or similar to the other method, the latter was used to fill in the series.

**Keywords:** deductive, inverse distance, regression.

\* Recibido: agosto de 2014  
Aceptado: enero de 2015

## Introducción

El clima es un recurso natural que afecta la producción agraria, por lo tanto es importante tener conocimiento de los recursos disponibles del ambiente. Las interacciones que se producen en el área del cultivo, debajo de la superficie del suelo, la interfase suelo-aire y la capa límite de la atmósfera, proporciona una guía esencial para las decisiones estratégicas en planificación a largo plazo de los sistemas agrícolas (WMO, 1981).

En las estaciones climatológicas es común que los registros estén incompletos situación que se presenta en la zona de estudio pues la información meteorológica es escasa, lo que ha limitado realizar estudios agroclimáticos cuyos resultados permitan aumentar la productividad, optimizar los recursos, reducir el riesgo de pérdidas en cosecha, planificación integral de la infraestructura de riego y drenaje y la posible predicción climática.

Para realizar un estudio que permita conocer de los elementos meteorológicos su variabilidad, frecuencia y probabilidad de valores críticos y relacionarlos con la exigencias agroclimáticas del cultivo se requiere que los registros históricos sean continuos y con coherencia, de esta manera se minimicen los riesgos de resultados erróneos con el fin de contar con datos precisos para evaluar y pronosticar los rendimientos de los cultivos y de la producción, por tanto para los alcances y nivel de detalle del presente trabajo es importante estimar los valores faltantes con el fin de minimizar los errores y no sesgar los resultados (Masseti, 2013).

En la guía de prácticas climatológicas de la Organización Meteorológica Mundial (WMO, 1983), se proponen los siguientes métodos estadísticos para el relleno de datos faltantes: regresión simple, múltiple (Degaetano *et al.*, 1995), razón q y razón-normal q (RN) (Paulhus y Kohler, 1952). Young (1992) presenta una modificación al método RN, donde la razón-normal ponderada es cambiada por la correlación entre las estaciones (NRWC) pero Yozgatligil *et al.* (2013) reportan que el promedio aritmético simple (SAA) y (NRWC) son métodos que son idénticos en algunos casos, si las correlaciones de estaciones de referencia son casi la misma.

McCuen (1998) recomienda el método promedio aritmético simple (SAA), cuando el valor anual en cada uno de los datos a promediar difiere por lo menos de 10%, pero según Yozgatligil (2013) presenta estimaciones fiables si la

## Introduction

The climate is a natural resource affecting agricultural production, so it is important to have knowledge of available resources of the environment. The interactions that occur in the area of crops, underneath the soil, soil-air interface and the boundary layer of the atmosphere, provides an essential guide for strategic decisions on long-term planning of agricultural systems (WMO, 1981).

In the weather stations is common that records are incomplete, situation present in the study area because weather information is scarce, which has limited agro-climatic studies whose results allow to to increase productivity, optimize resources, reduce the risk of losses at harvest, comprehensive planning of irrigation and drainage infrastructure and possible climate prediction.

To conduct a study that allows to know the meteorological elements, its variability, frequency and probability of critical values and relate them to the agro-climatic requirements of the crop requires historical records to be continuous and coherent, so the risk of erroneous results are minimized in order to have accurate data to assess and predict crop yields and production, for both scope and level of detail of the present work is important to estimate the missing values in order to minimize errors and avoid bias results (Masseti, 2013).

The Guide to Climatological Practices from the World Meteorological Organization (WMO, 1983), proposes the following statistical methods to fill in missing data: simple regression, multiple regression (Degaetano *et al.*, 1995), reason q and reason-normal q (RN) (Paulhus and Kohler, 1952). Young (1992) introduced an modification to RN method, where reason-normal weighted is changed by the correlation between stations (NRWC) but Yozgatligil *et al.* (2013) report that simple arithmetic average (SAA) and (NRWC) are methods that are identical in some cases, if the correlations of reference stations are almost the same.

McCuen (1998) recommends the simple arithmetic average (SAA) method, when the annual value in each data to average differs at least 10%, but according to Yozgatligil (2013) presents reliable estimates if the variable has no spatial variability and if the reference stations are

variable no tiene variabilidad espacial y si las estaciones de referencia están altamente correlacionadas. ASCE (1996), propone el método de ponderación de distancia inversa IDW, también conocido como el método U.S National Weather Service por su implementación en estudios hidrológicos y geográficos al igual que estudios realizados por Hubbard (1994); Sokol y Stekl (1994); Palomino y Martin (1995); Teegavarapu y Chandramouli (2005).

Aparicio (2011) y Campos (1998) indican que este método puede emplearse cuando se basa en registros simultáneos de tres estaciones que se encuentren lo más cerca posible a la estación en estudio. Una ventaja importante de la IDW es que se utiliza en cualquier paso de tiempo (Teegavarapu y Chandramouli, 2005). Existen otros métodos estadísticos entre ellos el análisis de componentes principales y el análisis de grupo (Huth y Nemesova, 1995), el método de Kriging (Saborowski y Stock, 1994) y la interpolación óptima (Bussieres y Hogg, 1989). Wagner *et al.* (2012) sugieren que estos no reporta mejoras con respecto a el inverso de la distancia (IDW) y que para su empleo requieren una cantidad suficiente de datos para producir un semivariograma fiable.

La regresión (mínimos cuadrados convencionales) (REG), se presentó como un método alternativo para utilizar en este estudio, implementado en las siguientes investigaciones para estimar los valores de datos faltantes de precipitación y temperatura (Wade, 1987; Eischeid *et al.*, 2000 y Hubbard, 2001). Xia *et al.* (1999); Kemp *et al.* (1983); Young (1992); Degaetano *et al.* (1995); Eischeid *et al.* (1995); You *et al.* (2008); Presti *et al.* (2010) y Kashani y Dinpashoh (2012) manifiestan que la REG es un método superior entre los tradicionales para las variables temperatura mínima, máxima y precipitación en diferentes condiciones climáticas.

Campos (1998) proponen el método empírico racional deductivo cuando no se disponen estaciones cercanas y se tiene más de 10 años de registros de la estación de interés, empleado por Puertas *et al.* (2011) y Guevara (2003) para estimar los datos faltantes de precipitación. Jiménez *et al.* (2004) utilizaron el método regresión lineal con las siguientes condiciones: distancia menor de 25 km, altitud de  $\pm 30$  m y con el mismo tipo de clima, y cuando no se cumplían utilizó el método Racional Deductivo. La finalidad de este trabajo fue determinar las confiabilidad de los métodos de relleno: U.S National Weather Service, deductivo racional, regresión múltiple y simple y utilizar el mejor para rellenar los datos faltantes de las series: precipitación, temperatura máxima, temperatura mínima de las estaciones dentro de la zona de estudio.

highly correlated. ASCE (1996) proposes the inverse distance weighting IDW, also known as the US National Weather Service method for its implementation in hydrological and geographical studies as well as studies made by Hubbard (1994); Sokol and Stekl (1994); Palomino and Martin (1995); Teegavarapu and Chandramouli (2005).

Aparicio (2011) and Campos (1998) indicate that this method can be used when it is based on simultaneous recordings of three stations that are as close as possible to the station under study. An important advantage of IDW is that is used at any time (Teegavarapu and Chandramouli, 2005). There are other statistical methods including principal component analysis and cluster analysis (Huth and Nemesova, 1995), the Kriging method (Saborowski and Stock, 1994) and the optimal interpolation (Bussieres and Hogg, 1989). Wagner *et al.* (2012) suggest that these do not report improvements regarding to the inverse distance weighting (IDW) and for its use requires a sufficient amount of data to produce a reliable semivariogram.

Regression (conventional minimum square) (REG), was presented as an alternative method to use in this study, implemented in the following investigations to estimate missing data of precipitation and temperature (Wade, 1987; Eischeid *et al.*, 2000 and Hubbard, 2001). Xia *et al.* (1999); Kemp *et al.* (1983); Young (1992); Degaetano *et al.* (1995); Eischeid *et al.* (1995); You *et al.* (2008); Presti *et al.* (2010) and Kashani and Dinpashoh (2012) state that REG is a superior method between the traditional ones for minimum and maximum temperature and precipitation variables under different climatic conditions.

Campos (1998) proposed the rational deductive empirical method when no stations are near and has over 10 years of records from the station of interest, used by Puertas *et al.* (2011) and Guevara (2003) to estimate missing data of precipitation. Jiménez *et al.* (2004) used the linear regression method with the following conditions: within 25 km, altitude  $\pm 30$  m and with the same kind of weather, and when conditions were not met, the rational deductive method was used. The purpose of this study was to determine the reliability of fill in methods: US National Weather Service, rational deductive, simple and multiple regression and use the best to fill in the missing data of the series: precipitation, maximum temperature, minimum temperature from the stations within the study area.

## Materiales y métodos

### Localización de zona de estudio

La zona bananera de Urabá, se localiza al Noroccidente del departamento de Antioquia e inicia al Oriente del municipio de Chigorodó, cubriendo los municipios de Carepa, Apartadó y Turbo, limita al Norte con el Río Guadualito en el corregimiento el Tres, al Sur con la comunal Santillana, al Oriente con el pie de monte de la serranía del Abibe y al Occidente con Río León y el Golfo de Urabá, está entre los  $7^{\circ} 43.779'$  a los  $8^{\circ} 01.035'$  de latitud norte y  $76^{\circ} 36.677'$  y  $76^{\circ} 44.802'$  de longitud Oeste (Salazar, 2012), la zona de estudio está comprendida entre los ríos Currulao y Apartadó (Figura 1).

### Estaciones meteorológicas utilizadas

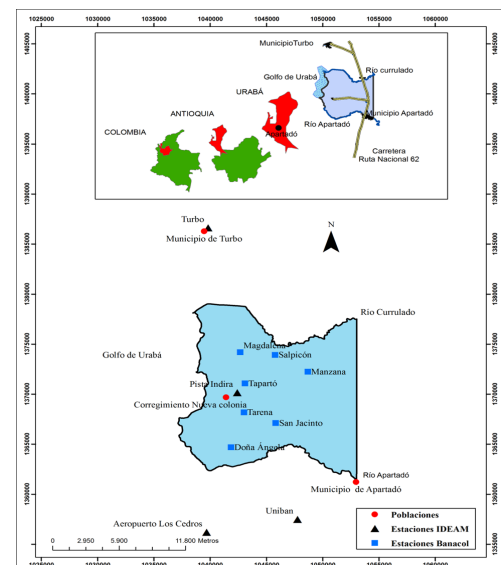
La información contó con datos diarios de las variables temperatura mínima, máxima y precipitación en el Cuadro 1 y en Figura 1 se presentan la ubicación de las estaciones dentro de la zona de estudio, la estación Pista Indira tiene un porcentaje de datos perdidos menor a las demás estaciones y su información fue proporcionada por el IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia) y la pérdida de información de las estaciones cuya información fue proporcionada por Comercializadora Internacional Banacol (C. I Banacol) tuvo un patrón de pérdida asociado a los días no laborables de las fincas en donde se encuentra ubicadas las estaciones climatológicas. En el Cuadro 2 se reportan las características de las estaciones del IDEAM ubicadas fuera de la zona de estudio (Figura 1) y en comparación con las estaciones de C.I Banacol (Cuadro 1) tienen un porcentaje de pérdida menor.

Para definir el método que presenta el error más pequeño, a la estación Pista Indira se le eliminó 31% de la información según el patrón de pérdida de las estaciones de C. I Banacol y para generarla se utilizaron las estaciones que se presentan en el Cuadro 2 (ubicadas fuera de la zona de estudio) y que se denominaron vecinas. Una vez que se definió el método con el que se tuvo el error más pequeño con este, se rellenaron las series de las estaciones C.I Banacol. En Cuadro 3 se presentan las estaciones vecinas (IDEAM) con una distancia menor a 24 km, localizadas dentro de la misma cuenca; asimismo, no tienen diferencia en altura mayor a los 20 m (Cuadro 2).

## Materials and methods

### Study area

The banana region from Urabá, is located to the Northwest of Antioquia and starts East from the Township of Chigorodó, covering the municipalities of Carepa, Apartadó and Turbo, bordered to the north by the river Guadualito in the village el Tres, south to the communal Santillana, the East with the foothills of the mountains from Abibe and west with Rio Leon and Gulf of Urabá, between  $7^{\circ} 43.779'$  at  $8^{\circ} 01.035'$  north latitude and  $76^{\circ} 36.677'$  and  $76^{\circ} 44.802'$  west longitude (Salazar, 2012), the study area is between the rivers Currulao and Apartadó (Figure 1).



**Figura 1. Zona de estudio (Norte del eje Bananero) y ubicación de las estaciones de IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia) y C. I Banacol (Comercializadora Internacional Banacol).**

**Figure 1. Study area (North from Banana axis) and location of stations from IDEAM (Institute of Hydrology, Meteorology and Environmental Studies of Colombia) and C.I. Banacol (Banacol International Marketing).**

### Meteorological stations used

The information counted with daily data of the variables minimum and maximum temperature and precipitation, Table 1 and Figure 1 shows the location of the stations within the study area, the station Pista Indira has a lower

**Cuadro 1. Ubicaciones y porcentajes de información de temperatura y precipitación en las estaciones de la zona de estudio.**  
**Table 1. Locations and percentages of information of temperature and precipitation in the stations from the study area.**

Estación	Altitud (m)	Latitud (grados)	Longitud (grados)	Precipitación		Temperatura	
				Periodo	(%) datos perdidos	Periodo	(%) datos perdidos
Doña Ángela*	9	7.89	-76.7	2004-2012	22	2004-2012	23
Magdalena*	11	7.98	76.69	2004-2012	25	2004-2012	25
Manzano*	26	7.96	-76.64	2004-2012	22	2004-2012	22
Salpicón*	22	7.98	-76.66	2004-2012	27	2004-2012	28
San Jacinto*	19	7.91	-76.66	2004-2012	28	2004-2012	28
Tapartó*	11	7.95	-76.68	2004-2012	27	2004-2012	26
Tarena*	10	7.93	-76.69	2004-2012	24	2004-2012	24
Pista Indira+	23	7.95	-76.7	2006-2009	2	2004-2009	3

Lat = Latitud; Long = Longitud, Fuente que proporciona la información C.I Banacol=\* e IDEAM=+.

**Cuadro 2. Estaciones de IDEAM ubicadas fuera de la zona de estudio.**  
**Table 2. IDEAM stations located outside the study area.**

Estación	Altitud (m)	Latitud (grados)	Longitud (grados)	Precipitación		Tmín y Tmáx		Tmín (%) datos perdidos	Tmáx (%) datos perdidos
				Periodo	(%) datos perdidos	Periodo	(%) datos perdidos		
Uniban	23	7.49	76.39	2004-2012	1.52	2006-2012	23.63	2.47	
Aeropuerto Cedros	19	7.49	76.43		0.55	2006-2012	0.07	0	
Turbo	37	8.05	76.42		13.64	2006-2012	42.88	43.08	
Prado Mar	38	7.59	76.38		2.08	-	-	-	

Tmín= temperatura mínima y Tmáx= temperatura máxima.

**Cuadro 3. Distancia entre estaciones (km).**  
**Table 3. Distance between stations (km).**

Estación (zona estudio)	Uniban	Aeropuerto	Prado Mar	Turbo	Pista Indira
Pista Indira	14.079	14.609	7.699	17.375	
Magdalena	17.626	18.36	5.988	13.749	3.773
Tapartó	14.533	154.783	6.617	16.843	1.085
Salpicón	16.917	18.898	2.988	15.041	4.896
Tarena	11.628	12.502	8.804	19.809	2.449
Doña Ángela	8.954	8.898	12.237	23.048	5.793
San Jacinto	10.12	12.756	8.263	21.441	4.863
Manzano	15.334	18.439	2.679	17.701	6.283

## Método

Se utilizaron los métodos: U.S National Weather Service (WS) según Chow *et al.* (1994), es un promedio con el inverso de la distancia al cuadrado como factor ponderador, el racional deductivo (RD) propuesto por Campos (1998) basado en el porcentaje de participación que tiene el dato perdido sobre la variable mensual y para la regresión lineal

percentage of missing data regarding to other stations and its information was provided by IDEAM (Institute of Hydrology, Meteorology and Environmental Studies of Colombia) and the loss of information from stations whose information was provided by Banacol International Marketing (C. I Banacol); having a pattern of data loss associated to non-working days from farms, where weather stations are located. Table 2 shows the characteristics of



(RL) y múltiple (RM) sugeridos en WMO (1983), donde la variable meteorológica dependiente es el valor del dato faltante en la estación (Pista Indira) y las independientes es el valor de la variable de las estaciones vecinas el mismo día (Turbo, Uniban, Aeropuerto) en el caso de la temperatura en la regresión múltiple, en la regresión lineal la estación más cercana (Aeropuerto), y para la variable precipitación en la regresión múltiple las estaciones vecinas (Uniban, Aeropuerto, Prado Mar) y en la regresión lineal la más cercana (Prado Mar).

### Estimación del error

Se empleó la raíz cuadrada del cuadrado medio del error (RCCME) (Rivas y Carmona, 2010; Teegavarapu y Chandramouli, 2005), el error medio absoluto (MEA) (Teegavarapu y Chandramouli, 2005 y Kashani y Dinpashoh, 2012), el error relativo (RE), coeficiente de determinación ( $R^2$ ) y Índice de concordancia de Willmott ( $d$ ) (Rivas y Carmona, 2010 y Willmott, 1981). El modelo perfecto es cuando  $R^2=1$  y  $MAE=RCCME=RE=0$  y el mejor modelo debe tender a los límites anteriores, siendo excelente cuando  $d \geq 0.95$ ,  $RE \leq 0.20$  y  $R^2 > 0.8$  (Cai *et al.*, 2007 y Pereira, 2004).

## Resultados y discusión

Para la variable precipitación se obtuvieron valores para la RCCME de: 16.57, 19.16, 14.08, 15.31 mm para los métodos WS, RD, RM y RL respectivamente, con errores similares a los encontrados por Kashani y Dinpashoh (2012) y superiores a los reportados por Teegavarapu y Chandramouli, (2005) y Di Piazza (2011), el método RD presentó el valor más alto, la RCCME para WS, RM y RL es similar, ya que no tienen una diferencia mayor a 2.48 mm. Los valores del MAE fueron de: 8, 9.8, 8.5 y 9.2 mm para WS, RD, RM y RL respectivamente, con el menor valor para WS.

Los valores de MAE son menores a los reportados en Kashani y Dinpashoh (2012) y superiores a los encontrados por Teegavarapu y Chandramouli (2005), Di Piazza (2011) y Xia *et al.* (1999). El estadístico RE presentó valores de 2.4, 2.8, 2.0 y 2.2 para los métodos WS, RD, RM y RL respectivamente, no están por debajo de 0.2, principalmente porque la variable precipitación es una variable discreta con fuertes frecuencias de valores cercanos a ceros, por ende presenta un promedio bajo.

IDEAM stations located outside the study area (Figure 1) and compared to CI Banacol stations (Table 1) have a lower percentage of loss.

To define the method with the smallest error, it was removed 31% of the information to the station Pista Indira according to the loss pattern of the stations from C. I Banacol and to generate it the stations that are presented in Table 2 (located outside the study area) were used and were called neighbors. Once defined the method with the smallest error, the series CI Banacol stations were filled. Table 3 shows neighboring stations (IDEAM) with a distance less than 24 km, located within the same watershed; also there is no difference in height greater than 20 m (Table 2).

### Method

The methods used are: US National Weather Service (WS) according to Chow *et al.* (1994), is an average with the inverse distance squared as a weighting factor, deductive rational (RD) proposed by Campos (1998) based on the percentage of participation that has the data lost on the monthly variable and for linear (RL) and multiple regression (RM) suggested in WMO (1983), where the weather dependent variable is the value of the missing data on the station (Pista Indira) and the independent is the value of the variable of neighboring stations on the same day (Turbo, Uniban, Aeropuerto) in the case of temperature on multiple regression, in linear regression the nearest station (Airport), and precipitation variable in multiple regression the neighboring stations (Uniban, Aeropuerto, Prado Mar) and in linear regression the nearest (Prado Mar).

### Standard error

The square root of the mean squared error (RCCME) (Rivas and Carmona, 2010; Teegavarapu and Chandramouli, 2005) was used; the mean absolute error (MEA) (Teegavarapu and Chandramouli, 2005 and Kashani and Dinpashoh, 2012); relative error (RE), coefficient of determination ( $R^2$ ) and Willmott concordance index ( $d$ ) (Rivas and Carmona, 2010 and Willmott, 1981). The perfect model is when  $R^2 = 1$  and  $MAE=RCCME=RE=0$  and the best model must tend to the above limits, being excellent when  $d \geq 0.95$ ,  $RE \leq 0.20$  and  $R^2 > 0.8$  (Cai *et al.*, 2007 and Pereira, 2004).

El método RD reportó el valor más alto de RE mientras que WS, RM y RL son similares no tienen una diferencia mayor a 0.4. El índice (d) tuvo valores de 0.5, 0.4, 0.3 y 0.1 para los métodos WS, RD, RM y RL respectivamente, que no están por encima de 0.95 como lo recomienda (Cai *et al.*, 2007; Pereira, 2004); sin embargo, se distingue el método WS al presentar el mayor valor del índice (d), por lo anterior y por tener los estadísticos RCCME, MAE y RE similares a RL y RM el método WS es que mejor estima el dato observado.

Adicional You *et al.* (2008) proponen que el método WS proporciona mejores estimaciones cuando la elevación de la estación de interés está cerca de la elevación de las estaciones utilizadas para rellenar el dato faltante, situación que se presenta ya que no hay diferencias mayores a 22 m entre ellas. Así mismo los siguientes autores: Teegavarapu y Chandramouli (2005); Wade, 1987; Eischeid *et al.* (1995); Eischeid *et al.* (2000) y Hubbard *et al.* (2005) reportan mejores resultados para WS.

El índice (d) presentó valores de 0.8, 0.4, 0.8 y 0.8 para WS, RD, RM y RL respectivamente aunque no se encuentren valores mayores a 0.95, los métodos WS, RM, RL tienen las estimaciones con el menor grado de error con los valores más altos de (d), por lo tanto se podría utilizar cualquiera de los tres métodos ya que el índice (d) es similar entre ellos y los demás estadísticos también. Al comparar los métodos RM y RL en todas las variables el método RM tuvo mejores estadísticos que RL situación que también fue objeto de algunos estudios que compararon el rendimiento de estos métodos y en la mayoría de los casos la regresión basada en dos o más estaciones vecinas resultó ser la más eficiente (Eischeid *et al.*, 1995; Xia *et al.*, 1999; You *et al.*, 2008 y Kashani y Dinpashoh, 2012).

En el Cuadro 4 se presentan los valores del coeficiente de determinación que variaron desde 0.14-0.56. La RM tiene un coeficiente de determinación para todas las variables mayor que el método RL situación también encontrada por Alfaro *et al.* (2000) quienes reportaron que el método RM presenta mejoras significativas con respecto aquellos métodos basados en la información de una sola estación vecina (RL). Sin embargo, Campos (2005) recomienda no utilizar los métodos de regresión si los coeficientes de determinación no tienen valores mayores a 0.8, debido a que en este estudio, son bajos se recomienda mejor utilizar WS que presentó estadísticos similares. En Cuadro 4 (b, c y d) representan el coeficiente en la RM de las variables independiente: (Turbo, Uniban y Aeropuerto) para

## Results and discussion

For the variable precipitation the RCCME values were: 16.57, 19.16, 14.08, 15.31 mm with WS, RD, RM and RL methods respectively, with similar errors to those found by Kashani and Dinpashoh (2012) and higher to those reported by Teegavarapu and Chandramouli, (2005) and Di Piazza (2011), the RD method showed the highest values, RCCME for WS, RM and RL are similar, since they don't have a difference higher than 2.48 mm. MAE values were: 8, 9.8, 8.5 and 9.2 mm for WS, RD, RM and RL respectively, with the lowest value for WS.

MAE values are lower than those reported in Kashani and Dinpashoh (2012) and higher than those found by Teegavarapu and Chandramouli (2005), Di Piazza (2011) and Xia *et al.* (1999). RE had of 2.4, 2.8, 2 and 2.2 for WS, RD, RM and RL respectively, these are not below 0.2, mainly because precipitation is a discrete variable with strong frequencies with near zero values, therefore shows a low average.

RD reported the highest value of RE while WS, RM and RL are similar, not having a difference higher than 0.4. Index (d) had values of 0.5, 0.4, 0.3 and 0.1 for WS, RD, RM and RL respectively, which are not above 0.95 as recommended (Cai *et al.*, 2007; Pereira, 2004); however, WS highlights by showing the highest index value (d), for the above and for having the RCCME, MAE and RE similar to RL and RM, WS is the best to estimate the observed data.

In addition You *et al.* (2008) propose that the WS method provides better estimates when the elevation of the station of interest is close to the elevation of the stations used to fill in the missing data, a situation that occurs, since there are no differences higher than 22 m between stations. Likewise the following authors: Teegavarapu and Chandramouli (2005); Wade, 1987; Eischeid *et al.* (1995); Eischeid *et al.* (2000) and Hubbard *et al.* (2005) report better results for WS.

The index (d) had values of 0.8, 0.4, 0.8 and 0.8 for WS, RD, RM and RL respectively although there are no values higher than 0.95; the WS, RM, RL methods have estimates with the lowest degree of error with the higher values of (d), so it could be used any of three methods, since index (d) is similar between them and also the other statistics. Comparing RM and RL in

temperatura y (Uniban, Aeropuerto y Prado Mar) en la variable precipitación, b representa la pendiente en RL con la variable independiente aportada por Aeropuerto para la temperatura y en la precipitación para Prado Mar.

all the variables, RM had better statistical than RL; situation that was subject of some studies comparing the performance of these methods and in most cases the regression based on two or more neighboring stations

**Cuadro 4. Coeficiente de determinación ( $R^2$ ) y coeficientes de regresión en RL y RM.**  
**Table 4. Coefficient of determination ( $R^2$ ) and regression coefficients of RL and RM.**

Variable	Método	$R^2$	n	Coeficientes de regresión			
				a	b	c	d
Temperatura mínima	RM	0.3224	167	10.78	0.09	0.2	0.27
	RL	0.2133	167	16.07	0.33		
Temperatura máxima	RM	0.5597	227	-2.98	0.18	0.25	0.67
	RL	0.3899	227	7.4	0.78		
Precipitación	RM	0.1391	436	3.35	0.24	0.12	0.002
	RL	0.0123	433	5.73	0.11		

b, c, d fueron estadísticamente significativos con ( $p < 0.05$ ), a es el intercepto de la recta en RM y RL, n= número de datos empleados.

Para la temperatura mínima en la Figura 2 se observa que valor estimado y observado por el método WS presenta un valor de la pendiente más cercano a la unidad y el intercepto a cero por lo que se aproxima a la recta 1:1, de esta se manera la estimación se acerca más al valor observado, mientras que para la temperatura máxima obtuvo el mayor acercamiento a la recta 1:1 por los método RM y WS ya que el valor de la pendiente se aproxima a la unidad y el intercepto a cero (Figura 3). En la variable temperatura al relacionar los valores estimados con los observados a través de una regresión se encontró bajos coeficientes de determinación para el método RD por lo tanto no presenta una relación significativa, igualmente para la precipitación en todos los métodos se presentan valores bajos de coeficientes de determinación por ende no muestra una relación significativa.

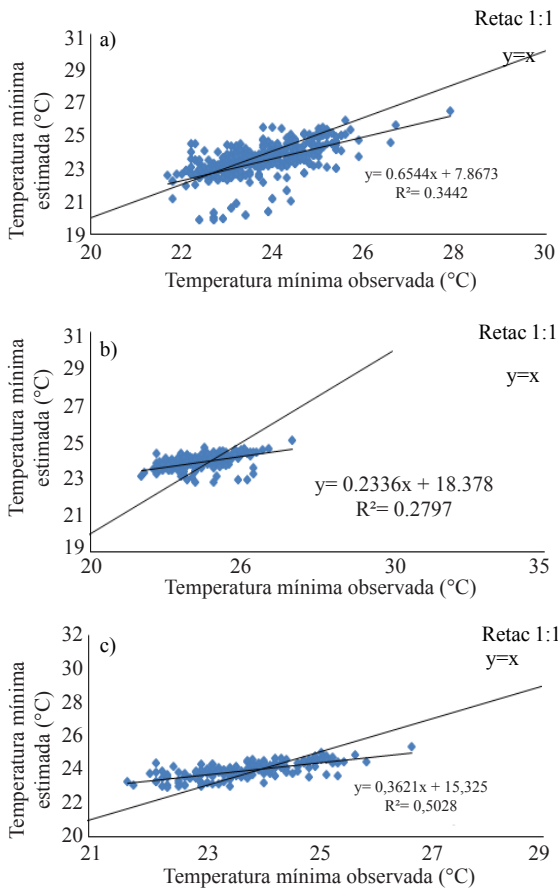
En la regresión entre los valores observado y estimado de los método RM y WS presentan un coeficiente de determinación mayor que el métodos RL y DR en las variables evaluadas, aunque los valores de  $R^2$  presentan valores inferiores a los reportados por Caí *et al.* (2007) como para considerar un buen ajuste del modelo, el coeficiente de determinación permite distinguir que existe dependencia entre los parámetros de ponderación o pesos y la distancia de las estaciones (WS), no siendo este un parámetro perfecto que por sí solo pueda predecir el comportamiento de las variables de estudio (Teegavarapu y Chandramouli, 2005), pero si mucho mejor que utilizar un modelo que solo dependa la información meteorológicas disponible (RM).

turned to be the most efficient (Eischeid *et al.*, 1995; Xia *et al.*, 1999; You *et al.*, 2008 and Kashani and Dinpashoh, 2012).

Table 4 shows the values of the coefficient of determination ranging from .14-.56. RM has a greater coefficient of determination for all variables than RL, situation also found by Alfaro *et al.* (2000) who reported that RM shows significant improvements over methods based on information from one neighboring station (RL). However, Campos (2005) recommends not to use regression methods if coefficients of determination do not have values higher than 0.8, because in this study are low, it is recommended to use WS that showed similar statistics. In Table 4 (b, c and d) represent the coefficient in RM of independent variables: (Turbo, Uniban and Aeropuerto) for temperature and (Uniban, Aeropuerto and Prado Mar) in the variable precipitation, b represents the slope in RL with an independent variable provided by Aeropuerto for temperature and in precipitation for Prado Mar.

For minimum temperature, Figure 2 shows that the estimated and observed value by WS has a value of the nearest slope to unity and the intercept to zero so it approaches the line 1: 1, in this way the estimate is closer to the observed value, whereas maximum temperature had the greater approach to line 1: 1 by RM and WS, since the value of the slope is close to unity and the intercept to zero (Figure 3). In the variable temperature when relating the estimate values with the observed ones through a

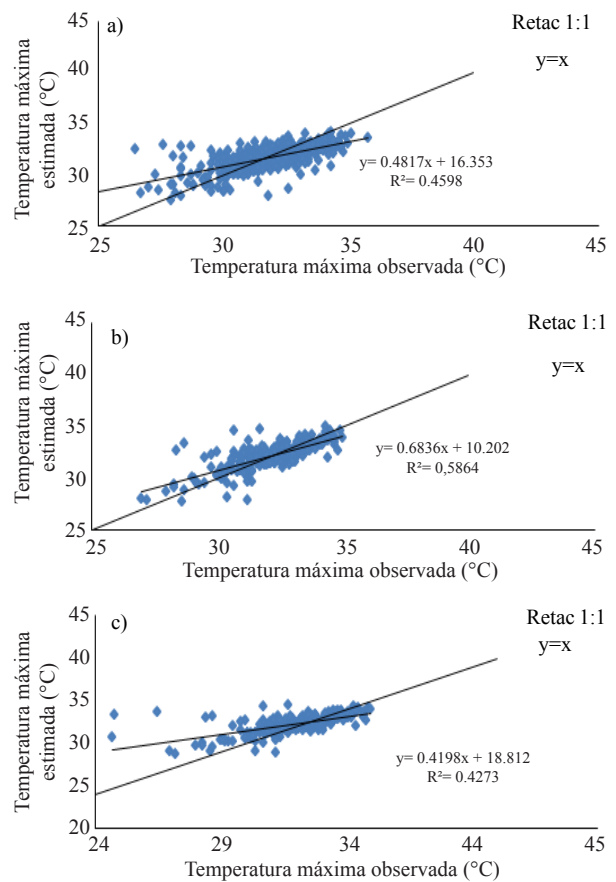




**Figura 2. a) Relación entre el valor estimado y observado de la temperatura mínima con WS; b) RM; y c) RL.**  
**Figure 2. a) Relationship between the estimated and observed value of the minimum temperature with WS; b) RM; and c) RL.**

Las Figuras 4, 5 y 6 se comparan las tendencias más robustas entre los valores estimados para las estaciones vecinas y las estaciones de C. I Banacol entre ellas en la (Figura 4a), se presenta la temperatura máxima generada con el WS para las estaciones dentro la zona de estudio, con un comportamiento similar entre sí, con excepción de la estación Doña Ángela, quien es la más cercana a la estación Aeropuerto (Estación vecina) y que, para el año 2009 tuvo los valores más bajos de temperatura máxima lo que hace sesgar el comportamiento de la estación Doña Ángela del comportamiento obtenido por las otras estaciones para ese año 2009.

En la Figura 4 b y 5 b se observa que los valores generado por el método WS para temperatura máxima y mínima están por encima a los valores de las estaciones utilizadas como vecinas, situación que no pasa con la variable precipitación (Figura 6b) en donde las estimaciones se encuentra dentro del rango de oscilación de las estaciones

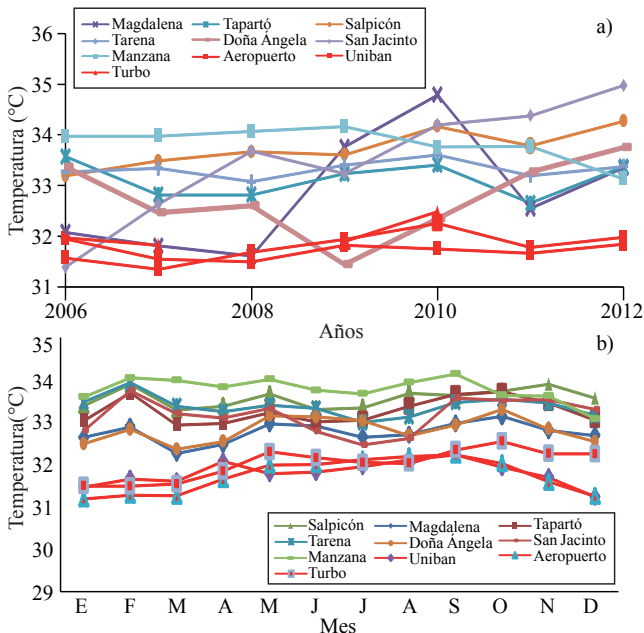


**Figura 3. a) Relación entre el valor estimado y observado de la temperatura máxima con WS; b) RM; y c) RL.**  
**Figure 3. a) Relationship between the estimated and observed value of maximum temperature with WS; b) RM; and c) RL.**

regression, were found low coefficients of determination for RD, therefore does not present a significant relationship, the same for precipitation in all methods, showing low values of coefficients of determination, therefore shows no significant relationship.

In the regression between observed and estimate values of RM and WS, show a coefficient determination higher than RL and DR in the variables evaluated, although  $R^2$  values had lower values than those reported by Cai *et al.* (2007) considering a good fit of the model, the coefficient of determination allows to distinguish that there is dependence between the weighting parameters and distance of the stations (WS), not being this a perfect parameter, that alone could predict the behavior of the study variables (Teegavarapu and Chandramouli, 2005), but much better than using a model that only depends on available weather information (RM).

vecinas, posiblemente porque la temperatura es una variable más localizadas y expuestas a la cercanía de la serranía Abibe (You *et al.*, 2008).



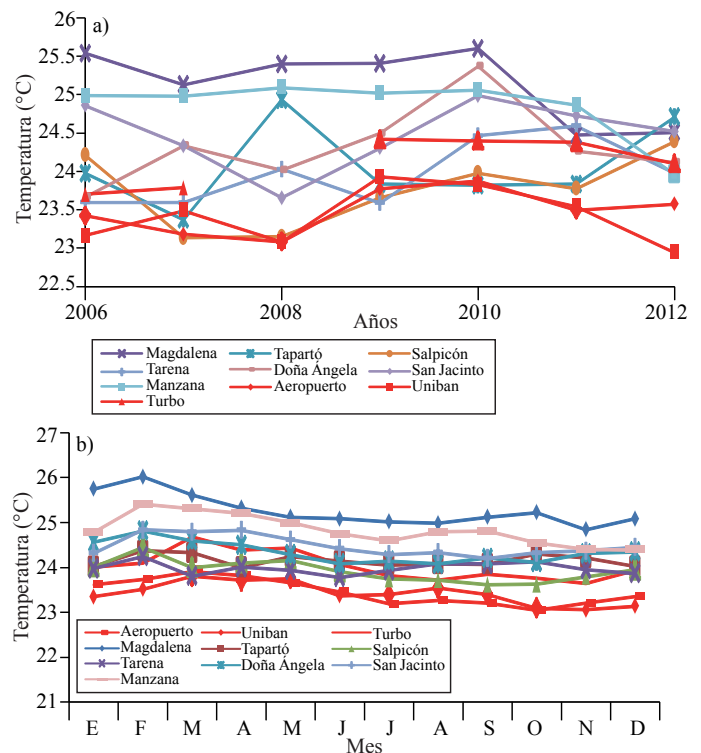
**Figura 4. a) Temperatura máxima media anual; y b) Temperatura máxima interanual generada con las estaciones de la zona de estudio.**  
**Figure 4. a) Maximum annual mean temperature; and b) Maximum annual temperature generated by the stations of the study area.**

### Conclusión

En este trabajo se utilizaron cuatro métodos para calcular datos faltantes, en series diarias de precipitación, temperatura mínima y máxima, para determinar su confiabilidad se evaluaron para la estación Pista Indira para el periodo 2006-2009, donde se contaba con registros continuos, para posteriormente aplicarlo a las 7 estaciones de la zona de estudio con valores faltantes. Las variable precipitación, temperatura máxima y temperatura mínima en los métodos RM, WS y RL presentaron estadísticos RE, RCCME y MAE similares, pero el valor mayor valor del índice (d) para la precipitación se tuvo con el método WS, la temperatura mínima con los métodos WS y RM fue similar y para la temperatura máxima RM, WS y RL presentaron un comportamiento semejante, por ende su utilización con llevaría a determinar errores similares, pero se utilizó

Figures 4, 5 and 6 compare the most robust trends between estimates values for neighboring stations and between the stations from C. I Banacol in (Figure 4a), the maximum temperature values generated with WS for the stations within the study area, with similar behavior in between, except the station Doña Angela, that is closest to the Aeropuerto (neighboring station) station and by 2009 had the lowest values for maximum temperature which makes bias the behavior of Doña Angela station obtained from the other stations for 2009.

Figure 4 b and 5 b shows that the values generated with WS for maximum and minimum temperatures are above the values from the stations used as neighboring, situation that does not happen with precipitation (Figure 6b) where estimates are within the oscillation range of neighboring stations, possibly because the temperature is a more localized variable and exposed to the proximity of the mountains Abibe (You *et al.*, 2008).



**Figura 5. a) Temperatura mínima media anual; y b) Temperatura mínima interanual generada con las estaciones de la zona de estudio.**  
**Figure 5. a) Minimum annual mean temperature; and b) minimum annual temperature generated by the stations of the study area.**

el método WS para el relleno de los datos faltante de 7 estaciones, ya que los métodos RL y RM tuvieron bajos coeficientes de determinación.

## Literatura citada

- Alfaro, R. y Pacheco, R. 2000. Aplicación de algunos métodos de relleno en series anuales de lluvia de diferentes regiones de Costa Rica. Instituto Meteorológico Nacional, Tópicos Meteorológicos Oceanográficos. 7(1):1-20.
- Aparicio, F. J. 2011. Fundamentos de hidrología de superficie. Limusa. 1ª. D. F., México. 304 p.
- American Society of Civil Engineers (ASCE). 1996. Hydrology handbook. 2ª. New York, EE.UU. 769 p.
- Bussieres, N. and Hogg, W. 1989. The objective analysis of daily rainfall by distance weighting schemes on a mesoscale grid. Atmosphere Ocean. 27:521-541.
- Campos, D. F. 1998. Procesos del ciclo hidrológico. Facultad de Ingeniería. Universidad Autónoma de San Luis Potosí (UASLP). 3ª. San Luis Potosí, México. 500 p.
- Campos, D. F. 2005. Agroclimatología cuantitativa de cultivos. Trillas. 2ª. D. F., México. 320 p.
- Chow, V. T.; Maidment, D. R. y Mays, L. W. 1996. Hidrología aplicada. McGraw Hill. 1ª. Bogotá, Colombia. 583 p.
- Degaetano, A. T.; Eggleston, K. L. and Knapp, W. W. 1995. A method to estimate missing maximum and minimum temperature observations. J. Appl. Meteorol. 34(2):371-380.
- Di Piazza, A.; Conti, F. L.; Noto, L. V.; Viola, F. and La Loggia, G. 2011. Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. Int. J. Appl. Earth Obser. Geoinf. 13(3):396-408.
- 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. J. Appl. Meteorol. 39(9):1580-1591.
- Eischeid, J. K.; Bruse, B. C.; Karl, T. R. and Díaz, H. F. 1995. The quality control of long-term climatological data using objective data analysis. J. Appl. Meteorol. 34(12):278-279.
- Guevara, J. M. 2003. Métodos de estimación y ajuste de datos climáticos. Imprenta Universidad Central de Venezuela, Ministerio del Ambiente y de los Recursos Naturales. Dirección de Hidrología y Meteorología. 2ª. Mérida, Venezuela. 74 p.
- Hubbard, K. G. and You, J. 2005. Sensitivity analysis of quality assurance using spatial regression approach. A case study of the maximum minimum air temperature. J. Atmospheric Oceanic Technol. 22(10):1520-1530.
- Hubbard, K. G. 2001. Multiple station quality control procedures. Automated weather stations for applications in agriculture and water resources management, AGM-3 WMO./TD No. 1074. High Plains Regional Climate Center. Lincoln. EE.UU. 248 p.
- Hubbard, K. G. 1994. Spatial variability of daily weather variables in the high plains of the USA. Agric. Forest Meteorol. 68(1):29-41.
- Huth, R. and Nemesova, I. 1995. Estimation of missing daily temperature: can a weather categorization improve its accuracy. J. Appl. Meteorol. 34:1901-1916.

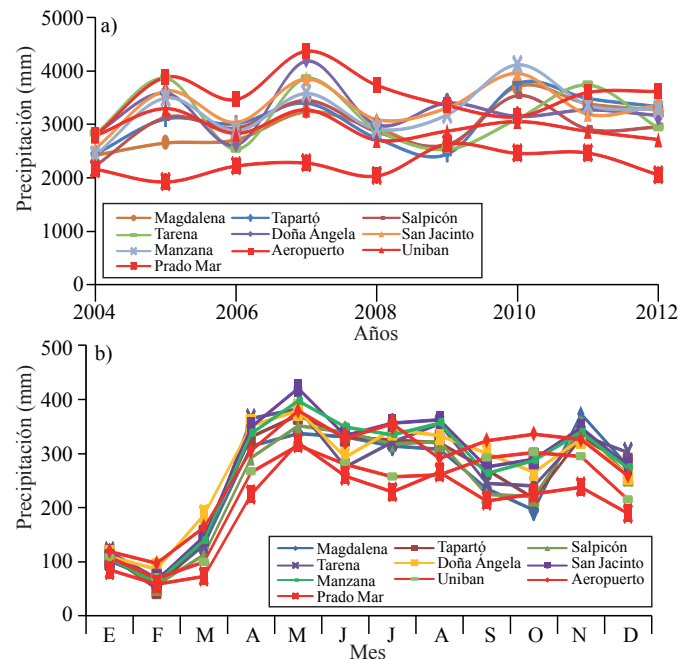


Figura 6. a) Precipitación anual; y b) Precipitación media interanual generada con las estaciones de la zona de estudio.

Figure 6. a) Annual precipitation; and b) average annual rainfall generated by the stations in the study area.

## Conclusion

In this paper four methods were used to estimate missing data, in daily series of precipitation, minimum and maximum temperature to determine their reliability; were evaluated for Pista Indira station during 2006-2009, having continuous records, used for 7 stations in the study area with missing values. The precipitation, maximum temperature and minimum temperature in RM, WS and RL showed similar statistical to RCCME and MAE, but the highest value of index (d) for the precipitation was with WS, the minimum temperature with WS and RM was similar and for maximum temperature with RM, WS and RL showed similar behavior, hence their use would lead to similar errors, but the WS method was used to fill in the missing data from 7 stations, since RL and RM had low coefficients of determination.

*End of the English version*



- Jiménez, C. A.; Vargas, T. V.; Salinas, C. W.; Aguirre, B. M y Cabrera, D. R. 2004. Aptitud agroecológica para el cultivo de la caña de azúcar en el Sur de Tamaulipas. Investigaciones Geográficas. Boletín del Instituto de Geografía. Universidad Nacional Autónoma Metropolitana (UAM). 53(53):58-74.

- Kashani, M. H. and Dinpashoh, Y. 2012. Evaluation of efficiency of different estimation methods for missing climatological data. *Stochastic Environmental Research and Risk Assessment*. 26(1):59-71.
- Massetti, L. 2013. Analysis and estimation of the effects of missing values on the calculation of monthly temperature indices. *Theor. Appl. Climatol.* 117(3-4):511-519.
- McCuen, R. H. 1998. *Hydrologic analysis and design*. Prentice-Hall. 3ª. NJ, EE.UU. 814 p.
- Presti, R. L.; Barca, E. and Passarella, G. 2010. A methodology for treating missing data applied to daily rainfall data in the Candelaro River Basin (Italy). *Environ. Monitoring Assessment*. 160(1-4):1-22.
- Puertas, O. O.; Carvajal, E. Y. y Quintero, M. 2011. Estudio de tendencias de la precipitación mensual en la cuenca alta-media del río Cauca, Colombia. *Dyna*. 78(169):112-120.
- Rivas, R. y 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.
- Saborowski, J. y Stock, R. 1994. Regionalization of precipitation data in the Harz mountains. *Allgemeine Forst und Jagdzeitung*. 165:117-122.
- Salazar, C. A. 2012. Los sistemas de riego en cultivo del banano: aplicado a las zonas bananeras de Colombia. Universidad Nacional de Colombia. 1ª. Medellín, Colombia. 591 p.
- Sokol, Z. and Stekl, J. 1994. 3-D mesoscale objective analysis of selected elements from SYNOP and SYRED reports. *Meteorologische Zeitschrift*. 3(4):242-246.
- Teegavarapu, R. S. and Chandramouli, V. 2005. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *J. Hydrol.* 312(1-4):191-206.
- Wade, C. G. 1987. A quality control program for surface mesometeorological data. *J. Atmospheric Oceanic Technol.* 4(3):435-453.
- Wagner, D. P.; Fiener, P.; Wilken, F.; Kumar, S. and Schneider, K. 2012. Comparison and evaluation of spatial interpolation schemes for daily rainfall in data scarce regions. *J. Hydrol.* 464-465:388-400.
- Willmott, C. J. 1981. On the validation of models. *Physical geography*. 2(2):184-194.
- WMO. 1981. Guide to agricultural meteorological practices. WMO-No. 134. Geneva, Switzerland. 799 p.
- WMO. 1983. Guide to climatological practices. WMO-No100. Ginebra, Suiza. 180 p.
- Xia, Y.; Fabian, P.; Stohl, A.; Winterhalter, M. 1999. Forest climatology: estimation of missing values for Bavaria, Germany. *Agric. Forest Meteorol.* 96(1):131-144.
- You, J. S.; Hubbard, K. G. and Goddard, S. 2008. Comparison of methods for spatially estimating station temperatures in a quality control system. *Int. J. Climatol.* 28(6):777-787.
- Young, K. C. 1992. A three-way model for interpolating monthly precipitation values. *Monthly Weather Review*. 120(11):2561-2569.
- Yozgatligil, C.; Aslan, S.; Iyigun, C. and Batmaz, I. 2013. Comparison of missing value imputation methods in time series: the case of Turkish meteorological data. *Theor. Appl. Climatol.* 112(1-2):143-167.