DOI: 10.24850/j-tyca-14-01-10

Notes

Data quality of automatic meteorological stations located in three states of Mexico

Calidad de datos de estaciones meteorológicas automáticas ubicadas en tres estados de México

Rocío Cervantes-Osornio[1], ORCID: https://orcid.org/0000-0003-2597-4517

Ramón Arteaga-Ramírez[2], ORCID: https://orcid.org/0000-0001-9459-3588

Mario Alberto Vázquez-Peña[3], ORCID: https://orcid.org/0000-0003-2084-7420

Waldo Ojeda-Bustamante[4], ORCID: https://orcid.org/0000-0001-7183-9637

[1]Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Cevamex, Texcoco, State of Mexico, Mexico, rcervanteso@hotmail.com

[2]Universidad Autónoma Chapingo, Departamento de Irrigación, State of Mexico, Mexico, arteagarr@gmail.com

[3]Universidad Autónoma Chapingo, Departamento de Irrigación, State of Mexico, Mexico, mvazquezp@gmail.com

[4]Instituto Mexicano de Tecnología del Agua, Jiutepec, Morelos, Mexico, w.ojeda@riego.mx

Corresponding author: Rocío Cervantes-Osornio, rcervanteso@hotmail.com

Abstract

The objective of the present work was to apply some procedures to evaluate the quality of data coming from different automatic meteorological stations (EMA) located in three states of the Mexican Republic. Daily meteorological data were obtained from automatic stations: one in the State of Mexico, five in Guanajuato, and one in Sinaloa. The procedures that were used were: for the maximum, average, and minimum temperature and global radiation, the flexible range test or meteorological event was applied; for the temperature and relative humidity variables, the internal consistency test was used; to the variable temperature, the thermal jump and the zeros test were applied; to the relative humidity, the internal consistency, and the persistence test. For the station located in Sinaloa, for the variable daily global radiation, on the flexible range test, it presented 23.25 % of erroneous data. The results of the procedures applied to historical data of the EMAs support determining the reliability of the meteorological information of these for use.

Keywords: Quality, meteorological data, relative humidity, temperature, solar radiation.

Resumen

El objetivo del presente trabajo fue aplicar algunos procedimientos para evaluar la calidad de datos provenientes de diferentes estaciones meteorológicas automáticas (EMA) ubicadas en tres estados de la república mexicana. Se obtuvieron datos meteorológicos diarios de estaciones automáticas ubicadas en los siguientes puntos: una en el estado de México, cinco en Guanajuato, y una en Sinaloa. Los procedimientos utilizados para la temperatura máxima, media y mínima, y radiación global fueron la prueba de rango flexible o efeméride meteorológica; para las variables temperatura y humedad relativa se empleó la prueba de consistencia interna; a la variable temperatura se le aplicó la prueba de salto térmico y de ceros; a la humedad relativa, la prueba de consistencia interna y la de persistencia. Para la estación ubicada en Sinaloa, la variable radiación global diaria, en la prueba de rango flexible, presentó 23.25 % de datos erróneos. Los resultados de los procedimientos aplicados a datos históricos de las EMA apoyan en determinar la confiabilidad de la información meteorológica de éstas para su uso.

Palabras clave: calidad, datos meteorológicos, humedad relativa, temperatura, radiación solar.

Received: 11/08/2020

Accepted: 18/09/2021

Introduction

Meteorological data are considered to be high-quality when they respond to certain norms, both explicit and implicit (WMO, 1996). These depend on the different applications the information will be used for (Estévez, 2008), some of which may be to distinguish temperature and rainfall anomalies under climate change scenarios (Campos-Aranda, 2020) or to discover patterns in weather data, particularly in temperature (Bulege-Gutiérrez & Custodio, 2020). Different methods are applied to meteorological data to ensure their quality: regular and adequate maintenance of the stations and sensors on the field, calibration of the latter, and data validation using statistical procedures (Estévez, Gavilán, & Girálvez, 2011). Obrien and Keefer (1985) proposed three rules based on information technology, which Meek and Hatfiel (1994) applied to validate meteorological data. These include calculating high, low, fixed, or dynamic limits for each variable, using a fixed or dynamic rate of the change limits for each variable, and identifying continuous changes not observed in a time limit. Estévez, García-Marín, Morábito, and Cavagnaro (2016) conducted a study in the province of Mendoza, Argentina, and they applied different methods (detection and marking of empty data, range or fixed and dynamic test, jump test, and internal consistency and persistence) to validate data from the variables: daily solar radiation; maximum, minimum and mean daily temperature; maximum, minimum and mean daily relative humidity; maximum, minimum and mean wind speed.

Shafer, Fiebrich, and Arndt (2000) applied for temperature (at 1.5 m and 9 m) and maximum relative humidity, the test of range, step or difference, persistence, spatial and comparison of similar instruments, on data registered every 5 minutes, and others every 15 and 30 minutes, in 115 stations in Oklahoma. In 1998, staff personnel found more than 50 "bursts of heat", relatively common events that affected one or two stations and appeared as abnormal conditions of heat and dryness during the nighttime. These were also registered as data with problems, which were automatically stored. However, a human judgment could distinguish between these indications, and the records of these events were stored for future studies.

The objective of the present study was to determine the quality of meteorological measurements with different procedures for automatic stations in three locations (Guanajuato, Sinaloa, and the State of Mexico).

Materials and methods

The used meteorological data consisted of global radiation (RG *o* R_s), maximum (T_{max}) mean (T_{med}) and minimum temperatures (T_{min}), maximum (HR_{max}), mean (HR_{med}), and minimum relative humidity (HR_{min}). Table 1 shows the latitude, longitude, height, and study period of the different stations used.

Table 1. Geographic coordinates and observation period of the study stations.

State	Station	Latitude	Longitude	Altitude, masl	Study period
State of México	Chapingo	19° 29′ 05″	98° 53′ 11″	2250	23/1/2004-31/12/2013
Guanajuato	Agroeduca	21° 00′ 39.2″	101° 39′ 4.9″	1777	1/1/2001-31/12/2006
Guanajuato	El Fuerte	20° 37′ 26.4″	101° 12′ 28.3″	1716	1/1/2001-31/12/2006
Guanajuato	El Cardonal	21° 11′ 39.7″	100° 46′ 39″	1912	1/1/2001-31/12/2006
Guanajuato	El Galomo	21° 04′ 2.7″	100° 27′ 14.3″	2048	1/1/2001-31/12/2006
Guanajuato	El Jardín	21° 15′ 52.5″	100° 35′ 30.1″	1968	1/1/2001-31/12/2006
Sinaloa	Santa Rosa 1	25° 45′ 03″	108° 57′ 21″	40	23/4/1997-15/5/2001

Before applying the validation of meteorological data, it was necessary to perform a prior analysis of the database, which meant detecting empty (blank) spaces. To evaluate the daily maximum (T_{max}), mean (T_{med}), minimum temperatures (T_{min}), global radiation (RG), and maximum (HR_{max}), mean (HR_{med}), and minimum relative humidity (HR_{min}), the rigid range test was used. The flexible range or meteorological ephemeride test was used for the maximum mean temperature and global radiation. No extreme data were found for relative humidity in the climatological normal, which made it impossible to carry out the flexible range test. The variables of temperature (T_{max}, T_{med} y T_{min}) and relative humidity (HR_{max}, HR_{med}, HR_{min}) underwent the internal

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

consistency test; temperature, the thermal jump test ($T_{max}(d) - T_{min}(d)$) and zero tests; the persistence test, relative humidity (HRmax, HRmed y HRmin), and global radiation. To apply the flexible range test or ephemerides, extreme values are needed, and since the automatic stations used in the study contain data aged 10 years or less, nearby stations were selected from the climatological normals (NC) published by the National Meteorological Service with similar characteristics in terms of altitude, latitude, and longitude, from which the necessary values were taken for this test. Table 2 shows the stations selected from the NC.

Table 2. Selected stations of the Climatological Normals (NC).

State	Station	NC Station	Latitude	Length	Altitude, masl
Guanajuato	Agroeduca	San Francisco Durán	21° 00′ 19″	101° 37′ 16″	1 790
Guanajuato	El Fuerte	Salamanca (DGE)	20° 33′50″	101° 11′ 54″	1 723
Guanajuato	El Cardonal	Peñuelitas	21° 06′ 16″	100° 52′ 33″	1 915
Guanajuato	El Galomo	San José Iturbide (DGE)	20° 59′ 52″	100° 23′ 07″	2 099
Guanajuato	El Jardín	Lourdes	21° 17′ 24″	100° 42′ 07″	1 995
Sinaloa	Santa Rosa 1	El Zopilote	25° 43′ 50″	108° 20′ 45″	37

Blanks test. The empty spaces or data remaining for the different variables were detected in the databases of the stations under study.

Rigid range and flexible range tests (ephemerides or events meteorological) (Estévez *et al*., 2011). This used the physical limits proposed by Feng, Hu, and Qian (2004) and UNE 500540 (2004) due to the lack of access to the data of the specifications on the respective sensors in the three study locations. In addition, this test verifies meteorological ephemerides (dynamic limit or flexible range), in which each value is compared with the extreme value historically registered in a particular place.

Flexible range test for solar radiation. To evaluate the solar radiation data for the three locations, daily data of the stations were used (R_s ó RG) represented graphically, with relation to the radiation expected under clear sky or sunny conditions (R_{so}), and the latter is calculated (Allen, 2008) for any given day with:

$$R_{so} = K_t \cdot R_a \tag{1}$$

Where K_t is the "clearness" index, and R_a is the extraterrestrial radiation, the same units of the R_s (MJ m^{-2} d^{-1} or MJ m^{-2} h^{-1} ó W m^{-2}). According to Allen (2008) and the ASCE-EWRI methodology (ASCE-EWRI, 2005), $K_t = K_B + K_D$, and K_B is the index of clearness of the direct radiation beam (dimensionless), K_D is the index of transmissivity for the diffused, dimensionless radiation, so:

$$R_{so} = (K_B + K_D) \cdot R_a \tag{2}$$

The estimation of K_B proposed by Majumdar, Mathur, and Kaushik (1972), Allen (1996), and Allen, Smith, Pereira, and Perrier (1998) is applied by ASCE-EWRI (2005) with the improvement of the coefficients developed in the evaluation of the ASCE Task Committee, for R_s data from nearly fifty meteorological sites in the United States, and for K_B we have:

$$K_B = 0.98exp\left[\frac{-0.00146P}{K_t \sin\beta} - 0.075\left(\frac{W}{\sin\beta}\right)^{0.4}\right] \tag{3}$$

Where:

K_t is the coefficient of turbidity (dimensionless), $0 < K_t \leq 1.0$, where $K_t = 1.0$ for clear air and $K_t \leq 0.5$ for turbid, dusty, or polluted air.

P is atmospheric pressure, [kPa], estimated with the site's elevation according to Allen *et al*. (1998).

β is the angle of the sun over the horizon (rad).

W is the precipitable water in the atmosphere (mm).

The value for K_t varies based on the time of the year and the cleanliness of the atmosphere due to precipitation. In general, to predict net radiation (R_n) and R_{so} envelopes, we recommend $K_t = 1.0$. The $\sin\beta$ in Equation (3) must be limited to values ≥ 0.01 for computer stability.

Precipitable water is estimated as:

$$W = 0.14 e_a P + 2.1 \tag{4}$$

Where e_a is the current water vapor pressure (at 2 m) (kPa).

The transmissivity index for diffused radiation is estimated from K_B (according to Allen, 1996):

$$K_D = 0.35 - 0.36 K_B \text{ to } K_B \geq 0.15$$

$$K_D = 0.18 + 0.82 K_B \text{ to } K_B < 0.15$$

For daily periods (24 horas), the mean value of β, is calculated, according to Allen (1996), as:

$$(\sin\beta)_{24} = \sin\left[0.85 + 0.3\varphi \sin\left(\frac{2\pi}{365}J - 1.39\right) - 0.42\varphi^2\right] \tag{5}$$

Where $(\sin\beta)_{24}$ is the daily average during the period of sunlight [rad]; φ is the latitude [rad], and J is the Julian day [dimensionless]

The variable $(\sin\beta)_{24}$ is used instead of $\sin\beta$ in Equation (3) and represents the average pondered of the sun's angle during daylight hours. The value for $(\sin\beta)_{24}$ must be ≥ 0.

Internal consistency test (Reek, Doty, & Owen, 1992). This ensures the meteorological coherence of the variable in question and also verifies that the maximum temperature (instant value) is above the average temperature (mean value), in turn, above the minimum value (instant value) (Equation (8)), Table 3). The same occurs for the maximum, minimum, and mean relative humidity variables (Equation (14)). This attempts to verify the coherence of the three variables stored daily (Estévez *et al*., 2011).

Table 3. Equations and tests applied in the different stations to the meteorological variables in the study locations.

Variable	Test of	Equation	Equation number
Maximum, minimum, and mean temperature	Rigid range	$-35 < T_{max}, T_{min}, T_{med} < 55$	(6)
	Flexible range (meteorological ephemeride)	$T_B < T_{max}, T_{min}, T_{med} < T_A$	(7)
	Internal consistency	$T_{max} > T_{med} > T_{min}$	(8)
		$T_{max}(d) > T_{min}(d-1)$	(9)
		$T_{min}(d) \leq T_{max}(d-1)$	(10)

	Daily termal jump	$dT(d) < 23.8\,°C$	(11)
	Zeros	$T_{min}(d) = T_{min}(d-1) = T_{min}(d-2) = 0$	(12)
Maximum, minimum, and mean relative humidity	Rigid range	$0.8 < HR_{max}, HR_{min}, HR_{med} < 100$	(13)
	Internal consistency	$HR_{max} > HR_{med} > HR_{min}$	(14)
	Persistence	$HR_{max}(d) \neq HR_{max}(d-1) \neq HR_{max}(d-2)$	(15)
		$HR_{min}(d) \neq HR_{min}(d-1) \neq HR_{min}(d-2)$	(16)
		$HR_{med}(d) \neq HR_{med}(d-1) \neq HR_{med}(d-2)$	(17)
Solar radiation	Persistence	$RG(d) \neq RG(d-1) \neq RG(d-2)$	(18)
	Flexible range or meteorological ephemeride	$RG < 1.1 \cdot R_{so}$	(19)

T_{max} is the maximum temperature (°C); T_{med} is the mean temperature (°C); T_{min} is the minimum temperature (°C); HR_{max} is the maximum relative humidity (%); HR_{med} is the mean relative humidity (%); HR_{min} is the minimum relative humidity (%); (d) is the day evaluated for the variable in question (T_{max}, T_{med}, T_{min}, HR_{max}, HR_{med}, HR_{min}); $(d-1)$ is the previous day evaluated for the variable in question (dimensionless); $(d-2)$ is the evaluation two days ago for the variable in question (dimensionless); $dT(d) = T_{max}(d) - T_{min}(d)$ is the daily thermal jump (°C); T_A is the value of the highest extreme temperature recorded in the SMN for the station in question (°C); T_B is the lowest minimum temperature value recorded in the SMN for the station in question (°C); RG is the global radiation (MJ $m^{-2}day^{-1}$); R_{so} is the daily global radiation under clear sky conditions (MJ $m^{-2}day^{-1}$).

Source: elaborated by Estévez et al. (2011); AENOR (2004); Reek et al. (1992).

Equations (9) and (10) (Table 3) verify that the maximum temperature is higher than the minimum of the previous day and that the minimum temperature is lower or equal to the maximum temperature of the previous day.

Daily thermal jump test. Verifies that the difference between T_{max} and T_{min} for any given day does not surpass a particular threshold (Araya-López, 2011; Reek *et al*., 1992).

Persistence test. This verifies if there are more than two records with the same value for maximum, mean, or minimum relative humidity or global radiation for the daily data (Meek & Hatfiel, 1994).

Zero test. To very continuous records of minimum temperatures equal to zero (Schroeder, Burgett, Haynie, & Sonmez, 2005).

Results and discussion

Blanks test. For all three locations, the number of blank records (empty fields) was detected in every case and grouped by year and variable. For the five stations in Guanajuato and the Santa Rosa I AC station in Sinaloa, no empty fields appeared (0 %) for all variables and the years studied, indicating that these stations have no missing information.

For the Chapingo station in the state of Mexico, 160 empty spaces were registered for each set of 3631 data on this station. No missing data were presented in five years, and the highest percentage of empty fields was in 2012. It should be noted that the number for each meteorological variable (RG, T_{max}, T_{med}, T_{min}, HR_{max}, HR_{med}, and HR_{min}) indicates problems with the EMA equipment.

Rigid range test. Regarding the results for this test in the stations in Guanajuato, the State of Mexico (Chapingo), and Sinaloa (Santa Rosa I AC) for the variables T_{max}, T_{min}, and T_{med}, no data was determined outside the range, according to Equation (6).

For the variables of maximum and mean relative humidity, the results of this test show that no data is outside the range of the condition of Equation (13) for the five stations in Guanajuato. However, for the minimum relative humidity, the station's El Galomo and El Jardín present a suspicious piece of information, each one: 22/05/2006 (0.05 %) and 07/09/2001 (0.05 %), respectively. For the station in Chapingo for variables HR_{max}, HR_{med}, and HR_{min} all the information in the period of study, except for blanks, fulfill the condition of the Equation (13), and in the Santa Rosa I AC station, for HR_{min} there is a piece of information (06/02/1999) that is suspicious (0.07 %).

Meek and Hatfiel (1994) report that there are patterns when data is being measured correctly by the sensor and that an unusual piece of information caused by an extreme weather event can be verified since there are patterns between related variables. These are shown, for example, in a low relative humidity figure, with one reported for that same day with a very high air temperature, and these patterns themselves in relatively nearby stations. Meek and Hatfiel (1994) report five days of lost figures since the sensor seems to have frozen in diverse intervals in that period. They mention that there are factors of relations that show this: 1) air temperature was reported to be below 0°; (2) vapor pressure was near saturation point; (3) the information for wind direction displays variation (4) the presence of ice on the tripod and instruments, and they claim that in practice, these data must not be deleted, but rather just flagged.

Flexible range or ephemeride. According to Table III, for this test, when considering Equation (7), it was found that for Guanajuato, variable T_{max} and the stations Agroeduca, El Cardonal, El Fuerte, El Galomo and El Jardin, all data fulfill the conditions of this equation, represented in Figure 1 and Figure 2.

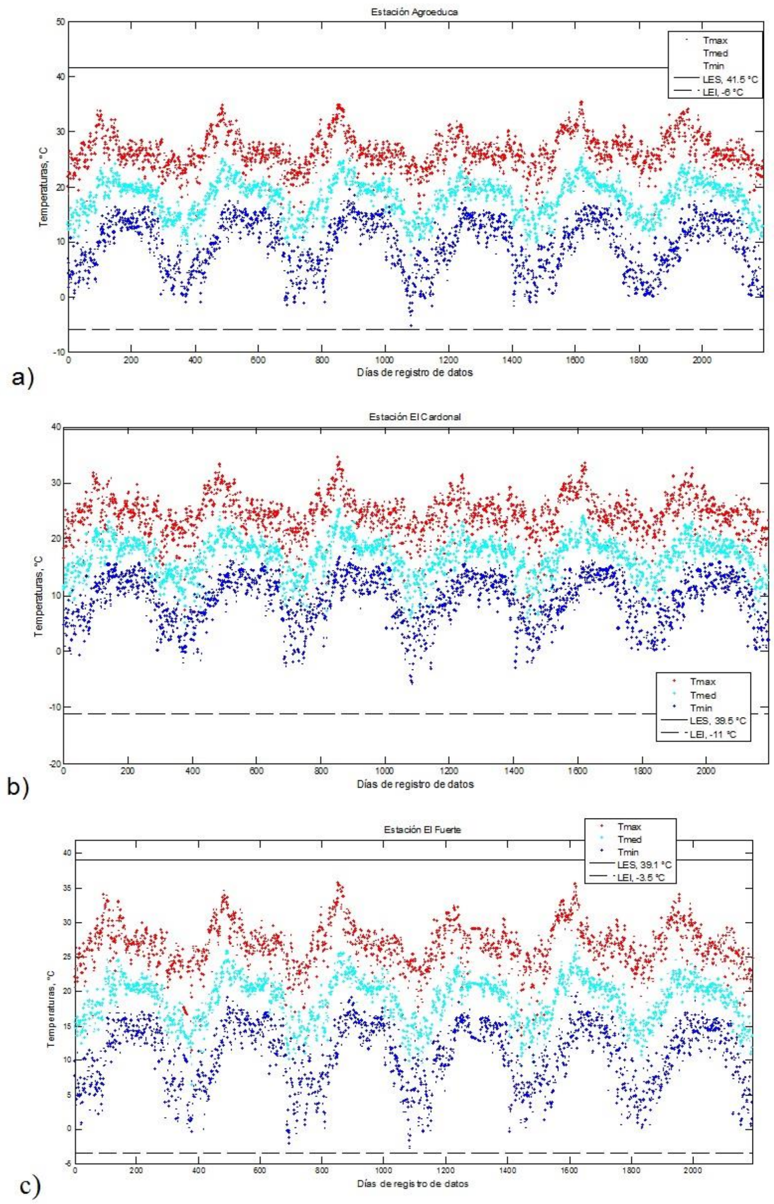


Figure 1. Variation of T_{max}, T_{med}, T_{min}, location of extreme values for the stations a) Agroeduca ($-6 > T_{max,min,med} > 41.5$), b) El Cardonal ($-11 > T_{max,min,med} > 39.5$) and c) El Fuerte ($-3.5 > T_{max,min,med} > 39.1$), LES Upper extreme limit, LEI Lower extreme limit.

Tecnología y Ciencias del Agua

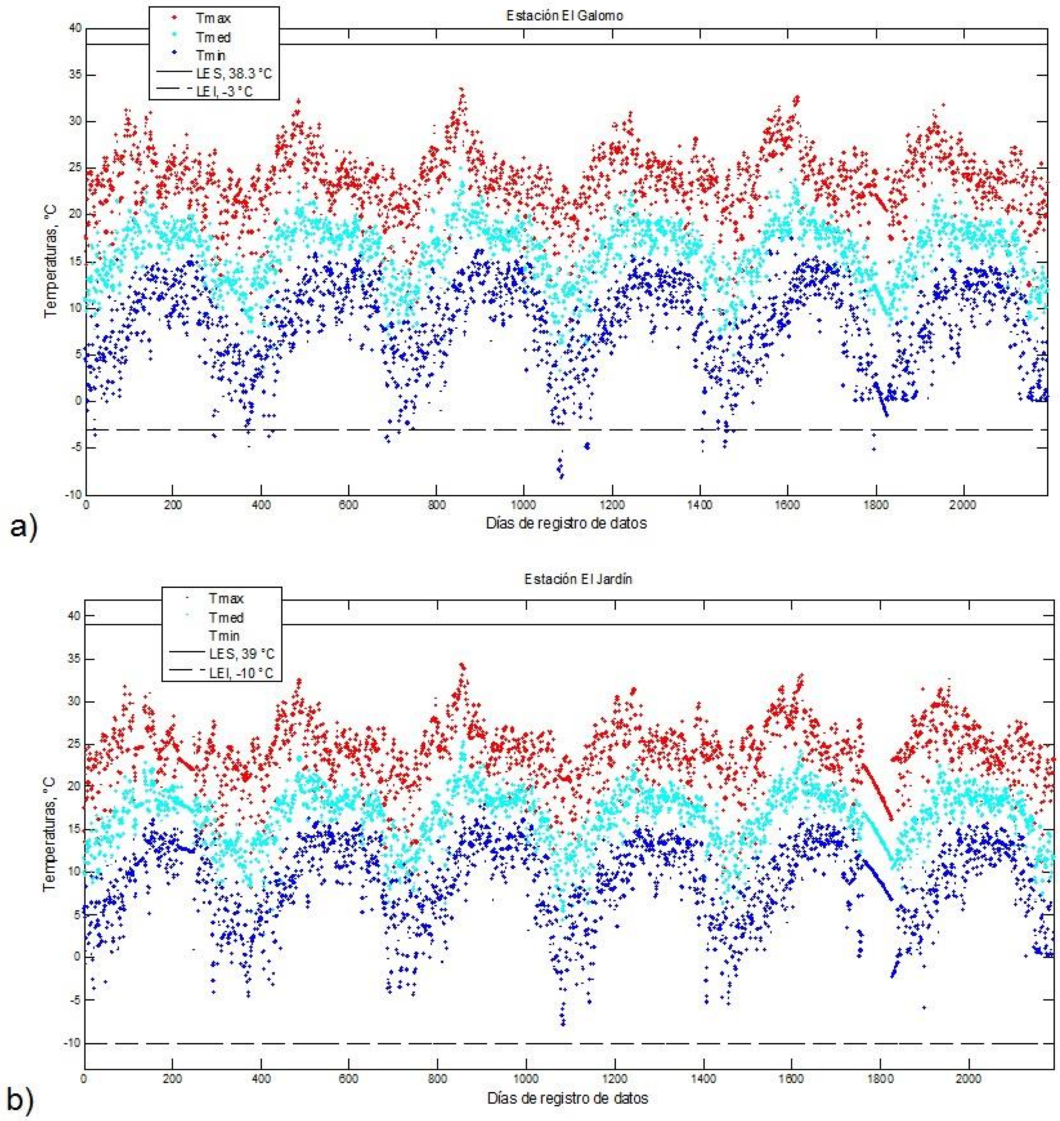


Figure 2. Variation of T_{max}, T_{med}, T_{min}, location of extreme values for stations a) El Galomo ($-3 > T_{max,min,med} > 38.3$) and b) El Jardín ($-10 > T_{max,min,med} > 39$). LES: Upper extreme limit; LEI: Lower extreme limit.

For the variable of minimum temperature (T_{min}) for the stations of Agroeduca, El Cardonal, el Fuerte and El Jardín, all data fulfill the conditions of extreme values of the climatological normals per station, although for the station El Galomo there are 34 data (1.55 %) that do not fulfill the condition, $-3 > T_{max,min,med} > 38.3$ (Equation (7)), placing them in the suspicious category. For Chapingo for T_{max} and T_{med}, all data fulfill the condition, $-9 < T_{max,min,med} < 38$ (Figure 3a). However, for T_{min} there is a spurious figure: 08/08/2011 (0.03 %). For Santa Rosa I AC, only the variable T_{med} did not display labelled data like suspicious, although the data of the variables T_{max} and T_{min} presented a suspicious figure each, with dates 14/10/1997 (0.07 %) and 6/02/1999 (0.07 %), respectively, which did not fulfill the condition, $0 < T_{max,min,med} < 43$ (Figure 3b). Meek and Hatfiel (1994) applied the LIM rule $(Ta_{min}(d) \leq Ta(d) \leq Ta_{max}(d))$ (which is the equivalent to Equation (7) in this work) at Walnut Creek, Treynor for Walnut Creek reported atypical data. Yet, they comment that it is valid since they analyzed a report of extreme weather (a severe thunderstorm) and a heavy snowstorm in the entire area experimented.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

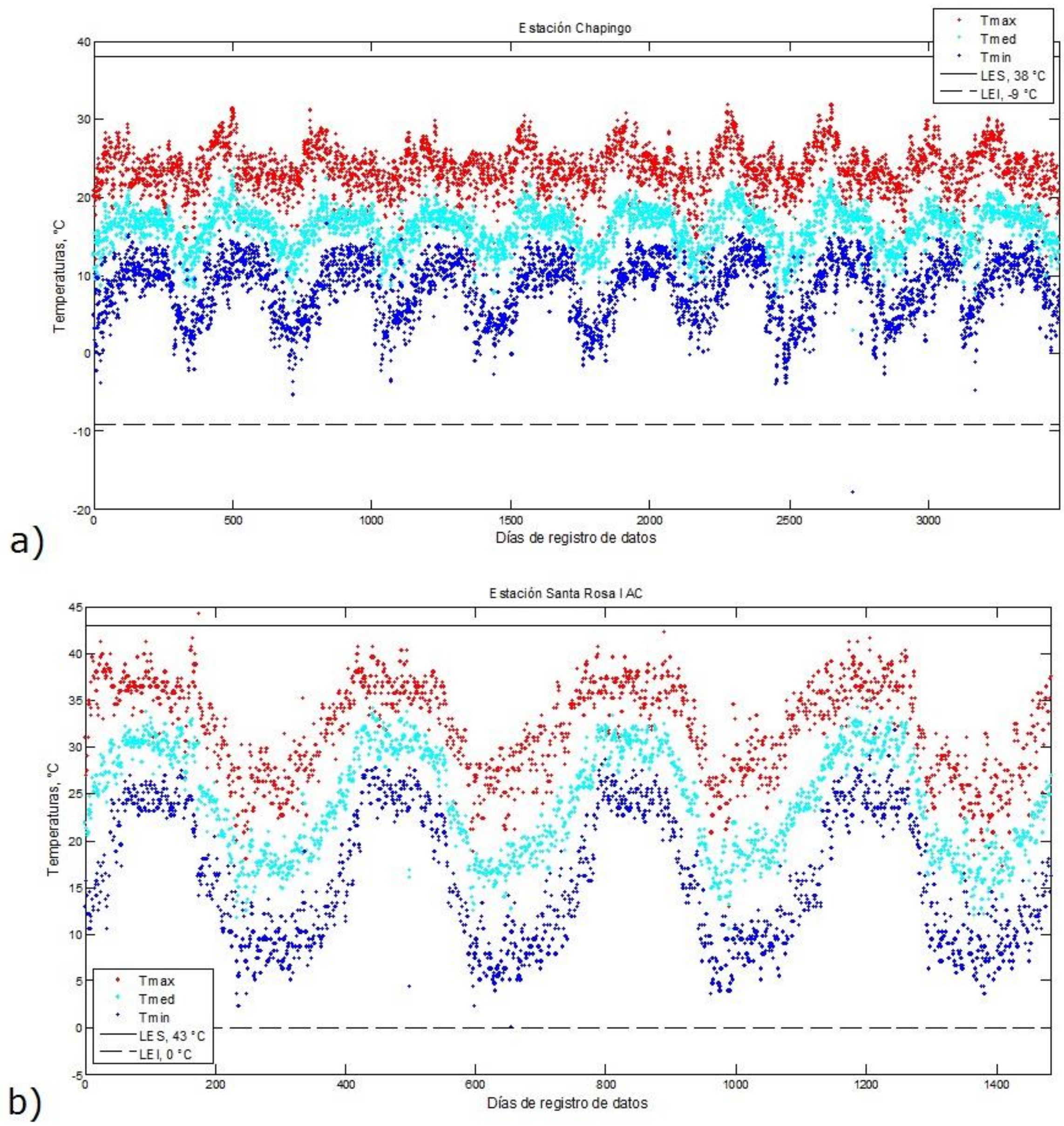


Figure 3. Variation of T_{max}, T_{med}, T_{min}, location of extreme values for stations a) Chapingo ($-9 < T_{max,min,med} < 38$) and b) Santa Rosa I AC ($0 < T_{max,min,med} < 43$), LES Upper extreme limit, LEI Lower extreme limit.

Likewise, Feng *et al*. (2004) report errors of 0.000035 % for T_{max}, for the test similar to Equation (7) that these authors use, $-89.4 < Td < 93.9\ °F$, and of 0.000009 % for T_{min}. The latter used the extreme temperature limits by Kubecka (2001) and Gleason (2002).

Flexible range test (ephemeride) for global radiation. Regarding the global radiation for this test (Equation (19)), values of R_{so} were obtained for the different stations in different locations. The envelopes of R_{so} are shown in Figure 4, Figure 5, and Figure 6. Here we can observe the condition of $RG < 1.1R_{so}$, used by Estévez *et al*. (2011), Allen (1996), and Moradi (2007), which considers the data above $1.1 \cdot R_{so}$ as suspicious.

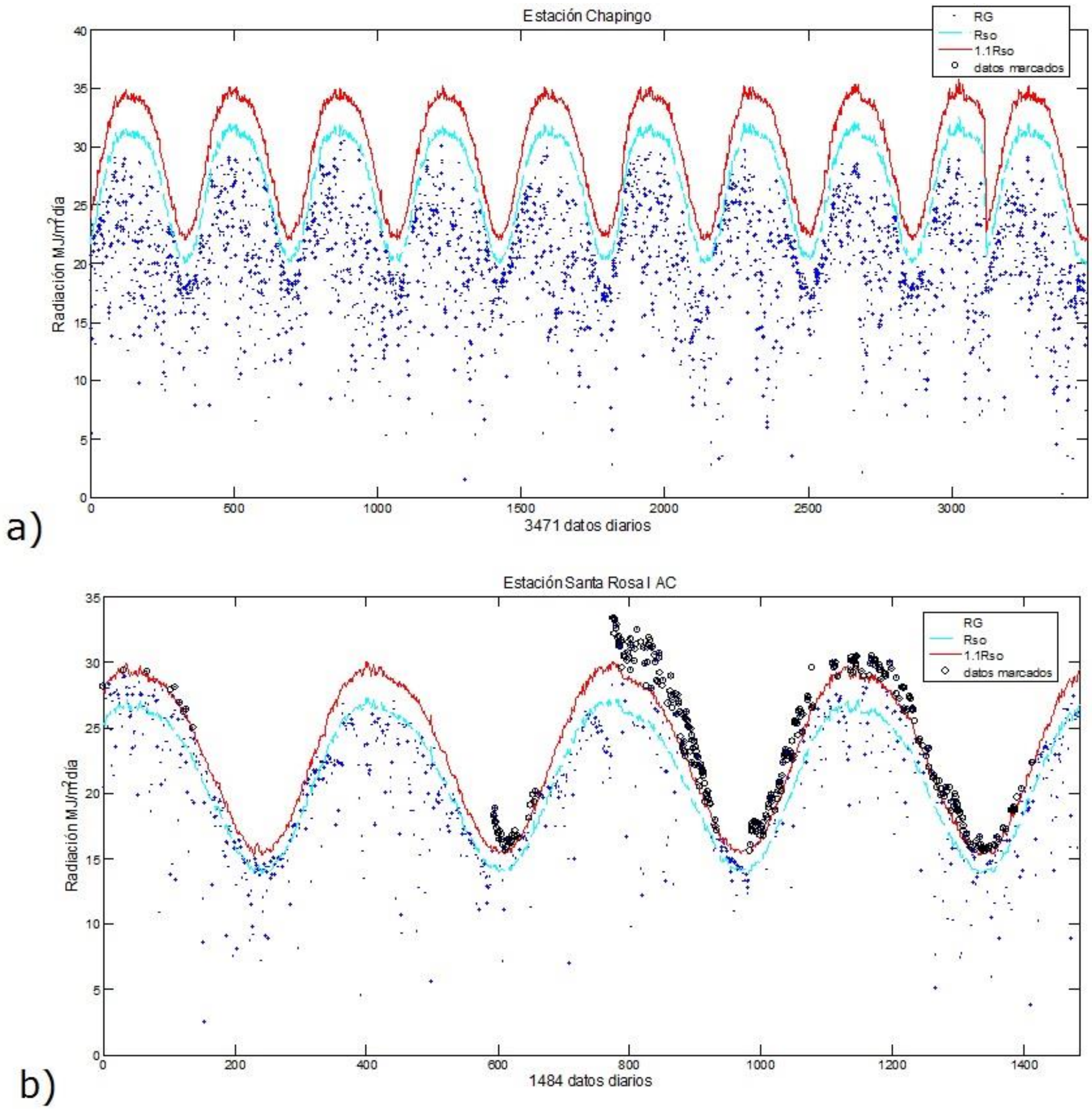


Figure 4. Envelope R_{so}, $1.1 \cdot R_{so}$ and RG for stations a) Chapingo and b) Santa Rosa I AC.

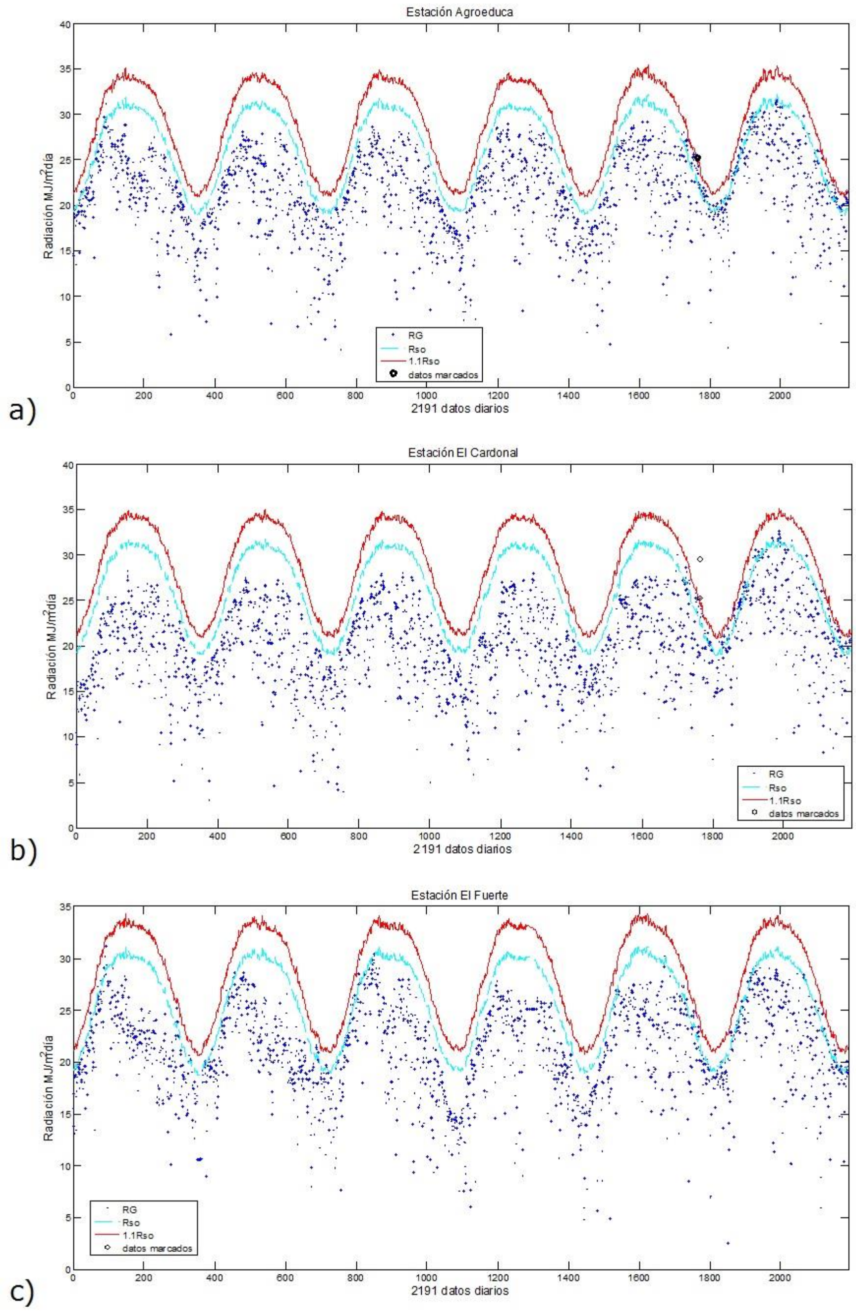
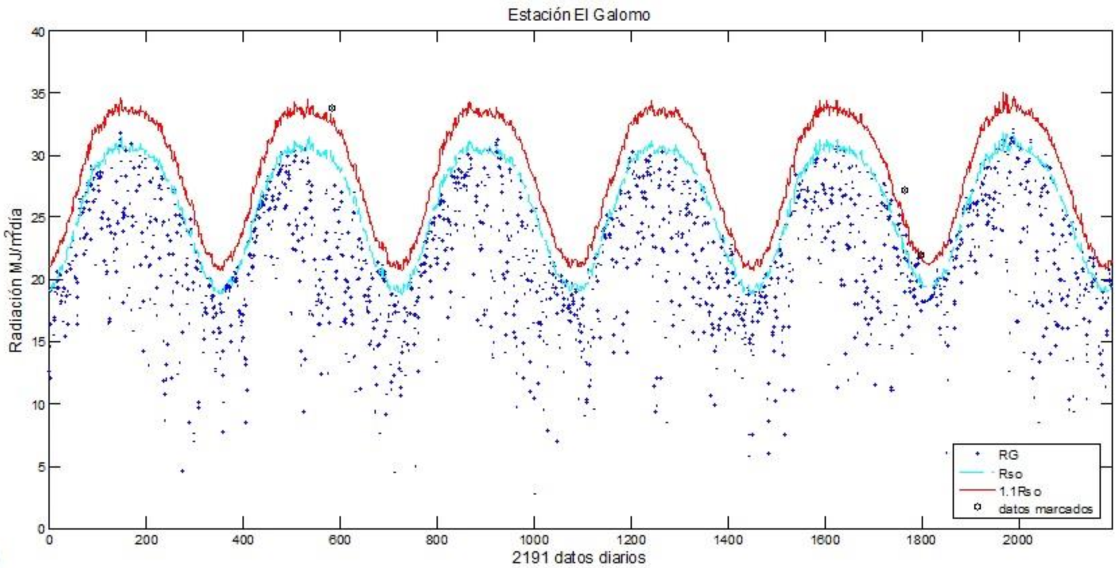
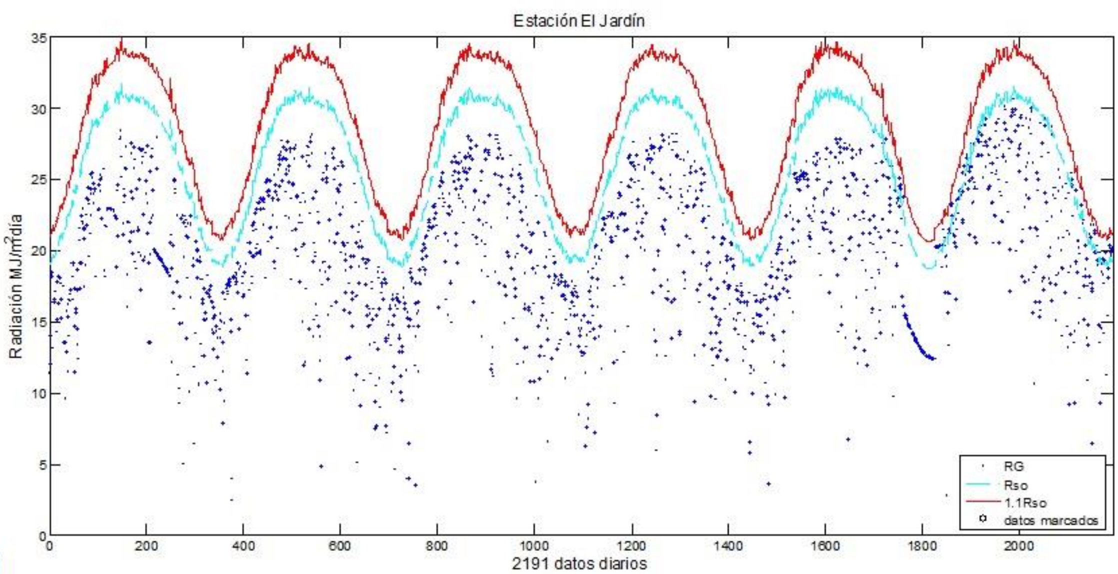


Figure 5. Envelope R_{so}, $1.1 \cdot R_{so}$ and *RG* for stations: a) Agroeduca, b) El Cardonal, and c) El Fuerte.



a)



b)

Figure 6. Envelope R_{so}, $1.1 \cdot R_{so}$ and RG for stations a) El Galomo and b) El Jardín.

For Chapingo (Figure 4a), they all fulfilled the condition for Equation (19). For the Santa Rosa I AC station (Figure 4b), 345 (23.25 %) data is unreliable (marked data). The sensor in this station began in 1998 to display malfunctions, specifically in the datum 775. In 1999, the sensor's damage was presented permanently since it marked solar radiation at night when it should have indicated zero values. In Las Cabezas of San Juan station in Seville. Estévez *et al*. (2011) found multiples measurements 20 % higher than the condition of Equation (19) between 2002 and 2006.

Regarding Guanajuato (Figure 5 and Figure 6), for the following stations, there are suspicious data that do not fulfill the condition of Equation (19): Agroeduca, one piece of data (0.05 %), 31/10/2005; El Cardonal, two data (0.09 %), 30/10/2005, 31/10/2005; El Galomo, three data: (0.14 %), 7/8/2002, 31/10/2005, 2/12/2005. Agroeduca, El Cardonal, and El Galomo are observed to report the same atypical piece of data (31/10/2005), which indicates that this condition presented itself in the general area, which may be due to very clear and clean skies. Therefore, it was not an erroneous measurement of the sensor, and these data must only be marked but are valid.

Internal consistency test. In the results for this test for Guanajuato, for the temperature variable, and all stations, all data fulfills the condition of Equation (8), with no reports of inconsistent data. Likewise, all data fulfill the condition of Equation (10). Yet, for Equation

(9), in the El Galomo station, there is one piece of data that does not fulfill the condition given, dated (10/01/2004), placing it in the category of inconsistent (0.05 %), whereas the other stations do fulfill it. For the Chapingo station, for the variable temperature, the objects of study do fulfill all the conditions of equations (8), (9) y (10). Regarding the Santa Rosa I station, for Equation (8), there are four inconsistent pieces of data (0.27 %) since they marked T_{max} equal to T_{med} and T_{min}. Also, for Equation (9), there is inconsistent data (0.07 %). For Equation (10), all data fulfill this condition.

For the relative humidity variable (Equation (14)), for all the stations in Guanajuato, all data are consistent with this equation, and the same occurs for Chapingo. However, for Santa Rosa I AC, for the relative humidity variable (Equation (14)), there are six pieces of inconsistent data (0.40 %).

Thermal jump (differences) test. In this test, for El Galomo, Agroeduca, El Cardonal, El Fuerte, and El Jardín, stations, the data (and percentages) Equation (11), are 95 values (4.34 %), 126 (5.75 %), 39 (1.78 %), 59 (2.69 %) y 133 (6.07 %), respectively. For Chapingo, there are 43 pieces of data (1.24 %), and for Santa Rosa I, 50 (3.37 %) are labeled as suspicious. Araya-López and Alfaro-Martínez (2008) applied an hourly thermal jump test on data from Manzanillo and Pavas in Costa Rica, are found that between hourly temperature data, there was an atypical value of the difference (15°C). They state that after comparing this piece of data with one form a nearby station, it became clear that the field of temperatures of the region underwent no variations, and it was, therefore, erroneous data.

Persistence test. For the variable of maximum relative humidity for Guanajuato, all stations presented data that do not fulfill the condition of Equation (15), likewise, in Chapingo and Santa Rosa I AC. Suspicious data are Agroeduca, 246 data (11.23 %); el Cardonal, 116 data (5.29 %); el Fuerte, 62 data (2.83 %); el Galomo, 972 (44.34 %); el Jardín, 40 (1.82 %); Chapingo, 603 (17.37 %); and Santa Rosa I AC, 1149 (77.43 %).

In the same way, for the variable of minimum relative humidity, all stations in all the locations presented suspicious data that did not fulfill the condition of Equation (16), and they are Agroeduca, 20 data (0.91 %); el Cardonal, 22 data (1.004 %); el Fuerte, 22 (1.004 %); el Galomo, 8 (0.36 %); el Jardín, 10 (0.46 %); Chapingo, 349 (10.05 %); and Santa Rosa I AC, 192 data (12.94 %).

Allen (2008) comments that relative humidity data below 30 % in sub-humid areas are not expected values.

For the mean relative humidity, the stations that presented suspicious data are el Cardonal, 6 data (0.27 %); El Fuerte, 4 data (0.18 %); el Galomo, 4 (0.18 %); El Jardín, 6 (0.27 %); Chapingo, 255 (7.35 %) and Santa Rosa I AC, 141 data (9.50 %). Agroeduca displayed not erroneous data for this variable. Santa Rosa I, in this persistence test, presented values of 100 % in continuous days (one after the other), $HR_{max}(d) = HR_{max}(d-1) = HR_{max}(d-2) = 100\ \%$ with 77.43 % values that do not fulfill the condition (Equation (15)). Although it is an argument that Santa Rosa I is an irrigation area, therefore it presents measurements of 100 % in diverse intervals of 5 minutes during one day, particularly at sunrise, when the dew point is similar to the minimum temperature. Therefore, humidity reaches 100 % or close to this value and is also close to the coast (Allen *et al*., 1998). In this station, the maximum and minimum values were obtained out of the data taken every five minutes for one day. Having mentioned this, the denomination of "erroneous" data must be reserved for this percentage, specifically for this test. This is corroborated by Meek and Hatfiel (1994) when they comment that, aside from the abrupt or severe events, the marked observations (flagged) are not frequent on a daily basis. They also comment that there is only one possible exception for the breaking of the higher LIM rule in *e*(*d*) (daily vapor pressure, which is directly related to the relative humidity), which corresponds to Equation (13) of rigid range on the right ($0.8 < HR_{max}, HR_{min}, HR_{med} < 100$) for this work, and they add that, nevertheless, for practical purposes, the air could be considered saturated in particular moments of dawn.

In this test for the variable of global radiation, in the different stations under study, the data that did not fulfill the condition of Equation (18) are for Agroeduca, 12 data (0.55 %); el Cardonal, 10 data (0.46 %); el Fuerte, 2 (0.09 %); El Galomo, 16 (0.73 %); El Jardín, 16 (0.73 %); Santa Rosa I AC, 16 (1.08 %) and Chapingo, 14 data (0.40 %).

Zero test. For the zero tests with the variable T_{min}, all the stations covered by this study valued the condition of Equation (12) as false for all data, meaning no data was marked as suspicious for this test.

Schroeder *et al*. (2005) apply this zero test and the range of step, persistence, and spatial tests for data from approximately 40 stations in Western Texas for observation periods of 5 and 15 minutes and 30-60 minutes. These authors accumulate errors in all tests. For the temperature variable for the different heights from ground levels 1.5, 9, and 2 m, they found errors of 0.003, 0.010, and 0.009 %, respectively.

In general terms, in the flexible range test, except for Santa Rosa I, for the variable solar radiation, the internal persistence test (23.25 %), the internal persistence test for the variable of maximum relative humidity in Santa Rosa I and El Galomo (Guanajuato, 44.36 %), the remaining stations did not present spurious data of over 20 % as reported by Araya-López (2011), who claims that the percentage of atypical relative humidity data was higher than that found for temperature data, which was never higher than 20 % of the total data from the station network that make up the National Meteorology Institute of Costa Rica, and reporting percentages of atypical data at different stations for maximum temperatures of up to 1.2 and 7.5 %, for minimum temperatures of 17.8 %; for relative minimum humidity, reports of atypical data are of 4.3 and 8.7 %, and relative maximum humidity of 4.3 %.

For all the tests applied in this study, some location or another presented data marked as suspicious, spurious, or inconsistent.

The rigid range test for HR_{max}, requires reconsideration, at least for Santa Rosa I AC, which is an irrigation area, since it can, in continuous days, reach HR_{max} of 100 %, as an exceptional condition for the same reason.

Finally, the data analyzed according to Figure 1, Figure 2, Figure 3, and Figure 4 present a seasonal and periodic behavior, but since less than 10 years of data are available for each and every one of the stations under study, it is not possible to observe a defined trend (at least 50 years are required according to WMO, 2007) and what according to the IPCC (2001) is increasing for temperature, therefore there would be a change in the extreme limits of Equation (7) of this work for each season. Then this would lead to the recalibration of these equations.

Conclusions

The results of the procedures applied to historical data support determining the reliability of the EMA's meteorological information and establishing if it can be used in the field of interest, having the confidence that correct inferences will be reached or discarded that station or certain variable. They also indicate that the sensors have worked within their range of accuracy and allow to determine suspicious, unreliable or inconsistent data, which helps to decide. Implementing these procedures in real time is recommended since they will indicate suspicious data that are easier to verify and, at the same time, define if a sensor begins to measure unreliable data.

References

AENOR, Asociación Española de Normalización y Certificación. (2004). *Redes de estaciones meteorológicas automáticas: Directrices para la validación de registros meteorológicos procedentes de redes de estaciones automáticas (Validación en tiempo real). UNE 500540:2004*. Madrid, España: Asociación Española de Normalización y Certificación.

Allen, R. G., Smith, M., Pereira, L. S., & Perrier, A. (1998). Crop evapotranspiration: Guidelines for computing crop water requirements. In: *FAO Irrigation and Drainage Paper No. 56*. Rome, Italy: Food and Agriculture Organization.

Allen, R. G. (1996). Assessing integrity of weather data for reference evapotranspiration estimation. *Journal of Irrigation and Drainage Engineering*, ASCE 122, 97-106. Recovered from https://doi.org/10.1061/(ASCE)0733-9437(1996)122:2(97)

Allen, R. G. (2008). Quality assessment of weather data and micrometeorological flux – Impacts on evapotranspiration calculation. *Journal of Agricultural Meteorology*, 64(4), 191-204. DOI: 10.2480/agrmet.64.4.5

Araya-López, J. L. (2011). Resultados de un control de calidad para datos de temperatura superficial de aire y humedad relativa. *Tecnología en Marcha,* 24, 33-49.

Araya-López, J. L., & Alfaro-Martínez, E. J. (2008). Prueba e implementación de algoritmos de control de calidad de datos de temperatura superficial del aire en un contexto operativo. *Tecnología en Marcha*, 21(3), 47-61.

ASCE-EWRI, American Society of Civil Engineers- Environmental & Water Resources Institute. (2005). *The ASCE Standardized Reference Evapotranspiration Equation. ASCE-EWRI Standardization of Reference Evapotranspiration Task Committee Report*. Recovered from hhttp://www.kimberly.uidaho.edu/water/asceewri/

Bulege-Gutiérrez, W., & Custodio, M. (2020). Cambio climático y retroceso glaciar en la cordillera Huaytapallana, Perú. *Tecnología y Ciencias del Agua*, 11(2), 229-261. DOI: 10.24850/j-tyca-2020-02-06

Campos-Aranda, D. F. (2020). Detección de registros homogéneos en 16 series amplias de precipitación anual del Altiplano Potosino, México. *Tecnología y Ciencias del Agua*, 11(3), 107-157. DOI:10.24850/j-tyca-2020-03-04

Estévez, G. J. (2008). *Diseño de un sistema integrado para el control de calidad de datos de estaciones meteorológicas automáticas* (disertación doctoral). Departamento de Agronomía, Universidad de Córdoba, España.

Estévez J., García-Marín, A. P., Morábito, J. A., & Cavagnaro, M. (2016). Quality assurance procedures for validating meteorological input variables of reference evapotranspiration in Mendoza province (Argentina). *Agricultural Water Management*, 172, 96-109. Recovered from https://doi.org/10.1016/j.agwat2016.04.019

Estévez, J., Gavilán, P., & Girálvez, J. V. (2011). Guidelines on validation procedures for meteorological data from automatic weather stations. *Journal of Hydrology*, 402, 144-154. DOI: 10.1016/j.jhydrol.2011.02.031

Feng, S., Hu, Q., & Qian, Q. (2004). Quality control of daily meteorological data in China, 1951-2000: A new dataset. *International Journal of Climatology*, 24, 853-870. DOI: 10.1002/joc.1047

Gleason, E. (2002). *Global daily climatology network, V 1.0*. Asheville, USA: National Climatic Data Center.

IPCC, Intergovernmental Panel on Climate Change. (2001). *Climate change: The scientific basis. Contribution of Working Group I to the third assessment report of the Intergovernmental Panel on Climate Change*. Cambridge, UK, and New York, USA: Cambridge University Press.

Kubecka, P. (2001). A possible world record maximum natural ground surface temperature. *Weather*, 56, 218-221. Recovered from https://doi.org/10.1002/j.1477-8696.2001.tb06577.x

Majumdar, N. C., Mathur, B. L., & Kaushik, S. B. (1972). Prediction of direct solar radiation for low atmospheric turbidity. *Solar Energy*, 13, 383-394. DOI: 10.1016/0038-092X(72)90004-7

Meek, D. W., & Hatfiel, J. L. (1994). Data quality checking for single station meteorological databases. *Agricultural and Forest Meteorology*, 69(1-2), 85-109. Recovered from https:doi.org/10.1016/0168-1923(94)90083-3

Moradi, I. (2007). Quality control of global solar radiation using sunshine duration hours. *Energy*, 34(1), 1-6. DOI: 10.1016/j.energy.2008.09.006

Obrien, K. J., & Keefer, T. N. (1985). Real-time data verification. Computer applications in water resources. In: *Procedures. ASCE Spec. Conf.* (pp. 764-770). Buffalo, New York, USA .

OMM, Organización Meteorológica Mundial. (2007). *Función de las normales climatológicas en un clima cambiante. Programa mundial de datos y vigilancia del clima*. Ginebra, Suiza: Organización Meteorológica Mundial.

Reek, T., Doty, S. R., & Owen, T. W. (1992). A deterministic approach to the validation of historical daily temperature and precipitation data from the Cooperative Network. *Bulletin of the American Meteorological Society*, 73, 753-762.

Schroeder, J. L., Burgett, W. S., Haynie, K. B., & Sonmez, I. (2005). The West Texas Mesonet: A technical overview. *Journal of Atmospheric and Oceanic Technology*, 22, 211-222. Recovered from https://doi.org/10.1175/JTECH-1690.1

Shafer, M. A., Fiebrich, C. A., & Arndt, D. S. (2000). Quality assurance procedures in the Oklahoma Mesonetwork. *Journal of Atmospheric and Oceanic Technology*, 17, 474-494. Recovered from https://journals.ametsoc.org/doi/abs/10.1175/1520-0426%282000%29017<0474%3AQAPITO>2.0CO;2

UNE 500540. (2004). *Redes de estaciones automáticas. Directrices para la validación de registros meteorológicos procedentes de redes de estaciones automáticas. Validación en tiempo real*. Madrid, España: Asociación Española de Normalización y Certificación.

WMO, World Meteorological Organization. (1996). *Guide to Meteorological Instruments and Methods of Observation* (WMO-No. 8). Geneva, Switzerland: World Meteorological Organization.