

Control de calidad de las series de temperatura diaria de las estaciones del INIA-Venezuela en el período 1950-2005

Quality control of the series of daily temperature for the stations of the INIA-Venezuela in the period 1950-2005

Raquel Parra¹, Adriana Cortez² y María F. Rodríguez²

¹Profesora. Universidad Central de Venezuela (UCV). Departamento de Ingeniería Agrícola. Núcleo Maracay.

²Investigadoras. INIA-CENIAP. Recursos Agroecológicos, Maracay 2105, estado Aragua. Venezuela.

Correos electrónicos: parrar@agr.ucv.ve, acortez@inia.gob.ve

RESUMEN

El control de calidad de los datos como una fase inicial en cualquier estudio que contemple procedimientos de cálculo, es un paso indispensable que garantiza la veracidad de los resultados e interpretaciones de los mismos. El problema más notable con el control de calidad de datos climáticos es la presencia de valores atípicos, ya que removerlos o mantenerlos puede afectar el análisis de eventos llevando a sobreestimaciones o subestimaciones. En el presente trabajo se aplicó una metodología para el control de calidad de datos de temperatura a través de la integración de distintos análisis que incluyen la identificación de valores atípicos con medidas estadísticas robustas, para lo cual se utilizaron los datos de 15 estaciones del Instituto Nacional de Investigaciones Agrícolas (INIA). En los resultados se encontró que la mayoría de las observaciones se ubicaron dentro de los umbrales esperados. En el análisis de las series temporales se detectaron casos de temperatura máxima ($T_{\text{máx}}$) $>50\text{ }^{\circ}\text{C}$ y temperatura mínima ($T_{\text{mín}}$) $>T_{\text{máx}}$ (OTD $<0\text{ }^{\circ}\text{C}$), que fueron eliminados. En la estación del Centro Nacional de Investigaciones Agropecuarias (CENIAP-Maracay), se encontraron valores muy elevados de FD10 (10 días en 1959 con $T_{\text{mín}} < 10\text{ }^{\circ}\text{C}$) y de SU35 (183 días en 1959 con $T_{\text{máx}} > 35\text{ }^{\circ}\text{C}$). Se demostró que a través de la construcción de límites con media y desviación bponderadas robustas se pueden detectar series de datos dudosos, como el caso (en la estación antes mencionada) de la serie de temperatura mayo-noviembre/1959, con $T_{\text{máx}} > \text{límite superior}$ y $T_{\text{mín}} < \text{límite inferior}$. Los valores cuestionados fueron reportados en la base de datos depurada.

Palabras clave. Control de calidad, estaciones climatológicas, valores atípicos, estadísticos bponderados, índices climáticos.

ABSTRACT

The data quality control as an initial phase in any study that involves calculation, it is a vital step that ensures the accuracy of the results and interpretations of them. The most notable problem with the quality control of climate data is the presence of outliers due that, removing them or keep them can affect the analysis of events leading to overestimates or underestimates. In the present work was implemented a methodology for quality control of temperature data through the integration of various analysis that include the identification of outliers with robust statistical measures in 15 climatic stations of INIA. It was found that the majority of the comments fell down within the expected thresholds. In the analysis of the temporal series, some cases were detected with $T_{\text{max}} > 50$ and $T_{\text{max}}\text{ }^{\circ}\text{C} < T_{\text{min}}$ (OTD $< 0\text{ }^{\circ}\text{C}$), which were eliminated. In the station CENIAP-Maracay it was found very high values of FD10 (10 days in 1959 with $T_{\text{min}} < 10\text{ }^{\circ}\text{C}$) and SU35 (183 days in 1959 with $T_{\text{max}} > 35\text{ }^{\circ}\text{C}$). It was demonstrated that through the construction of intervals with mean and deviation bi-weighted, it is possible to detect series with questionable data, as the temperature serie may-november/1959 of CENIAP-Maracay where $T_{\text{max}} >$ than the upper limit and $T_{\text{min}} <$ than the lower limit. The questioned values were reported in the refined data base.

Key words: Quality control, weather stations, outliers, bi-weighted statistical, climatic indexes.

INTRODUCCIÓN

El análisis de los patrones espacio-temporales futuros de la precipitación y la temperatura son incluidos en el análisis de aspectos relevantes, como el cambio climático y sus implicaciones en el confort térmico y la disponibilidad de agua a través de la construcción de escenarios, en gran parte fundamentados por información climática histórica de largo registro (MARN, 2005). De los 27 índices básicos recomendados por el equipo de expertos de CCI/CLIVAR (Índices para el estudio de Cambio Climático/ Variabilidad Climática y Previsibilidad, de sus siglas en inglés) para el monitoreo y detección de cambio climático (ETCCDMI), 16 están basados en temperatura (Zhang y Yang, 2004). Sin embargo, la calidad y veracidad de estos índices va a depender en gran medida de que los datos básicos tengan un control de calidad adecuado antes de que sean utilizados.

El Instituto Nacional de Investigaciones Agrícolas (INIA) posee una red de 20 estaciones diseminadas en estados de importancia agrícola para el país. Algunas de estas poseen más de 50 años de valores diarios de temperaturas, observadas en termómetros de máxima y de mínima (T_{\max} y T_{\min}), lo que significa más de 250 mil registros en la base de datos.

Rodríguez *et al.* (2011) en su trabajo titulado "Integración espacial y aplicación de los datos agroecológicos del INIA al norte del Orinoco de Venezuela" consolidó una base de datos climática con información de 263 estaciones climatológicas en todo el país provenientes de distintas instituciones.

En este trabajo se observó que el problema más notable en el control de calidad de las variables climatológicas, es la presencia de valores atípicos u outliers, ya que removerlos (siendo verdaderos) puede afectar el análisis final de eventos climáticos extremos, y mantenerlos (siendo erróneos) puede llevar a sobreestimaciones de los mismos.

Según Lobo *et al.* (2005), los valores atípicos pueden clasificarse en cuatro categorías: a) los que surgen de un error de procedimiento (transcripción a las planillas o base de datos); b) extremos asociados a un evento extraordinario;

c) atípicos que caen dentro del rango de las variables pero que son únicas en la combinación de los valores de dichas variables; d) atípicos para los que el investigador no tiene explicación.

Aquellos valores dudosos fácilmente detectables por su improbabilidad de ocurrencia, son eliminados de la base de datos y reportados en las observaciones; pero, como lo indica Edwards (1998), la eliminación de "outliers" no debería ser el objetivo principal del control de calidad, a menos que esté bien sustentado. Existen casos en los que se debe asociar una medida de tendencia central con una medida de dispersión que sean lo suficientemente robustas y determinen límites por fuera de los cuales las observaciones deben ser consideradas dudosas sin llegar a ser necesariamente eliminadas. Para tal fin, National Climatic Data Center (2002) propone calcular la media y desviación bponderadas (Lanzante, 1995) para la construcción de los límites de control.

El objetivo de este trabajo fue aplicar una metodología para el control de calidad de datos de temperatura de las estaciones climatológicas del INIA-Venezuela en el período 1950-2005, a través de la integración de distintos análisis que van desde los filtros de rutina, la utilización de índices climáticos, hasta la identificación de valores atípicos con medidas estadísticas robustas.

MATERIALES Y MÉTODOS

Evaluación y selección de las estaciones climatológicas para el estudio

Se comenzó por verificar y estructurar la metadata de las estaciones del INIA, incluyendo:

- Serial nacional: código con el que se identifica a la estación.
- Estación: nombre de la estación.
- Estado u ubicación: entidad en la que se encuentra.
- Cota: altitud en m s. n. m. a la que se encuentra la estación.
- Latitud y longitud: ubicación en grados, minutos y segundos con el fin de permitir la georeferenciación (Figura 1).

- Fecha de instalación y eliminación (si fuera el caso).
- Período de registro de las variables climáticas que se miden en la estación.

De las 20 estaciones climáticas del INIA, analizadas utilizando el criterio de: 1) período corto de registro y 2) más del 50% de datos faltantes, se seleccionaron 15 (Cuadro 1) en el período 1950-2005. Las restantes se omitieron por poseer un período de registros inconvenientemente corto, es decir, con menos de 10 años y más 50% de datos faltantes.

Se utilizaron los datos de temperatura tomados de los termómetros de mercurio de T_{máx} y T_{mín} registrados diariamente.

Proceso de evaluación de la calidad de los datos en las estaciones climáticas seleccionadas

Se consideraron los siguientes análisis de control de calidad:

Número de datos faltantes diarios: para la generación de datos mensuales se consideró la condición de que si existían más de 10 datos faltantes en el mes ($df > 10$), no se calculaba el valor mensual (se consideraba faltante). El valor mensual es el promedio de los valores dentro del mes (protocolo de control de calidad básico aplicado en la Unidad de Agrometeorología del Centro Nacional de Investigaciones Agropecuarias (CENIAP-Maracay), INIA).

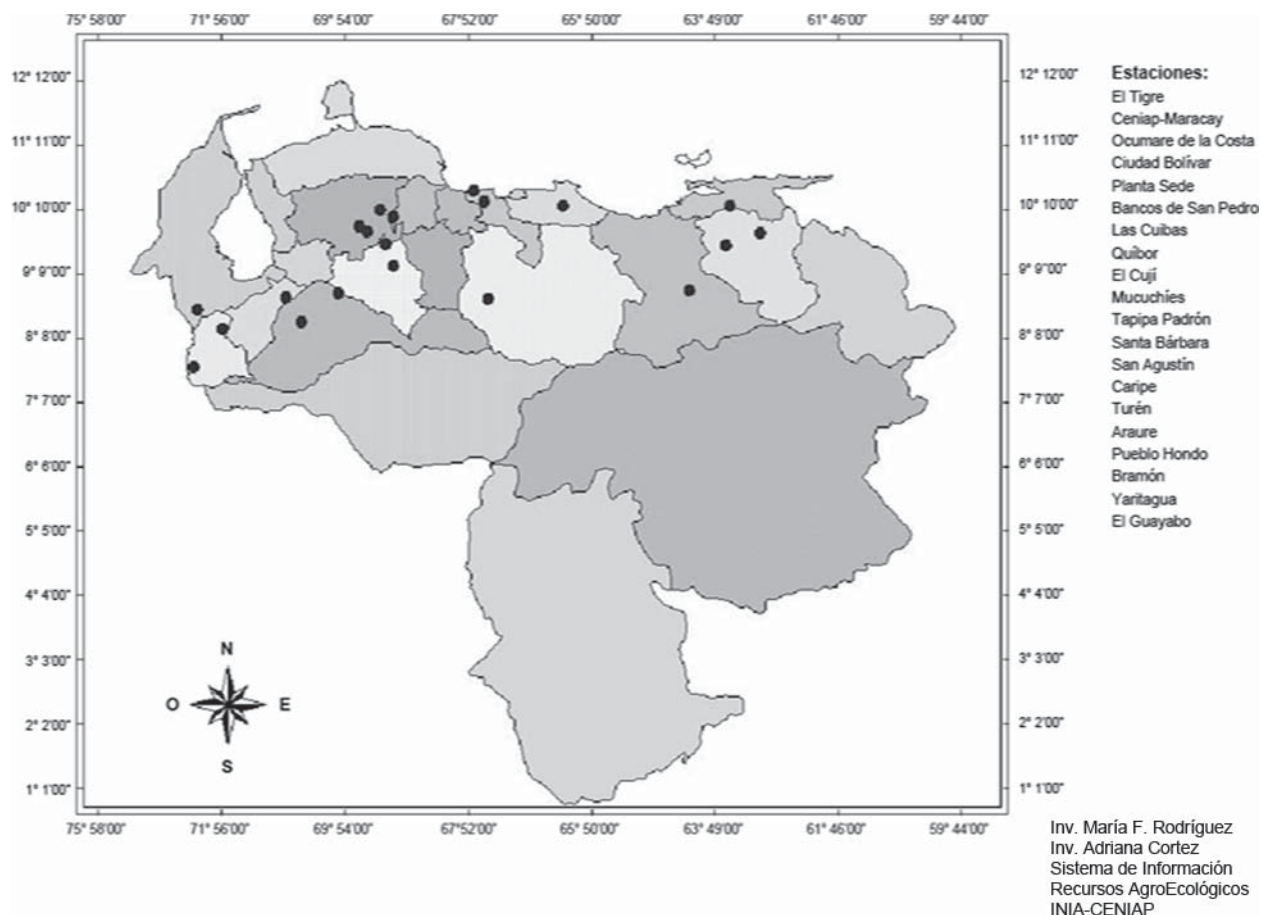


Figura 1. Red de estaciones climatológicas del Instituto Nacional de Investigaciones Agrícolas.

Cuadro 1. Lista de estaciones del Instituto Nacional de Investigaciones Agrícolas en Venezuela, sometidas al control de calidad de los datos de temperatura.

N°	Estación	Estado	Serial	Tipo	Altitud m.s.n.m.	Latitud gra° min' seg"	Longitud gra° min' seg"	Período de registro
1	El Tigre	Anz.	3715	CL	302	8° 51' 46"	64° 13' 17"	1971-2000
2	CENIAP-Maracay	Ara.	2604	CL	480	10° 17' 14"	67° 36' 2"	1950-2005
3	Ocumare de la Costa	Ara.	4701	CL	34	10° 27' 36"	67° 46' 22"	1963-2004
5	Ciudad Bolivia	Bar.	3503	CL	212	8° 22' 22"	70° 36' 52"	1988-2003
6	Bancos de San Pedro	Gua.	7502	CL	120	8° 44' 47"	67° 32' 6"	1958-2005
7	El Cuji	Lar.	1301	CL	605	10° 8' 51"	69° 18' 53"	1964-2003
9	Quibor	Lar.	2302	CL	705	9° 53' 14"	69° 39' 11"	1980-2003
10	Mucuchies	Mér.	7901	CL	3130	8° 45' 51"	70° 53' 9"	1961-2003
11	Tapipa-Padrón	Mir.	2308	CL	76	10° 13' 14"	66° 17' 57"	1977-2004
15	Araure	Por.	6201	CL	259	9° 36' 58"	69° 14' 10"	1994-2005
16	Colonia Turén	Por.	2001	CL	149	9° 15' 58"	69° 5' 54"	1970-2005
17	Bramón	Tác.	6303	CL	1150	7° 39' 3"	72° 24' 3"	1941-2003
18	Pueblo Hondo	Tác.	2902	CL	2418	8° 15' 9"	71° 54' 40"	1983-2003
19	Yaritagua	Yar.	0103	CL	336	10° 2' 38"	69° 5' 20"	1950-2004
20	El Guayabo	Zul.	6302	CL	61	8° 34' 1"	72° 20' 4"	1950-2004

Fuente: Inventario de estaciones climatológicas del INIA Venezuela.

Análisis de concentración y tendencia de las series temporales: tal como señala Shaw (1988), cuando se utilizan las series temporales para estudiar el comportamiento y la calidad de los datos, deben tomarse en cuenta los componentes que pueden estar definiendo su oscilación y tendencia:

- El componente de tendencia indica un incremento lineal, en términos absolutos, en el promedio a lo largo del tiempo.
- Por otro lado, la oscilación cíclica, para un retardo dado, determina el componente de estacionalidad.
- También se observan los valores picos aislados a lo largo de la serie fuera del rango estacional que son detectados como atípicos o extremos; y el componente estocástico aleatorio que representa una variación irregular, pero continua dentro de la serie y que debe tener alguna persistencia.

Para estudiar la tendencia y la ocurrencia de valores atípicos evidentes (improbables), se analizaron las series de T_{\max} y T_{\min} y oscilación térmica diaria (OTD) con relación a los años, y se aplicaron filtros lógicos en Excel, dependiendo de las condiciones particulares de cada estación. Por ejemplo, la OTD calculada como la diferencia de T_{\max} menos T_{\min} en el día, no puede arrojar valores negativos o iguales a cero.

Cálculo preliminar de los índices climáticos de temperatura:

Un conjunto de siete índices de temperatura fue seleccionado para este estudio y su descripción se muestra en el Cuadro 2. Vincent *et al.* (2005) utilizaron los índices para identificar valores de T_{\max} y T_{\min} extremas.

Los índices indicados en Cuadro 2 se basan en la máxima de las T_{\max} y T_{\min} (TXx y TNx), la

mínima de las T_{máx} y T_{mín} (TX_n y TN_n), y en el número de días en el año, donde la T_{mín} y T_{máx} son menores a 10 y 25 °C (FD10 y ID25, respectivamente) y la T_{máx} es mayor a 35 °C (SU35).

Construcción de los límites de control para la detección de valores atípicos

Para la construcción de los límites de control para la detección de valores atípicos (1), se utilizó la metodología de National Climatic Data Center (2002), que propone la incorporación de la media y desviación bponderadas (Lanzante, 1995), como se indica en la ecuación 1:

$$\bar{X}_{Bi} \pm 4s_{Bi} \quad (1)$$

Los estadísticos bponderados son promedios tales que la ponderación disminuye a medida que los datos se alejan del centro de la distribución (Feng *et al.*, 2004). En este caso tanto la media como la desviación se basan en estadísticos de orden como mediana (M) y MAD mediana de las desviaciones absolutas con respecto a M (MAD).

Un factor de ponderación $u_i = (X_i - M)/(c * MAD)$ es calculado para las n observaciones X_i y para

cualquier $|u_i| \geq 1$ se tiene que $u_i = 1$. La constante **c** representa un valor censor que controla la distancia, desde el centro de la distribución, en la que, la ponderación es equivalente a cero. En este caso se utilizó un $c=7,5$ por Lanzante (1995) y sugerido por National Climatic Data Center (2002). La media (\bar{X}_{Bi}) y la desviación bponderadas (s_{Bi}) se calculan como se indica en las ecuaciones 2 y 3, respectivamente.

$$\bar{X}_{Bi} = M + \left\{ \left[\frac{\sum (X_i - M)(1 - u_i^2)^2}{\sum (1 - u_i^2)^2} \right] \right\} \quad (2)$$

$$s_{Bi} = \left[\frac{n \sum (X_i - M)^2 (1 - u_i^2)^4}{\sum (1 - u_i^2)(1 - 5u_i^2)} \right]^{0.5} \quad (3)$$

Depuración de la base de datos

Esta fase consiste en la consolidación de una base de datos con la incorporación a la metadata de las observaciones y criterios de decisión debidamente justificados, para los valores eliminados (atípicos improbables) y dudosos (fuera de los límites robustos), según las metodologías utilizadas.

Para todos los análisis se utilizaron los paquetes estadísticos RCLimDex (1.0) y Excel ®.

Cuadro 2. Índices climáticos de temperatura.

ID	Nombre indicador	Elemento	Definición	Unidades
TXx	Máx. T _{máx} .	T _{máx} .	Valor mensual máximo de temperatura máxima diaria	°C
TNx	Máx. T _{mín} .	T _{mín} .	Valor mensual máximo de temperatura mínima diaria	°C
TXn	Mín. T _{máx} .	T _{máx} .	Valor mensual mínimo de temperatura máxima diaria	°C
TNn	Mín. T _{mín} .	T _{mín} .	Valor mensual mínimo de temperatura mínima diaria	°C
FD10*		T _{máx} .	Número de días en 1 año cuando la T _{mín} es <10 °C	Días
ID25*		T _{máx} .	Número de días en 1 año cuando la T _{máx} es <25 °C	Días
SU35*		T _{máx} .	Número de días en 1 año cuando la T _{máx} es <35 °C	Días

*Definidos para el presente estudio en casi todas las estaciones del INIA, exceptuando los climas de alta montaña (MARN, 2005), modificado de RCLimDex (1,0). Manual del usuario (Zhang y Yang, 2004).

RESULTADOS Y DISCUSIÓN

La implementación de un protocolo de calidad para las observaciones diarias de temperatura es de vital importancia, como lo confirman Wulfmeyer y Henning-Müller (2005), para detectar cambios y tendencias en las variables climáticas. No es suficiente con partir de una serie temporal de largo registro, sino contar con un apropiado nivel de calidad de ese conjunto de datos.

Con respecto a la proporción de datos faltantes, en los gráficos por estación se presentan dos ejemplos de la información obtenida:

- Cuando se habla de años perdidos, se hace referencia a que el valor anual es el resultado de promediar los valores mensuales, es decir, que si falta al menos 1 mes de registro de temperatura, según los criterios establecidos, no debería calcularse el promedio anual para el año en particular. Sin embargo, el promedio anual puede o no tener utilidad, dependiendo del tipo de estudio, por lo que se reporta pero no se descartan los valores mensuales en la base de datos.
- La serie diaria de T_{máx} 1950-2005 de la estación INIA-CENIAP ubicada en Maracay, estado Aragua, tiene menos 10% de datos faltantes para todos los meses (Figura 2a) lo que permitió que al generar los valores

mensuales con la condición $df > 10$ se conservaran 48 o más años de registros de las series mensuales de 51 años (Figura 2b). Esto indica que los datos de temperatura de esta estación posee una calidad adecuada, en términos de observaciones registradas.

- La serie diaria de T_{máx} 1980-2003 de la Estación Quibor (estado Lara) tiene de 60 a 70% de datos faltantes (Figura 3a). Al generar los valores mensuales con la condición $df > 10$ se conservaron entre 5 y 6 años de registros de las series mensuales de 22 años (Figura 3b).

Se realizó el mismo procedimiento para cada una de las estaciones y en el Cuadro 3 se presentan las frecuencias de datos observados y faltantes para T_{máx} y T_{mín}.

Se aclara que el promedio global de datos faltantes en el Cuadro 3 (25,9 y 26,6), no debe considerarse como un promedio simple de los porcentajes de datos faltantes, puesto que cada estación tiene una cantidad distinta de observaciones. Para este caso se calcula un promedio ponderado que considera el peso relativo, en cuanto a cantidad de información, de cada estación. Este promedio global podría tratarse como un indicador de calidad de la red de estaciones climatológicas de la Institución.

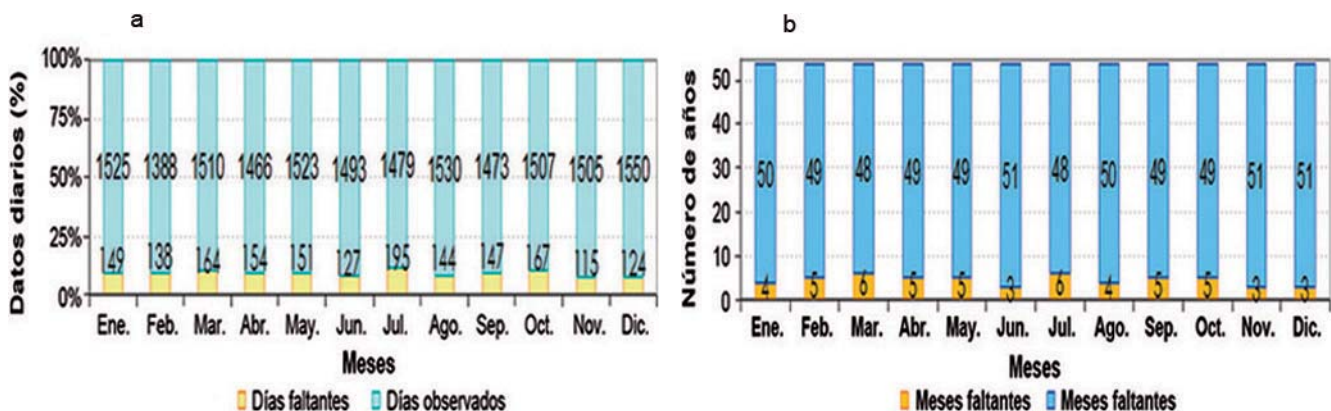


Figura 2. Número de valores faltantes y observados en temperatura máxima de la estación CENIAP-Maracay para el período 1950-2005. a. Observaciones diarias. b. Datos mensuales.

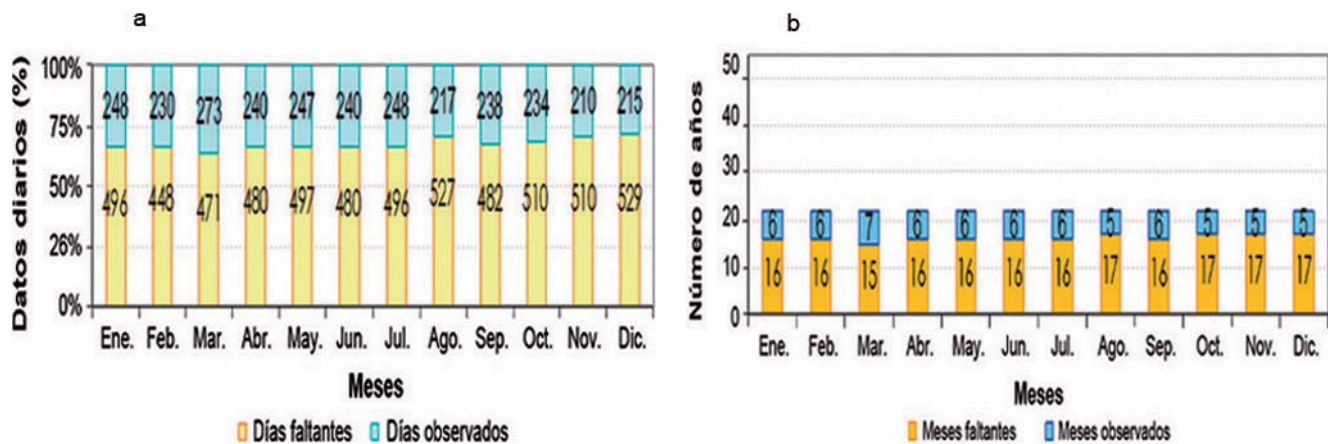


Figura 3. Número de valores faltantes y observados en temperatura máxima de la estación Quibor, estado Lara, para el período 1980-2003. a: Observaciones diarias. b: Datos mensuales.

El análisis de concentración y tendencia de las series temporales sirvió para la rápida detección del incremento o disminuciones en la media de la serie y los valores atípicos improbables. Tal es el caso de la serie 1951-2005 de la estación Turén, estado Portuguesa, donde se encontraron una $T_{máx}$ igual a $51,2\text{ }^{\circ}\text{C}$, $T_{mín}$ menores a los $10\text{ }^{\circ}\text{C}$ y OTD mayores a $25\text{ }^{\circ}\text{C}$ y menores a $0\text{ }^{\circ}\text{C}$ ($T_{mín} > T_{máx}$) (Figuras 4 a y b). Casos similares, considerados como valores atípicos de procedimiento por su ocurrencia improbable, fueron reportados y eliminados de la base de datos.

En el Cuadro 4 se muestra la frecuencia de observaciones improbables, que son valores con $T_{máx} > V$, $T_{mín} < U$ y $OTD \leq 0$, en el que V y U son umbrales que dependen del comportamiento de la variable para la región, donde se encuentra la estación. Se observa que el porcentaje total de casos improbables, rara vez excede 0,1% del total de observaciones de la serie.

Los valores fuera de rango podrían estar indicando errores humanos o problemas de calibración y mantenimiento de los equipos, por lo que si una estación en particular presenta un porcentaje elevado de casos improbables, deberían analizarse las posibles causas para realizar las correcciones pertinentes.

En el análisis de los índices climáticos se pudieron detectar casos como en la estación del

CENIAP-Maracay (Figuras 5 a y b) con valores muy elevados de FD10 (10 días en el año 1959 con $T_{mín}$ menor a $10\text{ }^{\circ}\text{C}$) y de SU35 (183 días en el mismo año con $T_{máx}$ mayor a $35\text{ }^{\circ}\text{C}$). Estos valores podrían ser el resultado de errores (mala calibración de los equipos o fallas en la lectura y en la transcripción) y estarse interpretando como eventos climáticos extremos de forma errónea, por lo que se recomienda en estos casos establecer procedimientos de control de calidad que sean realizados de forma periódica.

Se demostró que a través de la construcción de límites con media y desviación bponderadas robustas se pueden detectar series de datos dudosos por encima y por debajo del ciclo anual colectivo. La diferencia con otros métodos es que la ponderación asignada para cada observación depende de la distancia a la cual se encuentre de la mediana, es decir, que un valor extremo cuya distancia c sea $c \geq 7,5$ no tiene influencia sobre el cálculo de la desviación estándar (su ponderación es igual a cero).

Los límites permitieron detectar la causa del comportamiento atípico de los índices FD10 y SU35 en 1959 (Figuras 5 a y b) para la estación del Centro Nacional de Investigaciones Agropecuarias (Ceniap-Maracay), al revelar como atípica a la serie de temperatura mayo-noviembre/1959 con $T_{máx} > \text{límite superior}$ y $T_{mín} < \text{límite inferior}$ (Figuras 6a y 6b).

Cuadro 3. Frecuencia y porcentaje de datos faltantes en las series de temperatura máxima y mínima para las estaciones del Instituto Nacional de Investigaciones Agrícolas.

Estación	Serie	N° obs. de Tmáx en la serie	N° obs. de Tmín en la serie	N° obs. Tmáx faltantes	N° obs. Tmín faltantes	Tmáx de % faltantes	Tmín de % faltantes
El Tigre**							
CENIAP-Maracay	1950-2005	18 603	19 296	1 851	1 158	9,0	5,7
Ocumare de la Costa	1964-2004	13 257	13 251	1 718	1 724	11,5	11,5
Ciudad Bolivia	1988-2003	2 998	3 035	2 846	2 809	48,7	48,1
Banco San Pedro	1958-2005	14 158	13 749	3 374	3 783	19,2	21,6
El Cují	1964-2003	9 309	9 182	5 301	5 428	36,3	37,2
Quíbor	1980-2003	2 839	2 842	5 927	5 924	67,6	67,6
Mucuchíes	1961-2003	11 166	11 762	4 540	3 944	28,9	25,1
Tapipa-Padrón	1977-2004	5 822	5 320	4 405	4 907	43,1	48,0
Araure	1994-2005	3 145	2 528	1 238	1 855	28,2	42,3
Colonia Turén	1951-2005	15 580	15 470	4 509	4 619	22,4	23,0
Bramón	1941-2003	19 130	19 456	3 881	3 555	16,9	15,4
Pueblo Hondo	1983-2003	4 535	4 549	3 135	3 121	40,9	40,7
Yarituagua	1950-2004	16 318	15 092	3 771	4 997	18,8	24,9
El Guayabo	1974-1996	5 286	5 129	3 115	3 272	37,1	38,9
Total		142 146	140 661	49 610	51 095	25,9	26,6

Fuente: Inventario de estaciones climatológicas del INIA Venezuela. Cálculos propios. **No se incluye a la Estación El Tigre, para el momento del estudio no se pudo disponer del dato diario, únicamente de los mensuales.

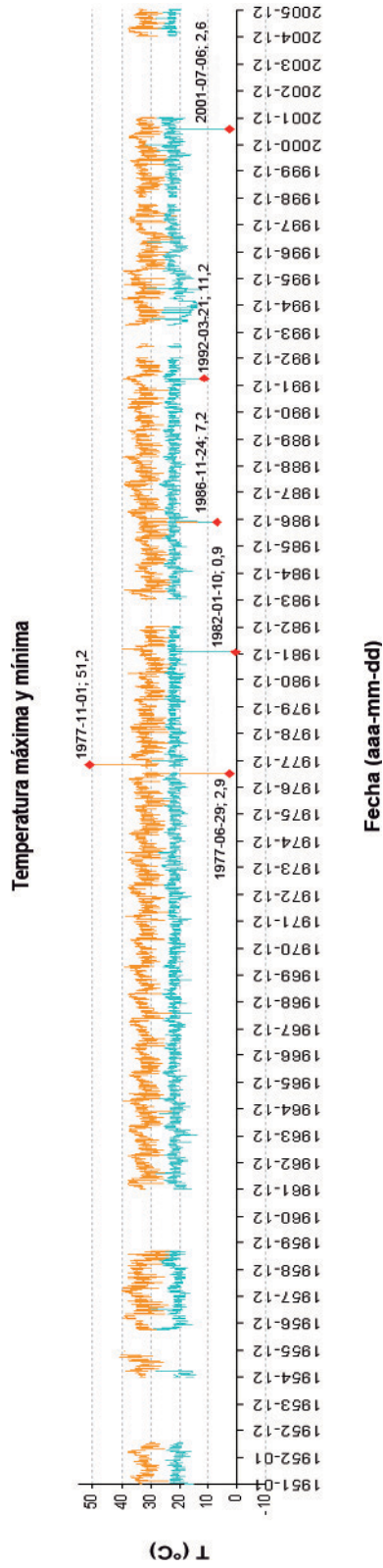


Figura 4a. Serie temporal de las Temperaturas máximas y mínimas para la estación Turén, estado Portuguesa, periodo 1951-2005.

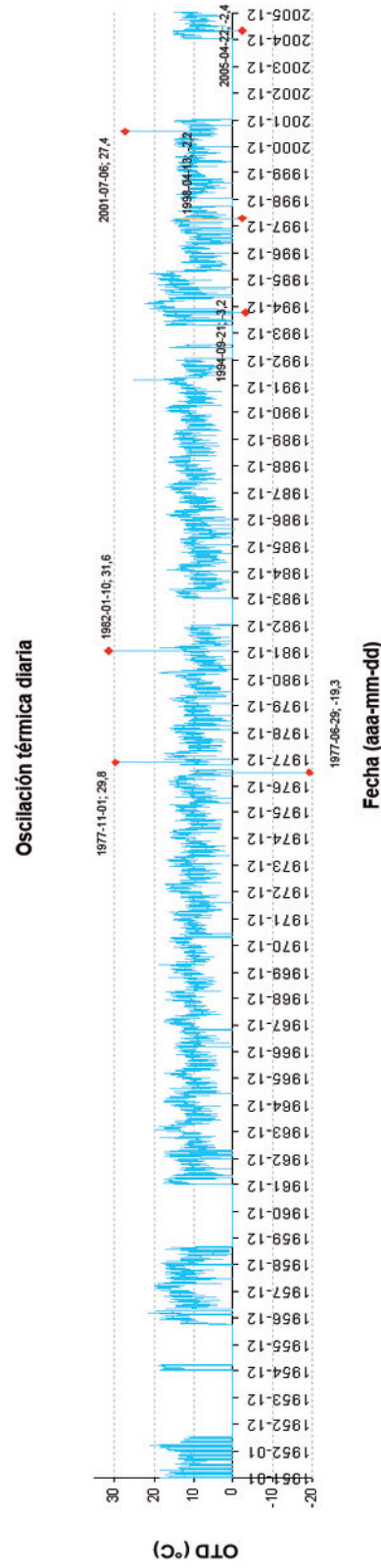


Figura 4b. Serie temporal de la oscilación térmica diaria (OTD) para la Estación Turén, estado Portuguesa, periodo 1951-2005.

Cuadro 4. Frecuencia y porcentaje de casos atípicos improbables para las estaciones del Instituto Nacional de Investigaciones Agrícolas.

Estación	Serie	N° obs. Tmáx en la serie	N° obs. Tmín en la serie	máx>V*	Tmín<U*	OTD≤0 °C	Total de casos improbables	Proporción de casos (%)
El Tigre**								
CENIAP-Maracay	1950-2005	18 603	19 296	1	37	2	40	0,11
Ocumare de la Costa	1964-2004	13 257	13 251	0	32	3	35	0,13
Ciudad Bolívar	1988-2003	2 998	3 035	0	0	5	5	0,08
Banco de San Pedro	1958-2005	14 158	13 749	0	1	6	7	0,03
El Cují	1964-2003	9 309	9 182	0	1	0	1	0,01
Quibor	1980-2003	2 839	2 842	0	8	0	8	0,14
Mucuchíes	1961-2003	11 166	11 762	0	0	1	1	0,00
Tapipa-Padrón	1977-2004	5 822	5 320	0	0	2	2	0,02
Araure	1994-2005	3 145	2 528	0	0	6	6	0,11
Colonia Turén	1951-2005	15 580	15 470	1	3	6	10	0,03
Bramón	1941-2003	19 130	19 456	2	0	37	39	0,10
Pueblo Hondo	1983-2003	4 535	4 549	1	0	0	1	0,01
Yaritagua	1950-2004	16 318	15 092	0	0	1	1	0,00
El Guayabo	1974-1996	5 286	5 129	1	1	2	4	0,04
Total		142 146	140 661	6	83	71	160	0,06

*Se utilizó V=50 °C y U=10 °C para todas las estaciones, a excepción de Mucuchíes, estado Mérida (V=30 °C y U=-10 °C); Bramón y Pueblo Hondo, estado Táchira (V=40 °C y U=0 °C). **No se disponía de registros diarios para el momento del análisis.

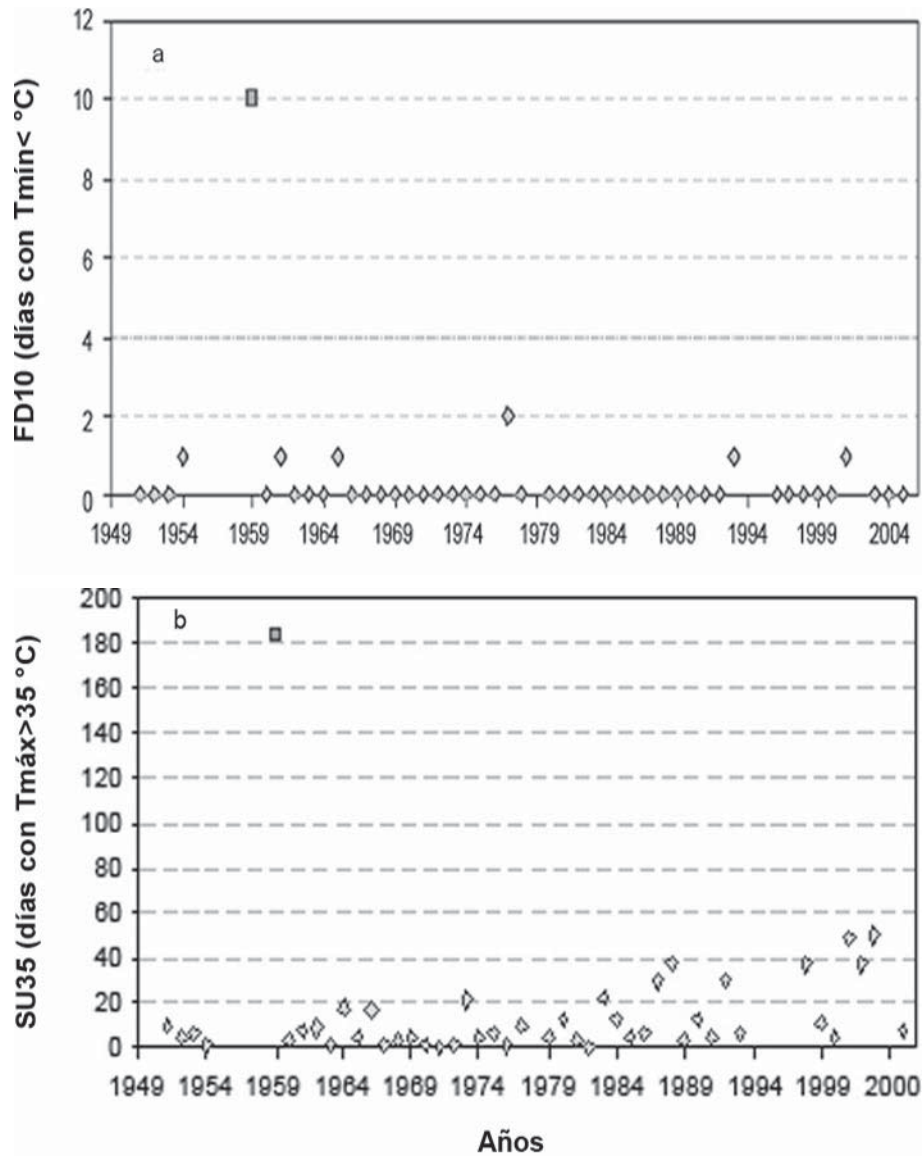


Figura 5. Comportamiento de dos índices climáticos para la estación del CENIAP-Maracay, estado Aragua, período 1950-2005. **a:** FD10 (frecuencia de días con $T_{mín} < 10$ °C); **b:** SU35 (frecuencia de días con $T_{máx} > 35$ °C).

En las Figuras 6 a y b, los colores representan: líneas sólidas = límites bponderados de detección de valores atípicos ($media \pm 4$ desviaciones); puntos verde claro = valores de temperaturas dentro de los límites; rombos azules = valores de temperatura fuera de rango (reportados como dudosos); puntos verde oscuro = valores medios de temperatura diaria. La media y desviación se calcularon según Lanzante (1996).

En todas las estaciones se consiguieron valores atípicos fuera de los límites calculados ($4s_{Br}$) no excediendo en ningún caso 2,7% de las observaciones (Cuadro 5). Los valores determinados como atípicos con este análisis fueron reportados como dudosos en la base de datos depurada.

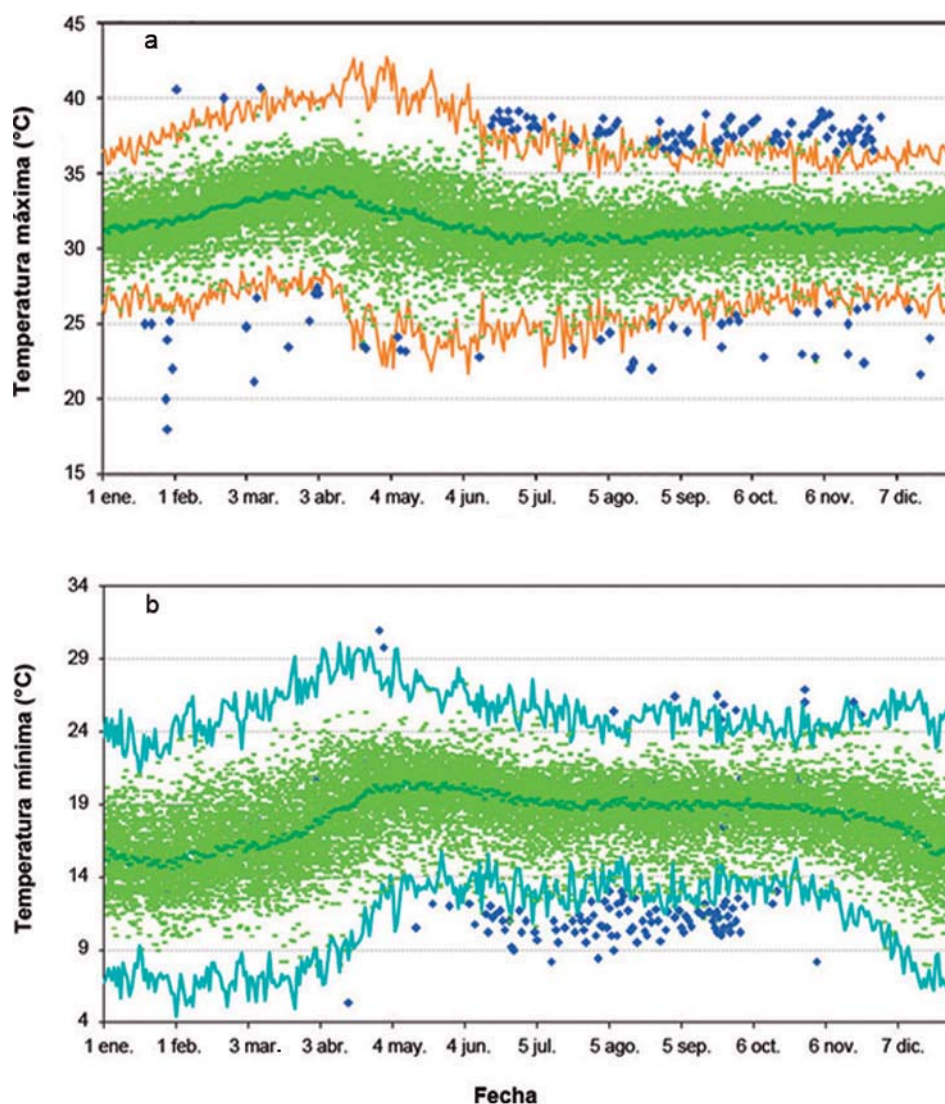


Figura 6. Temperatura máxima diaria (a); temperatura mínima diaria (b).
Para la estación del CENIAP-Maracay, 1950-2005.

CONCLUSIONES

En este trabajo se aplicó un protocolo de control de calidad y análisis de las series de $T_{máx}$ y $T_{mín}$, para 15 de las estaciones del INIA, lográndose la identificación de patrones y tendencias, la eliminación de datos dudosos improbables, y el reporte de datos dudosos cuestionables con base estadística robusta (estadísticos bponderados), por lo que se considera como la metodología más ajustada para la variable analizada.

El procedimiento utilizado permitió depurar la base de datos a través de la eliminación de aquellos valores de ocurrencia improbable por medio de la utilización de filtros lógicos, de acuerdo a las condiciones de cada estación ($T_{máx} > 50\text{ °C}$ y $T_{máx} < T_{mín}$). El cálculo de los índices climáticos y estadísticos bponderados permitió el análisis más exhaustivo de esta base depurada, tal fue el caso de la estación CENIAP-Maracay, donde se encontraron valores muy elevados de FD10 (10 días en 1959 con $T_{mín} < 10\text{ °C}$) y de SU35 (183 días en 1959 con $T_{máx} > 35\text{ °C}$).

Cuadro 5. Frecuencia de casos atípicos fuera de los límites $4s_{Bi}$, para las temperaturas máxima y mínima ($T_{máx} - T_{mín}$).

Estación	Serie	T _{máx}			T _{mín}		
		Nº obs. en la serie	Nº atípicos	% atípicos ($4s_{Bi}$)	Nº obs. en la serie	Nº atípicos ($4s_{Bi}$)	% atípicos
El Tigre**							
CENIAP-Maracay	1950-2005	18 603	213	1,14	19 296	193	1,00
Ocumare de la Costa	1964-2004	13 257	44	0,33	13 251	35	0,26
Ciudad Bolívar	1988-2003	2 998	53	1,77	3 035	16	0,53
Banco San Pedro	1958-2005	14 158	132	0,93	13 749	352	2,56
El Cují	1964-2003	9 309	40	0,43	9 182	43	0,47
Quibor	1980-2003	2 839	76	2,68	2 842	39	1,37
Mucuchíes	1961-2003	11 166	13	0,12	11 762	22	0,19
Tapipa-Padrón	1977-2004	5 822	44	0,76	5 320	110	2,07
Araure	1994-2005	3 145	75	2,38	2 528	46	1,82
Colonia Turén	1951-2005	15 580	105	0,67	15 470	237	1,53
Bramón	1941-2003	19 130	37	0,19	19 456	179	0,92
Pueblo Hondo	1983-2003	4 535	28	0,62	4 549	17	0,37
Yarituagua	1950-2004	16 318	173	1,06	15 092	114	0,76
El Guayabo	1974-1996	5 286	54	1,02	5 129	53	1,03
Total		142 146	1 087	0,76	140 661	1 456	1,04

*Se utilizó $V=50$ °C y $U=10$ °C para todas las estaciones, a excepción de Mucuchíes ($V=30$ °C y $U=-10$ °C), Bramón y Pueblo Hondo ($V=40$ °C y $U=0$ °C).

**No posee registros diarios.

Estos resultados pudieron haber sido ocasionados por fallas mecánicas o humanas reflejadas en los datos y se corre el riesgo de que sean interpretados como eventos climáticos extremos de forma errónea, pero no deberían eliminarse de forma definitiva de la base de datos. La construcción de límites con media y desviación bponderadas robustas permitieron detectar series de datos dudosos, como el caso (en la estación antes mencionada) de la serie de temperatura mayo-noviembre/1959, con $T_{\text{máx}} > \text{límite superior}$ y $T_{\text{mín}} < \text{límite inferior}$. Los valores cuestionados fueron reportados en la base de datos depurada.

Por otro lado, debe tomarse en cuenta, al momento de realizar cualquier análisis o procedimiento estadístico con datos climáticos, que la longitud de la serie debe determinarse más por el número de observaciones que por el número de años de la misma. Esto demuestra que series de 10 años o más (lo que se considera adecuado para la variable temperatura) reportaron más 50% de datos faltantes.

Se concluye que el aseguramiento de la calidad de una red y las comparaciones periódicas resultan ser aspectos fundamentales que deben ser tomados en cuenta para minimizar la aparición de datos atípicos derivados de particularidades instrumentales. Es importante contar con una base de meta-datos que permita una rápida verificación de las condiciones bajo las cuales el dato meteorológico es generado. Muchas de las decisiones que se puedan tomar sobre cómo tratar o catalogar datos atípicos dependerán de qué tan completa y disponible se encuentre la meta-información.

Finalmente, la metodología propuesta es sencilla y tanto los métodos como sus resultados pueden ser aplicados a un gran rango de situaciones para cualquier variable meteorológica.

LITERATURA CITADA

- Edwards, D. 1998. Data Quality Control/ Quality Assurance" In: Data and Information Management in the Ecological Sciences: A Resource Guide (W.K. Michener, J.H. Porter, and S.G. Stafford, Eds.), University of New Mexico, Albuquerque. pp. 33-40.
- Feng, S., Q. Hu. and W. Qian. 2004. Quality control of daily meteorological data in China, 1951-2000: A new dataset. *Int. J. Climatol.* 24(1):853-870.
- Lanzante, J. 1996. Resistant, robust, and nonparametric techniques for the analysis of climate data. Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.* 16(3):1.197-1.226.
- Lobo, D., D. Gabrield, F. Ovalles, F. Santibañez, M.C. Moyano, R. Aguilera, R. Pizarro, C. Sanguesa y U. Nelson. 2005. Guía metodológica para la elaboración del mapa de zonas áridas, semiáridas y subhúmedas secas de América Latina y el Caribe. CAZALAC, Chile. 59 p.
- Ministerio del Ambiente y Recursos Naturales (MARN). 2005. Primera Comunicación Nacional en Cambio Climático de Venezuela. MARN, Programa de las Naciones Unidas para el Desarrollo, Fondo Mundial para el Medio Ambiente. Venezuela. 164 p.
- National Climatic Data Center. 2002. Data documentation for Data Set 9300 (DSI-9300) Global Historical Climatology Network – Daily v. 1.0. Ashville, USA. 17 p.
- Rodríguez, M. F., A. Cortez, J. C. Rey, M. C. Núñez, F. Ovalles y R. Parra. 2011. Integración espacial y aplicación de los datos agroecológicos del INIA al Norte del Orinoco de Venezuela. Proyecto financiado por el FONACIT S1-20022000417. Instituto Nacional de Investigaciones Agrícolas (INIA.) 371 p.
- Shaw, E. 1988. Hydrology in Practice. 2a edición. Chapman & Hall, Londres-UK. 539 p.
- Vincent, L. A., T. C. Peterson, V. R. Barros, M. B. Marino, M. Rusticucci, G. Carrasco, E. Ramírez, L. M. Alves, T. Amvrizzi, M. A. Berlato, A. M. Grimm, J. A. Marengo, L. Molion, D. F. Moncunill, E. Rebello, Y. M. T. Anunciação, J. Quintana, J. L. Santos, J. Báez, G. Coronel, J. García, I. Trebejo, M. Bidegain, M. R. Haylock and D. Karoly. 2005. Observed Trends in Indices of Daily Temperature Extremes in South America 1960-2000. *Journal of Climate.* 18(23):5.011-5.023.

Wulfmeyer, V. and I. Henning-Müller. 2005. The climate station of the University of Hohenheim: Analyses of air temperature and precipitation time series since 1878. *Int. J. Climatol.* 26(1):113-138.

Zhang, X. and F. Yang. 2004. RclimDex (1.0) Manual del usuario. Climate Research Branch Environment Canada. Downsview, Ontario (Canadá). 22 p.