

Aplicación de procedimientos estadísticos para el control de calidad de las series de precipitación mensual en los llanos orientales venezolanos

Application of statistics procedures for the quality control of the monthly rainfall series in the Venezuelan Oriental Plains

B. Olivares¹, A. Cortez², R. Parra³, M.F. Rodríguez² y E. Guevara¹

¹Departamento de Agrometeorología. Instituto Nacional de Investigaciones Agrícolas del estado Anzoátegui.

²Investigadores. INIA. Centro Nacional de Investigaciones Agropecuarias (CENIAP). Laboratorio de sistemas de Información en Recursos Agroecológicos.

³Facultad de Agronomía. Universidad Central de Venezuela.

Resumen

La disponibilidad de datos confiables de precipitación representa un requisito indispensable para cualquier tipo de análisis hidrológico o agrometeorológico. Básicamente el error, presencia o ausencia de un dato en particular, puede tener efectos sobre las estimaciones de probabilidad con altos periodos de retorno de los eventos extremos, dificultades en el calendario agrícola y repercusiones en planificación y toma de decisiones en el negocio agrícola. El objetivo de este trabajo además de calcular la estadística descriptiva básica, es aplicar los procedimientos estadísticos para la detección de datos faltantes mensuales, la detección de valores extremos de lluvia mensual, la verificación de la homogeneidad de las series y el análisis de dispersión temporal a las series de precipitación mensual. Se utilizaron las series de precipitación mensual de 34 estaciones pertenecientes de los estados Anzoátegui, Bolívar, Guárico, Monagas y Sucre para el periodo de 1970-2000. De acuerdo al control de calidad de datos en las series de Aragua de Barcelona, Bergantín, Ciudad Bolívar, Cumana, El Tigre, Gurí, Maturín y Valle La Pascua, no se presentan valores atípicos que pudieran considerarse como dudosos, así mismo los registros no presentaron problemas de oscilación o falta de homogeneidad. Las series de precipitación con una gran cantidad de datos

Recibido el 16-3-2012 ● Aceptado el 21-6-2013
Autor de correspondencia e-mail: barlinolivares@gmail.com

faltantes tales como: Cachipo, Campo Mata, Cantaura, Urica y Viento Fresco ven seriamente afectada la representatividad de su información. La metodología propuesta para el control de calidad de datos es sencilla y puede ser aplicada a diferentes series de datos en otras regiones de la zona intertropical.

Palabras clave: control de calidad de series de precipitación, datos faltantes en registros de precipitación.

Abstract

The availability of reliable data of precipitation is a prerequisite for any hydrological or agrometeorological analysis. Basically the error, presence or absence of a particular data, may have effects on the probability estimates with high return periods of extreme events, difficulties in the agricultural calendar and implications for planning and decision-making in the agricultural business. The objective of this paper is to apply statistical procedures: 1) detection of missing data per month, 2) detection of extreme values monthly rainfall, 3) calculation of basic descriptive statistics, 4) verification of the homogeneity of the series and 5) temporal dispersion analysis of monthly precipitation series. Were used monthly precipitation series of 34 stations belonging to the states of Anzoategui, Bolivar, Guarico, Monagas and Sucre for the period 1970-2000. According to the quality control data series Aragua de Barcelona, Bergantín, Ciudad Bolívar, Cumana, El Tigre, Gurí, Maturín and Valle La Pascua, there are no outliers that could be considered as doubtful, likewise no records presented problems of oscillation or lack of homogeneity. The series of precipitation with a large amount of missing data such as: Cachipo, Campo Mata, Cantaura, Urica and Viento Fresco are seriously affected the representativeness of their information. The proposed methodology for data quality control is simple and can be applied to different data sets in other regions of the tropics.

Key words: quality control of precipitation series, missing data on precipitation records.

Introducción

La información climática es utilizada por productores, investigadores y profesionales del área agrícola para optimizar la toma de decisiones y evaluar el impacto del clima sobre diferentes aspectos de la producción agropecuaria. En este sentido, se requiere de caracterizaciones climáticas en diferentes escalas tales como anual, mensual y diaria que proporcionen una

Introduction

The weather information is used for producers, researchers and professionals of the agriculture area to optimize the decision-making process and evaluate the weather impact on different aspects of the livestock production. On this matter, different scale weather characteristics are required, such as, annual, monthly and daily, that would provide an

aproximación a las condiciones del ambiente físico de la región (Millano *et al.*, 2007; Cortéz *et al.*, 2005).

El estudio del régimen de precipitación constituye un aspecto primordial debido a que la lluvia determina el crecimiento de los cultivos y la ejecución o desarrollo de las actividades agrícolas en campo; así mismo los cambios climáticos en los últimos años requieren actualizar los registros de datos incorporando la información más reciente (Guenni *et al.*, 2008).

En Venezuela, la red de estaciones meteorológicas está integrada por: el Ministerio Popular para el Ambiente (MPPA), Ministerio del Poder Popular para la Defensa, Instituto Nacional de Investigaciones Agrícolas (INIA), Electrificación Del Caroní (EDELCA) y Universidades Nacionales.

Los diferentes trabajos vinculados al uso de datos climáticos requieren estrictamente de un procedimiento de análisis e integración de la información representada por la obtención de datos de manera adecuada, especialmente cuando se utiliza la variable precipitación debido a la alta variabilidad que ésta presenta (Parra y Cortéz, 2005; UNESCO, 2006).

La variabilidad temporal de la precipitación influye en la gestión y manejo de los recursos hídricos, la prevención de inundaciones y sequías, la planificación y operación de actividades agrícolas, la generación hidroeléctrica y el abastecimiento de agua a la población humana (Ablan *et al.*, 2008). La precipitación es probablemente uno de los parámetros climáticos más importantes; los diferentes usuarios de los servicios de meteorología e hidrología se interesan por obtener datos de llu-

approximation to the physical environment conditions of the region (Millano *et al.*, 2007; Cortéz *et al.*, 2005).

The investigation about the rainfall regime constitutes a main aspect, since rain determines the growth of the crops and the execution or development of the agriculture activities in the field; likewise, the weather changes in the last years require upgrading the data registers incorporating the most recent information (Guenni *et al.*, 2008).

In Venezuela, the meteorological station red is integrated by: The Environmental Popular Ministry (MPPA), The Defense Popular Ministry, National Institute of Agriculture Researches (INIA), Electrification of the Caroní (EDELCA) and National Universities.

The different researches linked to the use of weather data strictly require an adequate analysis and integration procedure of the information represented by the obtaining of the data, especially when the rainfall variable is used, due to the high variability that it has (Parra and Cortéz, 2005; UNESCO, 2006).

The temporal variability of the rainfall influences in the management and handling of the water resources, the prevention of floods and drought, planning and operation of agriculture activities, hydroelectric generation and the water provision of the human population (Ablan *et al.*, 2008). Rainfall is probably one of the most important weather parameter; the different users of the meteorological and water services are interested in obtaining daily, monthly, and annual

via diaria, mensual, anual, precipitación media, número de días lluviosos, número de días secos y precipitaciones extremas, con la finalidad de conocer la variabilidad de este elemento climático.

Básicamente cuando se establece el término de calidad de los datos se refiere a la representatividad de ese dato respecto a la magnitud real que se mide, a su valor en ese momento y además a la capacidad del dato climático de representar la condición meteorológica de la zona. La validación del dato puede realizarse a diferentes niveles con diferentes algoritmos o procedimientos en función de la disponibilidad del dato y de la exigencia del usuario final (WMO, 1989; Reilly y Kroll, 2003).

Las bases de datos meteorológicos de estaciones convencionales se caracterizan por presentar errores representados principalmente por varios aspectos; en primer lugar, por la ubicación de la estación y en segundo plano por la calibración, lectura y mantenimiento de los instrumentos de medición (Jones, 1987; Parra y Cortéz, 2005). En este sentido, es sumamente importante recurrir a pruebas de control de calidad de datos para disminuir considerablemente las inexactitudes, permitiendo verificar la amplitud de variación y consistencia de los datos climáticos.

Dentro de este contexto, Parra y Cortéz (2005) aplicaron el control de calidad a las series de precipitación diaria de las estaciones del INIA, mediante la determinación de datos faltantes, el cálculo de estadística descriptiva, análisis de concentración y tendencia de series temporales y aná-

rainfall data, mean precipitation, number of rainy days, number of dry days and extreme precipitations, with the aim of knowing this climatic element.

Basically, when the quality term of the data is established, it refers to the representation of this data regarding the real magnitude measured, its value in the moment and the weather data capacity to be represented in the meteorological levels of the area. The validation of the data can be done at different levels and with different algorithms or procedures in function of the data availability and the exigency of the final user (WMO, 1989; Reilly and Kroll, 2003).

The data of the conventional meteorological stations are characterized by presenting errors mainly represented by different aspects; firstly, the location of the station, secondly, the calibration, reading and maintenance of the measure instruments (Jones, 1987; Parra and Cortéz, 2005). In this matter, it is very important to appeal to data quality control tests in order to reduce the inaccuracies, allowing verifying the width of the variation and consistency of the climatic weather.

In this sense, Parra and Cortéz (2005) applied the quality control of the daily precipitation series of the INIA station, determining the remaining data, the descriptive statistical calculus, concentration analysis and the tendency of temporal series and dispersion analysis, concluding that the studied series did not require the implementation of adjustment and correction measures.

Ablan *et al.*, (2008) presented a methodological proposal to analyze the

lisis de dispersión, concluyendo que las series estudiadas no requerían la implementación de medidas de ajuste y corrección.

En este orden de ideas, Ablan *et al.* (2008) presentaron una propuesta metodológica para analizar datos climáticos históricos a ser utilizados en estudios de la variabilidad temporal de la precipitación en una región, determinando ciertos problemas de homogeneidad y consistencia de datos en las estaciones estudiadas. Así mismo, la UNESCO-ROSTLAC (1982) implantó el control de calidad de series de precipitación de América del Sur, estableciendo diferentes técnicas estadísticas para obtener un análisis gráfico y exhaustivo de la variable estudiada.

El objetivo principal de este trabajo es aplicar el control de calidad a las series mensuales de precipitación de las estaciones del estado Anzoátegui, Bolívar, Guárico, Monagas y Sucre mediante la detección de datos faltantes y valores extremos, cálculo de la estadística descriptiva básica, verificación de la homogeneidad de las series y análisis de dispersión temporal a las series temporales para el periodo de 1970-2000.

Materiales y métodos

Selección de las estaciones

Para la selección de las estaciones utilizadas en el estudio (cuadro 1), se formuló un inventario de todas las estaciones pluviométricas en la zona de estudio comprendida por los estados Anzoátegui, Bolívar, Guárico, Monagas, y Sucre; pertenecientes al INIA, MPPA, al Servicio de Meteorología

historical weather data to be used in the temporal variability researches of the precipitation in a region, determining some homogeneity problems and data consistency in the studied stations. Likewise, UNESCO-ROSTLAC (1982) applied the precipitation quality control series of South America, establishing different statistical techniques to obtain a graphic and exhaustive analysis of the studied variable.

The main objective of this research is to apply the quality control to the monthly precipitation series of the stations in Anzoátegui, Bolívar, Guárico, Monagas and Sucre, detecting the missing data and extreme values, calculus of the basic descriptive statistic, verifying the homogeneity of the series and the temporal dispersion analysis of the temporal series for the period of 1970-2000.

Materials and methods

Selection of the stations

An inventory of the rainy stations was formulated for selecting the stations used in this research (table 1) in the study areas of Anzoátegui, Bolívar, Guárico, Monagas and Sucre; belonging to INIA, MPPA, the Meteorological Service of FAV, EDELCA and Universidad Central de Venezuela (UCV). 99 stations were considered, with monthly registers corresponding 1970-2000. The information of the basic meta data was established, represented by:

National serial: code assigned by the responsible organism of the meteorological station

Cuadro 1. Listado de estaciones climatológicas seleccionadas.

Table 1. List of the selected climatic stations.

N	Estación	Estado	Serial	Tipo	Altura msnm	Latitude dec.	Longitude dec.
1	Anaco	AN	2757	PR	220	9.4833	-64.4833
2	Aragua de Barcelona	AN	1755	PR	110	9.4652	-64.8166
3	Barcelona	AN	80419	SB	7	10.1116	-64.6947
4	Bergantín	AN	1796	PR	310	10.0175	-64.3894
5	Cachipo	MO	2816	PR	30	9.9333	-63.0611
6	Campo Mata	AN	2701	PR	135	9.2022	-64.03472
7	Cantaura	AN	2766	PR	250	9.3097	-64.3644
8	Cariaco-Muelle	SU	1829	PR	10	10.5000	-63.6666
9	Ciudad Bolívar	BO	3882	A1	44	8.0669	-63.5508
10	Clarines	AN	2608	PR	12	9.9638	-65.16388
11	Cumana	SU	1741	SB	2	10.4502	-64.1316
12	El Chaparro Caserío	AN	1756	PR	120	10.2275	-64.5300
13	El Guamo	MO	1899	C3	435	10.0805	-63.6527
14	El Tigre	AN	3715	C1	302	8.8627	-64.2213
15	Guanape	AN	2605	PR	130	9.9194	-65.5019
16	Guaribe-Tenepe	AN	2647	PR	80	9.7613	-65.4855
17	Gurí	BO	8	C1	276	7.9969	-62.3716
18	Hacienda la Concordia	SU	1716	PR	2	10.4291	-63.4833
19	La Cerca	AN	1683	PR	4	10.1030	-65.2136
20	Los Yopales	AN	3719	PR	260	8.7455	-64.4363
21	Maturín	MO	2827	SB	66	9.7538	-63.2097
22	Onoto	AN	2648	PR	35	9.6016	-65.1969
23	Pariaguán	AN	3712	PR	240	8.8452	-64.7163

Cuadro 1. Listado de estaciones climatológicas seleccionadas (Continuación).

Table 1. List of the selected climatic stations (Continuation).

N	Estación	Estado	Serial	Tipo	Altura msnm	Latitud dec.	Longitud dec.
24	Puerto la Cruz	AN	1751	PR	15	10.2172	-64.6302
25	Querecual	AN	2615	PR	80	9.9500	-64.4833
26	San Diego Cabrutica	AN	3762	PR	140	8.4208	-64.8861
27	San Mateo	AN	2725	PR	132	9.7436	-64.5333
28	Santa Clara II	AN	3720	PR	60	8.4944	-64.6088
29	Santa Rosa	AN	2755	PR	280	9.5811	-64.2925
30	Upata	BO	3994	C2	327	8.0166	-62.3833
31	Úrica	AN	2730	PR	240	9.7172	-64.0055
32	Valle de la Pascua	GU	2589	C1	184	9.2166	-66.0000
33	Viento Fresco	MO	2802	PR	240	9.7736	-63.6694
34	Zuata	AN	2622	PR	58	8.3611	-65.1930

logía de la FAV, EDELCA y Universidad Central de Venezuela (UCV). Se consideraron 99 estaciones con registros mensuales del periodo 1970-2000. Se estableció la información de la metadata básica representada por:

Serial nacional: código asignado por el organismo responsable de la estación meteorológica.

Estación: nombre de la estación.

Ubicación de la estación: Estado, municipio y sitio

Tipo de estación: (C1: mide todos los elementos, C2: mide todos los elementos excepto dirección y velocidad del viento, C3: mide solo temperatura del aire y precipitación, PR: registra solo precipitación, A1: estación sinóptica, realiza observaciones de los principales elementos meteorológicos en horas establecidas internacionalmente, SB: Estación Sinóptica Básica (a la vez Estación Climatológica Principal C1).

Cota: altitud en msnm

Latitud y Longitud: Ubicación en grados, minutos y segundos para obtener la georeferenciación (figura 1).

Variables climáticas medidas

Fecha de inicio y fecha de culminación

El período de datos climáticos disponibles.

Después de estructurar la metadata con la información básica de las estaciones, solo se seleccionaron aquellas estaciones climatológicas con registros largos de precipitación. De acuerdo a la Organización Meteorológica Mundial (WMO, 1989), se requiere un mínimo de 40 años de datos para estudios de variabilidad de la precipitación en regiones tropicales planas continentales tales como los Llanos

Station: name of the station

Location of the station: state, county and place

Type of station: (C1: measures all the elements, C2: measures all the elements except direction and wind velocity, C3: only measures the air temperature and precipitation, PR: only registers precipitation, A1: synoptic station, carries out observations of the main meteorological elements in the hours established internationally, SB: basic synoptic station (at the same time main climatic station C1).

Height: altitude represented in masl

Latitude and longitude: location in degrees, minutes and seconds, to obtain the geo-reference (figure 1)

Climatic measure variables

Starting date and end date

The period with the available climatic data

After structuring the data with the basic information of the stations, were only selected those climatic stations with long precipitation registers. According to the World Meteorological Organization (WMO, 1989), a minimum of 40 years of data is required, for precipitation variability researches in flat continental tropical regions, such as Venezuelan Oriental Plains. Only 34 stations were selected, due to 25% of the total of the stations presented continuous and extended periods of missing data inside the historical register.

Detection of missing data

To determine the proportion of the monthly missing data in the precipitation series, was considered the

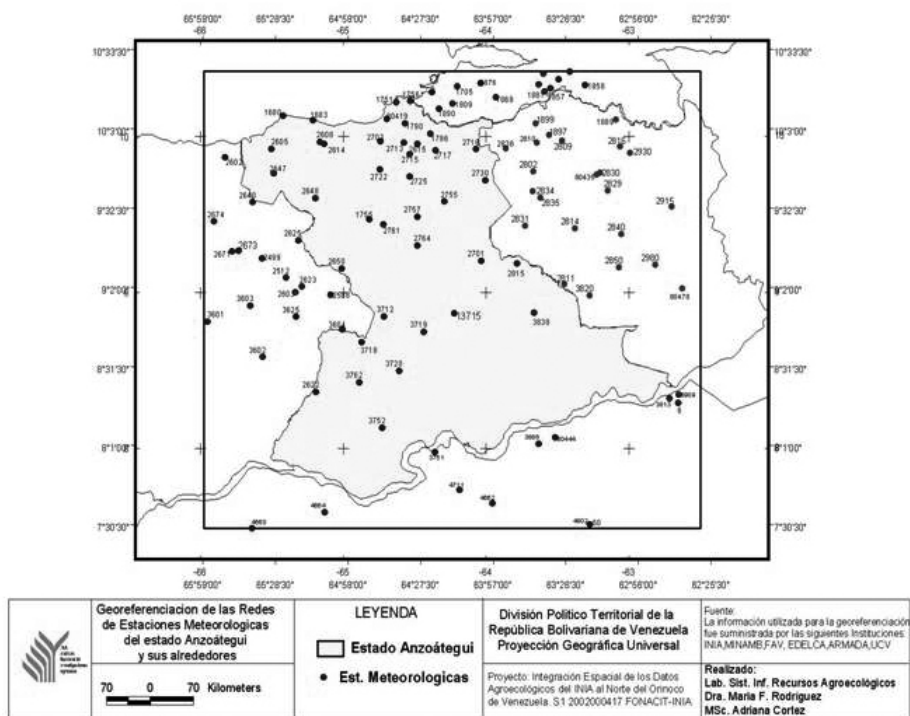


Figura 1. Distribución geográfica de la red de estaciones meteorológicas en el estado Anzoátegui, Bolívar, Guárico, Monagas y Sucre.

Figure 1. Geographic distribution of the weather stations red in states Anzoátegui, Bolívar, Guárico, Monagas and Sucre.

orientales venezolanos. Solo se seleccionaron 34 estaciones debido a que el 25% del total de las estaciones presentaban periodos ininterrumpidos y extensos de datos faltantes dentro del registro histórico.

Detección de datos faltantes

Para determinar la proporción de datos faltantes mensuales en las series de precipitación se consideró el criterio utilizado por Parra y Cortez (2005); el cual establece que si existe por lo menos una observación faltante (df) en el mes, no se calcula el valor mensual (se considera perdido), debi-

criteria used by Parra and Cortez (2005); which establishes that if there is, at least one missing observation (df) in the month, the monthly value is not calculated (is considered lost), since the monthly value is calculated as the sum of the daily values or total rainy amount in the month. The missing data in these series can affect the representation of the data used in the research, at the same time; it would allow the selection of stations with an adequate historical register longitude for different types of climatic and water analyses.

do a que el valor mensual se calcula como la sumatoria de los valores diarios o lluvia total caída en el mes. Los datos faltantes en estas series pueden afectar la representatividad de los datos utilizados en el estudio, a su vez permitiría la selección de estaciones con una longitud de registro histórico adecuado para diferentes tipos de análisis climático e hidrológico.

Detección de valores extremos

La teoría de valores extremos permite describir el comportamiento de los puntos muy alejados de la media en una distribución; en este sentido resulta de mucha utilidad para modelar eventos extremos de precipitación por el impacto que estos pueden tener en el cambio climático global afectando de diversas maneras la integridad humana, ambiental y económica (Wulfmeyer y Henning-Müller, 2005; Javelle *et al.*, 2003).

Para detectar los valores extremos mensuales en las series de precipitación se usó el criterio basado en la desviación absoluta de la mediana (*mad*) de acuerdo con Crawley (2002) y Ablan *et al.* (2008). La *mad* representa un estimador robusto de la variabilidad de los datos considerado menos sensible a los valores extremos. En este sentido, cuando el valor de la desviación estándar es mayor que cuatro veces el valor de *mad*, se considera que el conjunto de datos posee valores extremos. Después de haber identificado los posibles valores extremos, para cada estación se grafican los datos mensuales con su respectiva media y desviación estándar.

Estadística descriptiva básica

Para cada estación seleccionada, se calcularon los parámetros estadís-

Detection of extreme values

The theory of extreme values allows describing the behavior of the numbers that were far away of the mean in a distribution; in this sense, it is very useful to model extreme precipitation events by the impact that these might have in the global climatic change, affecting differently the human, environmental and economical integrity (Wulfmeyer and Henning-Müller, 2005; Javelle *et al.*, 2003).

The criteria based in the absolute deviation of the media (*mad*) according to Crawley (2002) and Ablan *et al.*, (2008), was used for detecting the monthly extreme values in the precipitation series. *Mad* represented the consistent estimator of the data variability, considered less sensitive to the extreme values. In this sense, when the value of the standard deviation is four times higher than the *mad* value, is considered that the data has extreme values. After indentified the possible extreme values, the monthly data is showed in a graphic, with the correspondent mean and standard deviation.

Basic descriptive statistics

For each of the selected station, were measured the descriptive statistical parameters; monthly media precipitation, monthly maxima, standard deviation, variation coefficient, percentiles 25%, 50%, 75%, asymmetry, kurtosis and absolute deviation of the mean.

Homogeneity verification of the series

To verify the homogeneity of the studied stations was carried out the homogeneity test and random trait, to

tivos descriptivos: precipitación media mensual, máxima mensual, desviación estándar, coeficiente de variación, percentiles 25%, 50%, 75%, la asimetría, kurtosis y desviación absoluta de la mediana.

Verificación de la homogeneidad de las series

Para verificar la homogeneidad de las estaciones estudiadas se realizó la prueba de homogeneidad y carácter aleatorio para reconocer patrones de oscilación de Wald-Wolfowitz (WMO, 1989; Sneyers, 1990). Esta prueba de homogeneidad establece que la hipótesis nula consiste en asumir que todas las observaciones de la serie provienen de la misma población y que son independientes, o lo que es equivalente a que la serie se comporta de forma aleatoria.

Este test contrasta si dos muestras con datos independientes proceden de poblaciones de la misma distribución. Si esto es así lógicamente los parámetros poblacionales de ambas muestras son los mismos basándose en la prueba de rachas, la cual consiste en ordenar todos los casos de ambos grupos de forma conjunta, dispuestos en orden, se cuentan las rachas pertenecientes al mismo grupo para determinar la homogeneidad de las series (Arnell, 2003; Griffis *et al.*, 2004).

En este sentido aquellas estaciones que resulten homogéneas, pueden ser consideradas en un análisis de variabilidad climática, debido a que todas las observaciones de la serie provienen de la misma población (Yue y Pilon, 2004). La no homogeneidad de las series, puede ser atribuida directa o indirectamente al personal que efectúa la medición en las estaciones o en

recognize the Wald-Wolfowitz oscillation patterns (WMO, 1989; Sneyers, 1990). This homogeneity test establishes that the null hypothesis consists on assuming that all the observations of the series come from the same population, and are independent, which means that the series behaves randomly.

This test contrasts if two samples with independent data come from populations of the same distribution. If this is supported, the population parameters of both samples are logically going to be the same, based on the run test, which consists on ordering all the cases of both groups in order, the runs belonging to the same group are counted, to determine the homogeneity of the series (Arnell, 2003; Griffis *et al.*, 2004).

In this sense, those stations which result to be homogeneous, are considered in a climatic variability analysis, due to all the observations of the series come from the same population (Yue and Pilon, 2004). The lack of homogeneity of the series can be attributed directly or indirectly to the personnel that carry out the measures in the stations or in the operation of these places; as well as to the localization changes of the station or local derivations proper of the region.

Dispersion analysis of the temporal series

Parra and Cortéz (2005) establish the inter-annual variability with a temporal dispersion analysis of the monthly precipitation series. This dispersion analysis of the temporal series was done with the graphic representation of "Boxplots",

las operaciones de las mismas, así como también, a los cambios de localización de las estaciones o derivaciones locales características de la región.

Análisis de dispersión de las series temporales

Parra y Cortéz (2005) establecen la variabilidad interanual mediante análisis de dispersión temporal de las series de precipitación mensual. Este análisis de dispersión de las series temporales se realizó a través de la representación gráfica de "Boxplots" con determinación de valores atípicos moderados y severos, ya que estos pueden tener un impacto potencial de análisis estadísticos. Se utilizó el paquete estadístico INFOSTAT versión 9.0 (2008), para la determinación de todos los análisis del control de calidad de las series de precipitación.

Resultados y discusión

Detección de datos faltantes

En relación con la detección de datos faltantes se presentan dos ejemplos contrastantes. La serie mensual 1970-2000 de la estación Cantaura (figura 2a) muestra claramente la ocurrencia de datos faltantes mensuales durante todo el año; en general, esta estación se ubica por debajo de la longitud de registros recomendada para los análisis aplicados. Por su parte, la serie mensual 1970-2000 de la estación Bergantín (figura 2b) presenta entre 28 y 31 años sin datos mensuales faltantes, en consecuencia los datos de precipitación de esta estación poseen una calidad considerada como adecuada en función al bajo número de observaciones faltantes. Esta serie de datos puede ser confiable para realizar estu-

determining atypical values from mild to severe, since these may have a potential impact of statistical analyses. The statistical software INFOSTAT, version 9.0 (2008) was used, for the determination of all the quality control analyses of the precipitation series.

Results and discussion

Detection of the missing data

In relation to the detection of the missing data are presented two contrasting examples. The monthly series 1970-2000 of the Cantaura station (figure 2a) clearly shows the occurrence of the missing monthly data during all the year; in general, this station is under the registers longitude recommended for the applied analysis. On the other hand, the monthly series 1970-2000 of the Bergantín station (figure 2b) has from 28 and 31 years without missing monthly data, consequently the precipitation data of this station has a quality considered as adequate in function to the low number of missing observations. This series of data might be trustable to carry out researches related to the regional characterization of rain, but is not recommendable for analyzing extreme events that require the series with many years of registers.

Detection of extreme values

The stations with extreme monthly values were: Anaco, Barcelona, Cantaura, Clarines, El Chaparro Caserío, Guanape, Pariaguan, Puerto La Cruz, San Diego de Cabrutica, San Mateo, Santa Rosa and Úrica. The occurrence of this data is mainly due to: monthly values accumulated in the series of precipitation data,

dios relacionados con la caracterización regional de la lluvia pero no sería recomendada para análisis de eventos extremos que requieren de series con muchos años de registros.

Detección de valores extremos

Las estaciones que presentaron valores extremos mensuales fueron: Anaco, Barcelona, Cantaura, Clarines, El Chaparro Caserío, Guanape, Pariaguan, Puerto La Cruz, San Diego de Cabrutica, San Mateo, Santa Rosa y Úrica. La ocurrencia de estos datos se debe principalmente a: valores mensuales acumulados en las series de datos de precipitación, errores de transcripción a las bases de datos y a lluvias muy intensas de corta duración. De acuerdo a los resultados, el uso de estas estaciones estaría únicamente vinculado con estudios preliminares o caracterizaciones climáticas que no requieran de un alto nivel de detalle, debido principalmente a que son estaciones en las cuales se ve afectada la representatividad de la información, es decir las series de precipitación de las estaciones nombradas anteriormente no pueden ser utilizadas para la construcción de isoyetas, cálculos de probabilidades de ocurrencia de umbrales definidos, entre otras aplicaciones.

Estadística descriptiva básica

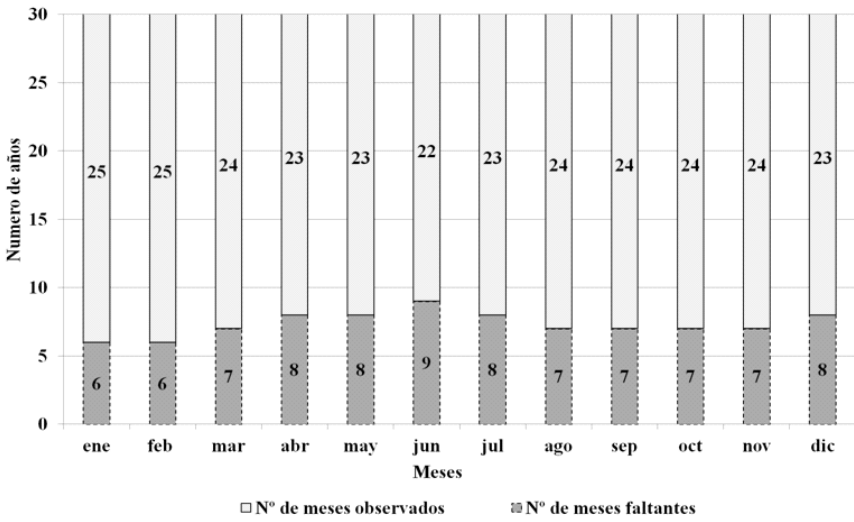
Las representaciones gráficas de cada estación constituidas por: la media, máximo, mínimo, percentiles 25%, 50% y 75%, mostraron valores aceptables y lógicos dentro del rangos reportados por la Primera Comunicación Nacional de Cambio Climático en Venezuela (MARN, 2005). Una característica de esta región es la marcada estacionalidad de la precipitación, de-

transcription errors in the database and very intense rains with short lasting period. According to the results, the use of these stations might be only related to the preliminary researches or climatic characterizations that do not require a high level of detail, mainly due to these stations are affected by the representation of the information, that is, the precipitation series of the latter stations cannot be used for the construction of the isohyets, calculus of the occurrence probabilities of defined thresholds, among other applications.

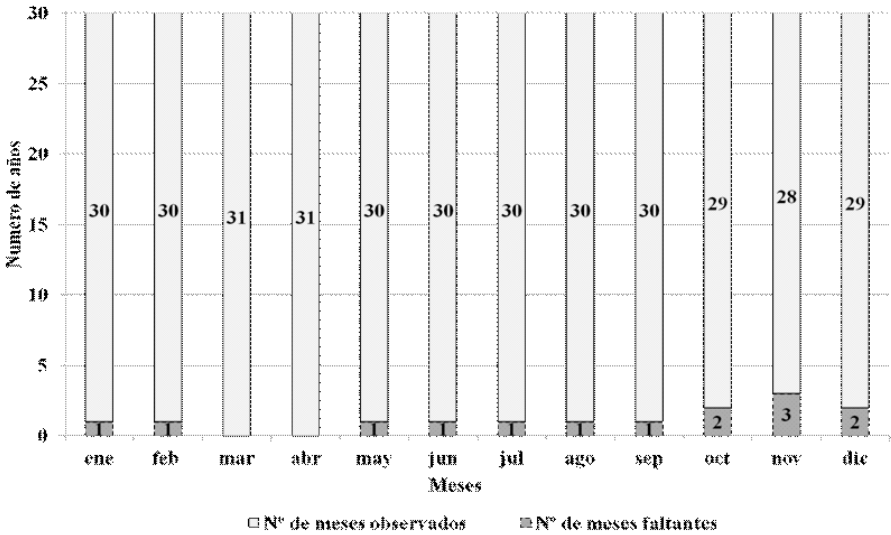
Basic descriptive statistics

The graphic representations of each station constituted by: the mean, maximum, minimum, percentiles 25%, 50% and 75%, showed acceptable and logic values inside the ranks reported by the First National Communication of Climatic Change in Venezuela (MARN, 2005). A characteristic of this region is the marked station of the precipitation, due to the alternated action of the inter tropical convergence zone. The rainy period involves more than 85% of the annual total of precipitation; since in the dry months there are little precipitations, the monthly precipitation is inferior to the reference evapotranspiration; it is very frequent that in February and March, the registered rain reaches zero millimeters (Martelo, 2004).

Figure 3 shows examples of the descriptive statistics applied in two stations of the Venezuelan Oriental Plains. The Onoto station (figure 3a) presents a station behavior typical of the Plains, represented by a rainy station that covers from six to seven months, and a dry station from six to



2a



2b

Figura 2. Número de valores faltantes y observados en la serie de precipitación mensual. 2a. Estación Cantaura. 2b. Estación Bergantín.

Figure 2. Number of missing and observed values in the monthly precipitation series. 2a. Cantaura station. 2b. Bergantín station.

bida a la acción alternada de la zona de convergencia intertropical. El periodo lluvioso concentra más del 85% del total anual de precipitación, mientras que en los meses secos llueve muy poco siendo la precipitación mensual inferior a la evapotranspiración de referencia; es muy frecuente que en los meses de febrero y marzo la lluvia registrada sea de cero milímetros (Martelo, 2004).

En la figura 3 se muestran ejemplos de la estadística descriptiva aplicada a dos estaciones de los Llanos Orientales Venezolanos. La estación Onoto (figura 3a) presenta un comportamiento estacional típico de los Llanos representado por una estación lluviosa que abarca entre seis a siete meses y una estación seca entre seis y cinco meses. Con relación a la estación Cumana (figura 3b) presenta precipitaciones escasas a lo largo del año, concentradas mayormente de diciembre a marzo debido a la influencia de los frentes fríos que llegan a las zonas costeras del país.

Homogeneidad de las series

Con relación a la prueba de Wald-Wolfowitz, el cuadro 2 muestra valores de probabilidad de algunas de las estaciones ubicadas en el estado Anzoátegui. Las estaciones estudiadas presentaron valores de probabilidad que favorecen a la hipótesis nula; es decir, que la serie de datos se comporta de manera aleatoria ($P > 0,05$). Se observó que solo en algunos meses del año mostraron valores de probabilidad que no favorecen la hipótesis nula. En este caso se recomienda la aplicación de procedimientos de homogeneización de las series en aquellas estaciones en las cuales no se hayan encontrado su-

five months. In relation to the station Cumana (figure 3b), it presents scarce precipitations throughout the year, mainly focused from December to March, due to the influence of cold periods that reach to the coast areas of the country.

Homogeneity of the series

In relation to the Wald-Wolfowitz test, table 2 shows the probability values of some of the stations located in Anzoátegui state. The studied stations presented probability values that favor the null hypotheses; that is, that the data series behaves at random ($P > 0.05$). In some months of the year were observed probability values which did not favor the null hypotheses. In this case, is recommended the application of homogenization procedures of the series in those stations without enough evidence found for the rejection of the null hypothesis.

Two figures are presented about the temporal series for the stations where the test failed in accepting the null hypothesis. Figure 4a shows that the lack of randomness in the variable precipitation is due to the change of the mean in determined periods of time, presenting a tendency to increase. On the other hand, figure 4b shows the temporal series of August, being this one of the wettest month of the year, some dry years appeared, influencing in the behavior of the mean. Finally, was seen that the lack of randomness was not related to a cyclic alternation of the precipitation series.

Dispersion analysis of the temporal series

Figure 5a shows the dispersion of the monthly precipitation series for the station Guanape, the atypical

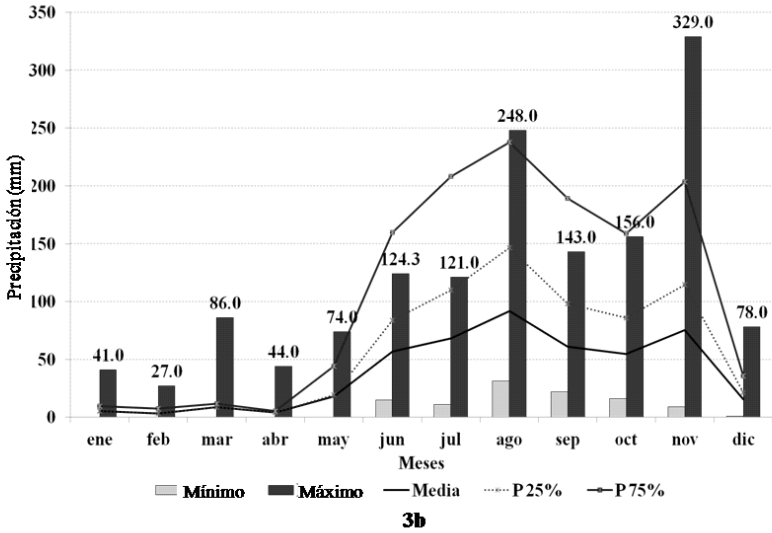
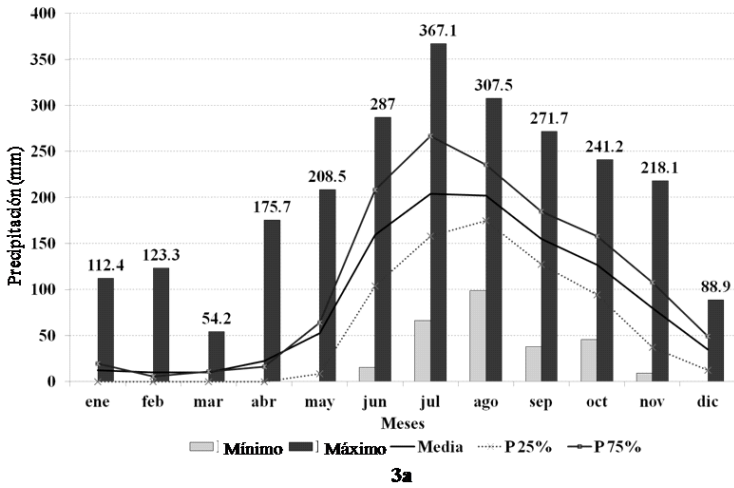


Figura 3. Comportamiento de la precipitación mensual (mm) para el periodo (1970-2000). 3a. Estación Onoto. 3b. Estación Cumana.

Figure 3. Behavior of the monthly precipitation (mm) for the period (1970-2000). 3a. Onoto station. 3b. Cumana station.

Cuadro. 2. Valores de probabilidad a la derecha del estadístico de prueba P (U>U(r)) para la prueba de correlación serial (homogeneidad y carácter aleatorio) Según Wald-Wolfowitz (Sneyers, 1990).

Table 2. Probability values on the right of the statistical test P (U>U(r)) for the serial correlation test (homogeneity and randomized trait) according to Wald-Wolfowitz (Sneyers, 1990).

Estación	P (U>U(r)) para la prueba de correlación serial (homogeneidad y carácter aleatorio)											
	ENE	FEB	MAR	ABR	MAY	JUN	JUL	AGO	SEP	OCT	NOV	DIC
San Mateo	0,8956	0,6081	0,6781	0,5526	0,5108	0,0256	0,5108	0,0465	0,3935	0,6746	0,8885	0,7000
El Tigre	0,4423	0,0505	0,2459	0,8956	0,1461	0,4153	0,0703	0,8531	0,9707	0,6432	0,4461	0,8956
Los Yopales	0,2004	0,2003	0,1805	0,5509	0,5565	0,4004	0,6627	0,9218	0,1688	0,1710	0,9139	0,8989
Pariaguan	0,0067	0,5171	0,1805	0,0937	0,8821	0,9222	0,1710	0,5000	0,8312	0,0144	0,2154	0,8989
San Diego	0,1560	0,5198	0,7000	0,8307	0,4618	0,7632	0,2042	0,8983	0,8925	0,8307	0,6070	0,3671
Clarines	0,7979	0,0730	0,0730	0,0576	0,1057	0,8725	0,6797	0,5623	0,9647	0,2949	0,1160	0,2004
Guaribe T.	0,1244	0,0048	0,1244	0,0912	0,9525	0,6636	0,0228	0,9304	0,9911	0,3928	0,0305	0,1244
Campo Mata	0,3056	0,4423	0,1244	0,5960	0,8117	0,4755	0,3126	0,6469	0,4648	0,2036	0,2382	0,0405
Aragua de Barcelona	0,6085	0,9999	0,4423	0,3965	0,2454	0,8429	0,6614	0,0694	0,6614	0,9332	0,7846	0,7979

Nota: Las celdas sombreadas representan valores de P (U>U(r)) <alfa (alfa=0,05), lo que indica que la serie no cumple con el carácter aleatorio.

ficientes evidencias para el rechazo de la hipótesis nula.

A continuación se presentan dos gráficos de las series temporales para las estaciones en las que la prueba falló en aceptar la hipótesis nula. La figura 4a muestra que la falta de aleatoriedad en la variable precipitación, se debe al cambio de la mediana en determinados periodos de tiempo, presentando una tendencia a aumentar. Por su parte la figura 4b muestra la serie temporal del mes de agosto, siendo este mes uno de los más húmedos del año, se presentaron ciertos años secos, repercutiendo en el comportamiento de la mediana. En síntesis se evidenció que la falta de aleatoriedad no estuvo asociada a una alternancia cíclica de las series de precipitación.

Análisis de dispersión de las series temporales

La figura 5a muestra la dispersión de la serie de precipitación mensual para la estación Guanape, los valores atípicos tales como la ocurrencia de eventos extremos de lluvia, errores en la toma de datos, errores en la transcripción a la base de datos o falta de calibración del equipo; estos valores atípicos se concentraron en los meses de la época seca (enero a marzo) y de transición (abril y noviembre) en la región de los Llanos orientales. Los diagramas de los meses húmedos (mayo a octubre) presentaron una mayor variabilidad que en los meses secos (figura 5b). En general, la ocurrencia de estos valores atípicos se pueden explicar por el sesgo que produce la ocurrencia de muchos valores ceros en las observaciones.

El cuadro 3 presenta un resumen de los procedimientos estadísticos aplicados a las series de precipitación, el

values, such as the occurrence of extreme events of rain, errors in the obtaining of the data, errors in the transcription of the data to the database or lack of calibration of the equipment; these atypical values were concentrated during the dry-month season (January to March) and transition (April to November) in the Oriental Plains. The diagrams of the wet months (May to October) presented higher variability than in the dry months (figure 5b). Generally, the occurrence of these atypical values can be explained by the bias that the occurrence produces in many zero values in the observations.

Table 3 presents a resume of the statistical procedures applied to the precipitation series, the main problem, which is characteristic of the series, is constituted by the monthly missing data, most of the stations have low values (<10%), however, there are some stations with a high proportion of missing data (10-37%), this lack of data somehow affects the longitude of the series about to be analyzed for different purposes, becoming an important problem on the representation of the data, consequently, in the accuracy of the results product of the analysis obtained.

Conclusions and recommendations

The quality control of the data using the applied analyses, allowed verifying the width of the variation and the consistency of the registered data in the climatic stations. Every procedure applied in this research allowed

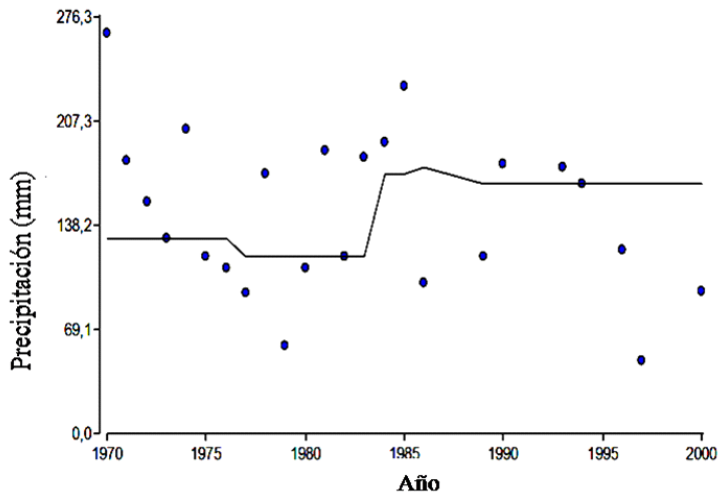
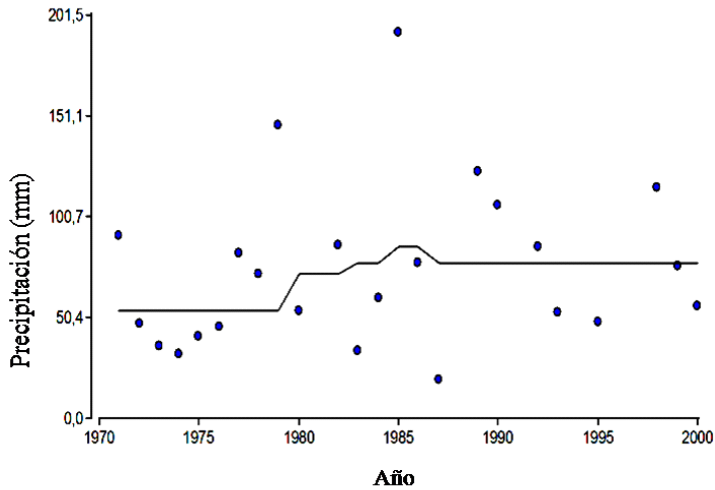


Figura 4. Gráficos de mediana móvil con período igual a 10 años para las series que resultaron no aleatorias en la prueba de Wald -Wolfowitz. 4a. Estación Guaribe-Tenepe para el mes de noviembre. 4b. Estacion San Mateo para el mes de agosto.

Figure 4. Mobile Mean figures with a 10-year period for the not-randomly series in the test Wald -Wolfowitz. 4a. Station Guaribe-Tenepe in November. 4b. Station San Mateo during August.

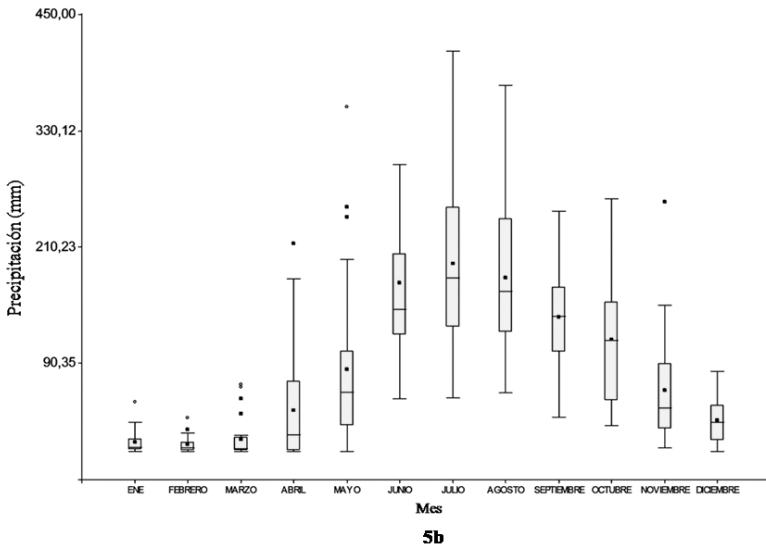
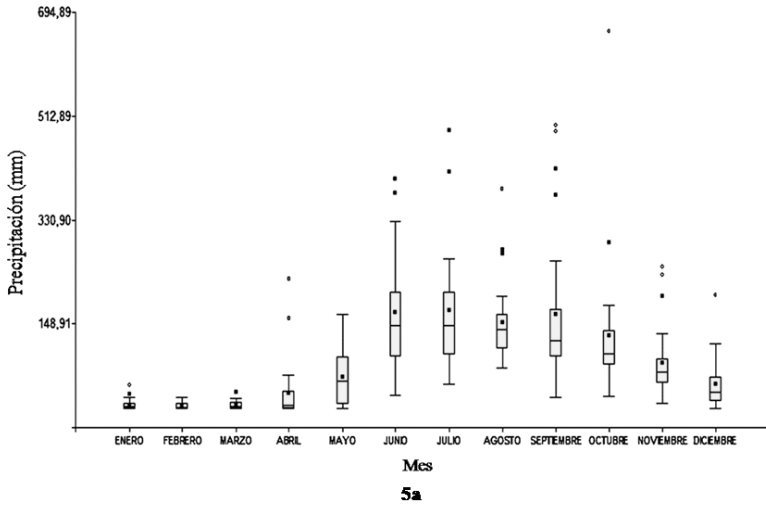


Figura 5. Diagramas “Boxplot” de la precipitación mensual (mm) para el periodo 1970-2000/ 5a. Estación Guanape, 5b. Estación El Tigre.

Figure 5. Boxplot diagrams of the montly precipitation (mm) for the period 1970-2000. 5a. Guanape station. 5b. El Tigre station.

Cuadro 3. Resultados de la aplicación de los procedimientos para el control de calidad de las series de precipitación de las estaciones climatológicas estudiadas.

Table 3. Results of the application procedures for the quality control of the precipitation series of the studied climatic stations.

N	Estación	Datos faltantes mensuales (%)	Datos extremos	Estadística descriptiva	Nº de series aleatorias	Análisis de dispersión
1	Anaco	10	Si	N	12	N
2	Aragua de Barcelona	2	No	N	12	N
3	Barcelona	0	Si	N	12	N
4	Bergantín	4	No	N	12	N
5	Cachipo	37	No	N	11	N
6	Campo Mata	21	No	N	11	N
7	Cantaura	24	Si	N	10	N
8	Cariaco-Muelle	9	No	N	11	N
9	Ciudad Bolívar	2	No	N	12	N
10	Clarines	15	Si	N	12	N
11	Cumana	1	No	N	12	N
12	El Chaparro Caserío	6	Si	N	10	N
13	El Guamo	16	No	N	11	N
14	El Tigre	0	No	N	12	N
15	Guanape	14	Si	N	10	N
16	Guaribe-Tenepe	15	No	N	9	N
17	Gurí	0	No	N	12	N
18	Hacienda la Concordia	4	No	N	9	N

Nota: (N: Patrón Normal esperado)

Cuadro 3. Resultados de la aplicación de los procedimientos para el control de calidad de las series de precipitación de las estaciones climatológicas estudiadas (Continuación).

Table 3. Results of the application procedures for the quality control of the precipitation series of the studied climatic stations (Continuation).

N	Estación	Datos faltantes mensuales (%)	Datos extremos	Estadística descriptiva	N° de series aleatorias	Análisis de dispersión
19	La Cerca	15	No	N	10	N
20	Los Yopales	16	No	N	12	N
21	Maturín	0	No	N	12	N
22	Onoto	14	No	N	9	N
23	Pariaguan	4	Si	N	10	N
24	Puerto la Cruz	12	Si	N	12	N
25	Querequal	17	No	N	10	N
26	San Diego Cabrutica	12	Si	N	12	N
27	San Mateo	15	Si	N	10	N
28	Santa Clara II	16	No	N	9	N
29	Santa Rosa	13	Si	N	9	N
30	Upata	2	No	N	11	N
31	Urica	27	Si	N	9	N
32	Valle de la Pascua	1	No	N	12	N
33	Viento Fresco	21	No	N	11	N
34	Zuata	18	No	N	10	N

Nota: (N: Patrón Normal esperado)

principal problema característico de las series lo constituyen los datos faltantes mensuales, la mayoría de las estaciones presentan valores bajos (<10%), sin embargo hay ciertas estaciones que poseen una alta proporción de datos faltantes (10-37%), esta ausencia de datos de alguna manera afecta la longitud de las series que se quieren analizar para diversos fines, resultando un problema importante sobre la representatividad de los datos y en consecuencia, en la confiabilidad de los resultados producto de los análisis obtenidos.

Conclusiones y recomendaciones

El control de calidad de datos mediante los análisis aplicados en este estudio permitió verificar la amplitud de variación y la consistencia de los datos registrados en las estaciones climatológicas. A través de cada uno de los procedimientos aplicados en este estudio se obtuvo una información importante acerca de la proporción de datos faltantes y detección de valores extremos, una descripción estadística básica, homogeneidad de las series de precipitación y la dispersión de las series temporales mediante representaciones gráficas, lo cual indica que esta metodología se ajusta a la variable precipitación de acuerdo a las condiciones intertropicales. A manera general el mayor problema de las series está representado por los datos faltantes, seguido de la presencia de valores extremos y la falta de homogeneidad de algunas series.

La no homogeneidad de las series se debió principalmente a la interven-

obtaining important information about the proportion of the missing data and detection of extreme values, a basic statistical description, and homogeneity of the precipitation series and the dispersion of the temporal series using graphic representations, which indicates that this methodology is adjusted to the variable precipitation, according to the inter-tropical conditions. Generally, the biggest problem of the series is represented by the missing data, followed by the presence of extreme values and the lack of homogeneity of some series.

The lack of homogeneity of the series was mainly due to the intervention of two factors, such as the existence of the values originated by a procedure error and the presence of atypical or extreme values, which can be considered or not in the analysis in function of the objectives. In the case of the extreme values, it is recommended to replicate the analysis with or without atypical data, with the aim of analyzing and determining their influence on the final results.

The studied data series (Aragua de Barcelona, Bergantín, Ciudad Bolívar, Cumana, El Tigre, Gurí, Maturín and Valle La Pascua) does not require the implementation of the adjustment and correction measures. On the other hand, the rest of the series present some extreme values or lack of homogeneity in the data.

The series Cachipo, Campo Mata, Cantaura, Urica and Viento Fresco, presented a high percentage of missing data (>20%). This indicates that these missing observations affect the longitude of the series and might have a potential impact on the

ción de dos factores tales como la existencia de valores originados por un error de procedimiento y la presencia de valores atípicos o extremos el cual puede ser considerado o no en el análisis en función de los objetivos. En el caso de los valores extremos se recomienda replicar el análisis con y sin los datos atípicos con la finalidad de analizar y determinar su influencia sobre los resultados finales.

Las series de datos estudiadas (Aragua de Barcelona, Bergantín, Ciudad Bolívar, Cumana, El Tigre, Gurí, Maturín y Valle La Pascua) no ameritan la implementación de medidas de ajuste y corrección. Por su parte, el resto de las series presenta ciertos valores extremos o la falta de homogeneidad en los datos.

Las series Cachipo, Campo Mata, Cantaura, Urica y Viento Fresco, presentaron un alto porcentaje de datos faltantes (>20%). Lo que indica que estas observaciones pérdidas afectan la longitud de las series y pueden llegar a tener un impacto potencial sobre la representatividad de los datos y la confiabilidad de los resultados que se generan de sus análisis.

En este sentido, se recomienda utilizar series de datos de precipitación a los cuales se les hayan realizado las pruebas iniciales que se consideraron pertinentes sin presentar datos atípicos, observaciones faltantes, ni tendencia fuera de lo normal, de esta manera el investigador contara con un conjunto de datos confiables evitando el uso de información con errores de estimación.

representation of the data and accuracy of the results that are generated from the analyses.

In this sense, it is recommended to use series of precipitation data with initial tests that were considered important, without presenting atypical data, missing observations, or abnormal tendency; thus, the researcher will have a complex of trustable data, avoiding the use of information with estimation errors.

End of english version

Literatura citada

- Ablan, M., R. Andressen, M.P. Vargas y M. Acevedo. 2008. Propuesta metodológica para el control de calidad de datos de precipitación. *Agronomía Trop.* 58(1): 57-60. [En línea] Disponible en: http://sian.inia.gov.ve/repositorio/revistas_ci/Agronomia%20Tropical/at5801/pdf/ablan_m.pdf [Consulta: 2011, 02 de octubre].
- Arnell, N.W. 2003. Relative effects of multi-decadal climate variability and changes in the mean and variability of climate due to global warming: future streamflows in Britain. *En Journal of Hydrology* 270(3-4), 195-213.
- Cortéz, A., F.A. Ovalles, M.F. Rodríguez, J.C. Rey y M.C. Núñez. 2005. Análisis geoestadístico de la variabilidad temporal, a una escala anual y mensual, de parámetros climáticos en un campo experimental, estado Aragua, Venezuela. *Rev. Agronomía Trop.* 55(3): 327-342. Disponible en: http://sian.inia.gov.ve/repositorio/revistas_ci/Agronomia%20Tropical/at5503/pdf/cortez_a.pdf. [Consulta: 2011, 02 de octubre].
- Crawley, M. 2002. *Statistical Computing: An Introduction to Data Analysis using S-plus*. John Wiley & Sons, England. 761 p.

- Griffis, V.W., J.R. Stedinger y T.A. Cohn. 2004. LP3 Quantile Estimators with Regional Skew Information and Low Outlier Adjustments. En *Water Resources Research* 40(10) 145-156.
- Guenni, L., E. Degryze y K. Alvarado. 2008. Análisis de la tendencia y la estacionalidad de la precipitación mensual en Venezuela. *Revista Colombiana de Estadística*. 31(1) 41 – 65. Disponible en <http://www.emis.ams.org/journals/RCE/ingles/body/v31n1a03GuenniDegryzeAlvarado.pdf>. [Consulta: 2011, 02 de octubre].
- Infostat. 2008. Infostat for Windows Version 9.0. Grupo Infostat. Inc. Facultad de Ciencias Agrícolas. Universidad Nacional de Córdoba. Argentina.
- Javelle, P., T.B.M.J. Ouarda y B. Bobée. 2003. Flood regime definition using the flood-duration frequency approach: Application to the Provinces of Quebec and Ontario, Canada. En *Hydrological Processes*, 17(18), 3717–3736.
- Jones, P. 1987. Current availability and deficiencies in data relevant to Agro-ecological studies in the Geographic Area Covered by IARCs. In: *Agricultural Environments. Proceedings of the Rome Workshop on Agro-ecological classification and mapping*. CAB, International. Wallingford, UK. P. 69-111.
- Martelo, M.T. 2004. Consecuencias ambientales generales del cambio climático en Venezuela, Trabajo de ascenso, Universidad Central de Venezuela, Facultad de Agronomía, Maracay, Venezuela. 172 p
- MARN. 2005. Primera Comunicación Nacional de Cambio Climático en Venezuela. MARN, Programa de las Naciones Unidas para el Desarrollo, Fondo Mundial para el Medio Ambiente. Venezuela, 164 p.
- Millano, J.L., F. Paredes y I. Vivas. 2007. Efecto de la Oscilación Meridional (ENSO) y la temperatura superficial del océano Atlántico sobre la distribución espacio-temporal de las lluvias en el estado Cojedes. *Revista Agrorranía*. 4: 103-116
- Parra, R. y A. Cortez. 2005. Control de calidad de series de precipitación de las series de precipitación del INIA Venezuela en el periodo 1970-2000. *Rev. Arg. de Agrometeorología*, (5-6): 63-73.
- Reilly, C.F. y C.N. Kroll, 2003. Estimation of 7-day, 10-year low-streamflow statistics using baseflow correlation. En *Water Resources Research*, 39(9), 1236.
- Sneyers, R. 1990. On the statistical analysis of series of observations. WMO-N° 415, Technical Note N°143. Ginebra, Suiza. 193 p.
- UNESCO. 2006. Guía metodológica para la elaboración del mapa de zonas áridas, semiáridas y subhúmedas secas de América Latina y El Caribe. CAZALAC. Documentos Técnicos del PHI_LAC, N° 3. La Serena, Chile. 49 p.
- UNESCO-ROSTLAC. 1982. Guía metodológica para la elaboración del balance hídrico de América del Sur. Estudios e Informes en Hidrología. UNESCO-ROSTLAC. Montevideo, Uruguay. 38 p.
- WMO. 1989. Proceedings of international workshop on precipitation measurements. Switzerland. 584 p.
- 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: 113-138.
- Yue, S. y P. Pilon, 2004. A comparison of the power of the t test, Mann-Kendall and bootstrap tests for trend detection. En *Hydrological Sciences*, 49(1), 21–37.