

PROPUESTA METODOLÓGICA PARA EL CONTROL DE CALIDAD DE DATOS DE PRECIPITACIÓN¹

METHODOLOGICAL PROPOSAL FOR THE CONTROL OF QUALITY OF DATA OF PRECIPITATION¹

Magdiel Ablan*, Andressen Rigoberto**, Mary Pili Vargas***, Miguel Acevedo****

¹ El proyecto fue financiado por el FONACIT a través de la subvención S1-97000206.

* Investigadores. Universidad de los Andes, Facultad de Ingeniería, Centro de Simulación y Modelos (CESIMO), **Centro de Investigaciones Atmosféricas y del Espacio (CIAE), ***Escuela de Ingeniería de Sistemas, Mérida, Venezuela. ****Actualmente en el Instituto de Ciencias Aplicadas y Departamento de Geografía, E.E.U.U. Email: randss@ula.ve

RESUMEN

Este trabajo tiene como objetivo presentar una propuesta metodológica para analizar datos climáticos históricos a ser utilizados en estudios de la variabilidad temporal de la precipitación en una región. Específicamente, en este caso, se trata de los Llanos venezolanos. La propuesta con la selección de las estaciones a ser utilizadas, continúa con la detección de posibles valores extremos o aberrantes, la estimación de valores faltantes y la verificación de la homogeneidad de las series de datos históricos. Aunque existen un gran número de técnicas para tratar cada uno de los pasos anteriores, en este trabajo, se hace énfasis en técnicas numéricas sencillas que puedan ser fácilmente entendidas y aplicadas en esta u otra región de características geográficas similares. Además, los análisis son realizados en el lenguaje R (<http://www.r-project.org>), que es software libre, por lo que está disponible para todos los usuarios. Debido a las importantes aplicaciones de los estudios sobre variabilidad climática y a la falta de disponibilidad de datos en Venezuela, esta propuesta viene acompañada de un portal Web (<http://cesimo.ing.ula.ve/ENOS/>) donde se pueden consultar y descargar los datos utilizados en el proyecto, así como los programas asociados con cada uno de los métodos.

Palabras Clave: Variabilidad climática; métodos de análisis de datos climáticos; Llanos venezolanos.

SUMMARY

This paper presents a methodological proposal for the quality control of historical rainfall data, a step previous in any applied climatological analysis, like for example, in studies of the temporal variability of the precipitation in a region; specifically, in this case, it is applied to the Venezuelan 'Llanos'. The proposal begins with the selection of the stations to be used, then continues with the detection of possible extreme or aberrant values, following with the estimation of missing data and then testing for homogeneity of the historical data series. Although there are various techniques for dealing with the previously mentioned steps, in this paper emphasis is placed on simple numerical methods that could be easily understood and applied in this or another region of similar geographical characteristics. Also, the analyses are carried out in "R" (<http://www.rproject.org>), a free software that has the advantage of being available for all users. Due to the importance of data quality control and to the limitations on available climate data in Venezuela, this proposal is accompanied with a Website (<http://cesimo.ing.ula.ve/ENOS/>) from where the data and the programs that are used in this project can be consulted and downloaded.1.

Key Words: Variabilidad climática; métodos de análisis de datos climáticos; Llanos venezolanos.

INTRODUCCIÓN

La variabilidad temporal de la precipitación influye en múltiples actividades humanas, sobre todo 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. Debido a su importancia, en este trabajo se propone una metodología o protocolo que permite verificar la calidad de los datos. Este protocolo se siguió como un primer paso en un estudio que tenía como objetivos estudiar la variabilidad temporal de la precipitación en una región determinada.

MATERIALES Y MÉTODOS

Selección de las estaciones

Se elaboró un inventario de todas las estaciones pluviométricas en la zona de estudio. Estas estaciones pertenecen al MPPA y al SM-FAV. Se consideraron 257 estaciones con registros mensuales. Para cada estación se determinó, en lo posible, la información de metadata básica (ubicación de la estación, variables medidas, fecha de inicio, fecha de culminación, etc.) y el período de datos disponibles.

En la selección del período óptimo de datos se trató de conseguir un compromiso entre el número de estaciones disponibles y el número de años de registro de datos sin períodos extensos de datos faltantes. De acuerdo a la Organización Meteorológica Mundial (WMO, 1969), para estudios de variabilidad de la precipitación, en regiones tropicales planas continentales, tales como los Llanos venezolanos, se requiere un mínimo de 40 años de datos.

Por otro lado, es necesario considerar un número suficiente de estaciones para poder dilucidar si existe un patrón geográfico determinado. Finalmente se seleccionaron 68 estaciones, con un registro de 38 años. La mayoría de las estaciones finalmente seleccionadas (78%) poseían 10 o menos valores faltantes.

Selección de las estaciones vecinas

Para cada una de las estaciones consideradas, se realizó una selección de las estaciones vecinas que serían usadas como referencia para la determinación de los valores extremos y el cálculo de valores faltantes. Las estaciones

fueron seleccionadas entre las más cercanas, de manera tal, que en lo posible estuvieran orientadas en los diferentes sectores de arcos de 120 grados de un círculo centrado en la estación dada (por ejemplo una estación en la dirección noreste, otra al sur y otra en la dirección noroeste), siguiendo la recomendación propuesta por Paulhus y Kohler (1952). La mediana de la distancia mínima aproximada entre un par cualquiera de las estaciones consideradas fue de 18 km y la de la máxima fue de 33 km.

Detección de valores extremos

Para detectar la posible presencia de valores extremos en un conjunto de datos pluviométricos se usó una función basada en la desviación absoluta de la mediana, *mad*, un estimador robusto de la variabilidad de los datos que es menos sensible a los valores extremos que los estimadores tradicionales como la desviación estándar. 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 (Crawley, 2002). Una vez identificado un conjunto de datos con posibles valores extremos, éstos se identificaban graficando los datos mensuales de cada estación con relación a su media y desviación estándar, y estudiando la relación de este valor con los valores registrados en estaciones vecinas para el mismo período de tiempo. Los valores que se consideren extremos reales o aberrantes, se estimarán luego usando el mismo procedimiento que para datos faltantes.

Estimación de datos faltantes

Para seleccionar el método a ser usado para la estimación, se probaron tres de los procedimientos más sencillos, como se indica en Searcy y Hardison (1966), a saber: el método del promedio, el método de regresión lineal múltiple y el método de las normales de precipitación. Para cada estación se determinó el mejor método a usar siguiendo el siguiente procedimiento:

- 1) Se seleccionó una muestra aleatoria de 30 valores de la estación. Si alguno de los valores seleccionados resultaba ser faltante, este se descartaba y se procedía con la selección del siguiente valor.
- 2) Se procedió a estimar cada uno de los valores seleccionados con los tres métodos considerados.
- 3) Se calculó el error cometido en la estimación en cada caso y globalmente para la estación. Para la estimación

del error se utilizó el error absoluto y el error cuadrático medio. Una vez determinado el mejor método a usar en cada caso, se procedió a estimar los valores faltantes en cada una de las estaciones.

Verificación de la homogeneidad de las estaciones

Para verificar la homogeneidad de las estaciones utilizadas en este estudio, se siguió el método no paramétrico propuesto por la OMM, que está basado en las series de secuencias (Thom, 1966). En principio, las estaciones no homogéneas no se deben considerar en un análisis de variabilidad climática, ya que podría, erróneamente, interpretarse como evidencia de un cambio climático, lo que en realidad puede ser considerado como consecuencia de cambios en la tecnología, en el personal que efectúa la medición, en los procedimientos empleados en la operación de las estaciones, o cambios de localización de las estaciones o efectos locales. Sin embargo, aunque existen casos en los que las series no son homogéneas, y se pudiera verificar que dicha situación no se debe a errores de observación o de procedimientos, estos no se consideraron en este trabajo, debido a la falta de información de metadata.

RESULTADOS Y DISCUSIÓN

Para la selección de las estaciones a considerar en el estudio, la comparación relativa de los períodos de datos disponibles arrojó como el período más apropiado, el comprendido entre 1960 y 1997, lo que incluye 38 años de registros, apenas un poco por debajo de lo requerido por la Organización Meteorológica Mundial; y una selección de 68 estaciones, con datos mensuales. De las estaciones consideradas, aproximadamente el 72% de las estaciones fueron identificadas como sospechosas de poseer por lo menos un valor extremo.

Sin embargo, luego del procedimiento de comparación con las estaciones vecinas sólo 55 valores del conjunto de datos, aproximadamente el 12%, se consideraron como errores de medición. El análisis de los de datos faltantes reveló que aproximadamente 78% poseían menos de 10 datos faltantes, es decir, menos del 3% con respecto al total de los valores considerados. El mejor método de estimación de los valores faltantes, utilizando como criterio el error cuadrático medio fue el método de regresión lineal (en 47,06% de las estaciones); el método de la normal de precipitación en un 39,71% de los casos y el del promedio simple en el restante 13,23%.

Con respecto a la homogeneidad, de las 68 estaciones consideradas en este trabajo, 16 resultaron ser no-homogéneas.

Por limitaciones de espacio, el siguiente Cuadro muestra algunos de los resultados obtenidos para las estaciones consideradas. Detalles sobre los resultados pueden consultarse en la página web del proyecto <http://cesimo.ing.ula.ve/enos>.

CUADRO. Muestra algunos de los resultados de homogeneidad para las 68 estaciones consideradas.

| Serial | Nombre | Ext. | Falt. | Hom. |
|--------|-----------------------|------|-------|------|
| 1751 | Puerto La Cruz | si | 4 | si |
| 1756 | El Chaparro-Caserío | si | 2 | no |
| 1773 | Barcelona-Aeropuerto | si | 4 | si |
| 1796 | Bergantín | no | 5 | si |
| 2170 | Guafas | no | 14 | no |
| 2171 | Suruguapo | no | 0 | si |
| 2239 | Agua Blanca | si | 0 | si |
| 2246 | Hacienda Camburito | no | 15 | si |
| 2253 | Hacienda Guache | si | 3 | si |
| 2259 | Payara | si | 0 | si |
| 2260 | Potreros | si | 0 | si |
| 2261 | Chabasquen | no | 19 | si |
| 2265 | Ospino | si | 2 | si |
| 2266 | Ospino-La Estación | no | 0 | si |
| 2267 | Píritu | si | 1 | si |
| 2273 | Biscucuy | no | 2 | si |
| 2277 | Colonia Turén- IAN | si | 22 | si |
| 2282 | Desembocadero | no | 15 | si |
| 2286 | Hacienda San Rafael | si | 5 | si |
| 2295 | Campamento Las Marías | si | 2 | si |
| 2331 | San Rafael de Onoto | no | 0 | no |
| 2348 | Morita-Caserío | si | 0 | si |
| 2350 | Galera | si | 3 | si |
| 2425 | Lezama | si | 9 | si |
| 2426 | La Yeguera | si | 4 | si |
| 2448 | Ortiz | si | 0 | si |
| 2469 | El Sombrero | si | 8 | si |
| 2477 | Encrucijada-Calabozo | si | 7 | si |
| 2505 | Guatopo | no | 2 | no |

./... continúa

../ Continuación Cuadro.

| Serial | Nombre | Ext. | Falt. | Hom. |
|--------|--------------------------|------|-------|------|
| 2507 | San Francisco de Macaira | no | 22 | si |
| 2528 | Paso Real | si | 2 | si |
| 2601 | Soublette | no | 10 | si |
| 2602 | San José de Guaribe | si | 43 | no |
| 2605 | Guanape | si | 2 | si |
| 2608 | Clarines | si | 1 | no |
| 2640 | Kilómetro 133 - Coromoto | si | 9 | si |
| 2647 | Guaribe- Tenepe | si | 10 | si |
| 2648 | Onoto | si | 2 | si |
| 2650 | El Chaparro-Pueblo | si | 14 | no |
| 2703 | San Benardino | si | 3 | no |
| 2713 | Curataquiche | si | 0 | si |
| 2715 | Quiamare | no | 12 | no |
| 2717 | Santa Inés | si | 2 | no |
| 2722 | El Carito | si | 5 | no |
| 2725 | San Mateo | si | 3 | si |
| 2755 | Santa Rosa | si | 1 | si |
| 2757 | Anaco | si | 6 | si |
| 2764 | San Joaquín | si | 4 | no |
| 2766 | Cantaura | si | 29 | no |
| 3126 | Barinitas | no | 13 | si |
| 3128 | Barrancas | no | 2 | no |
| 3136 | Quebrada Seca | no | 2 | si |
| 3149 | Altamira - Barinas | no | 77 | si |
| 3158 | Torunos | si | 2 | si |
| 3162 | El Corozo-Palmitas | no | 3 | si |
| 3173 | San silvestre | si | 1 | si |
| 3212 | Sabaneta | si | 2 | si |
| 3214 | Puerto Las Animas | si | 1 | si |
| 3216 | Sipororo | no | 0 | no |
| 3222 | San Hipólito | si | 2 | si |
| 3261 | El Real | si | 31 | si |
| 3283 | Santa Lucia | si | 7 | si |
| 3403 | Los Naranjos | si | 2 | si |
| 3454 | Corozo Pando | si | 1 | si |
| 3502 | Palenque-Dispensario | si | 9 | si |
| 3712 | Pariguan | si | 6 | si |
| 3762 | San Diego de Cabrutica | si | 11 | no |
| 4404 | San Fernando de Apure | si | 0 | si |

Extr.: Extremos; Falt.: Faltantes; Hom.: Homogénea.

CONCLUSIÓN

- Se desarrollaron e implementaron un conjunto de métodos y programas estadísticos que permiten estudiar y analizar la variabilidad de la precipitación en los Llanos venezolanos. Los métodos, datos y resultados de los análisis se encuentran disponibles en un portal web (<http://cesimo.ing.ula.ve/ENOS>) que permite un mayor uso de los mismos.
- La metodología propuesta es sencilla y tanto los métodos como sus resultados pueden ser aplicados a un gran rango de situaciones donde es de interés caracterizar la variabilidad de la precipitación.

AGRADECIMIENTO

Nuestro agradecimiento a la Lic. Cherly Guerrero por su ayuda con el procesamiento de los datos.

BIBLIOGRAFÍA

- Crawley, M. 2002. Statistical Computing: An Introduction to Data Analysis using S-plus. John Wiley & Sons, England. 761 p.
- Paulhus J. L. H and M. A. Kohler, 1952. Interpolation of Missing Precipitation Records. Monthly Weather Review. 80 p.
- Searcy, J. K. and C. H. Hardison. 1966. Double-mass curves. Manual of Hydrology: Part 1. General surface-water techniques. Washington: Geological Survey Water-Supply Paper 1541-B. 66 p.
- Thom, H. 1966. Some Methods of Climatological Analysis. Secretaria of the World Meteorological Organization. Geneva, Switzerland.
- Organización Meteorológica Mundial (WMO). 1969. A note on climatological normals. Technical Note No. 84. WMO-No. 208.TP.108. Geneva, 19 p.