

# Metodologías indirectas para la detección y corrección de series de tiempo heterogéneas

## Indirect Methodologies to Identify and Adjust inhomogeneities in Time Series

Omar Javier Daza Leguizamón\*  
Rigaud Sanabria Marin\*\*

### Resumen

Para la realización de estudios hidrológicos y climáticos, que involucran series de tiempo, es necesario que estas se encuentren libres de errores que afecten la calidad de los resultados. Se describen las metodologías indirectas objetivas para la detección y corrección de heterogeneidades en series de tiempo, compuestas por cuatro fases: revisión de la metadata, control de calidad de las series, elaboración de la serie de referencia y la identificación y corrección de heterogeneidades. Se presentan criterios propuestos por diferentes autores en la aplicación de las diferentes fases descritas. Finalmente se listan algunos programas de computador, de uso libre, diseñados para la corrección de heterogeneidades o que cuentan con herramientas para evaluar la homogeneidad.

**Palabras clave:** Series de tiempo, Datos Hidroclimáticos, Hidrología, Climatología, Software libre.

### Abstract

During the hydrologic and climatologic studies is necessary to have times series without inhomogeneities that could affect the quality of the results. This work describes the indirect objective methodologies for the detection and correction of the heterogeneous time series. This methodology has four steps: The metadata inspection, the quality control, the building of a reference time series and the breakpoints or the trend identification and data adjustment.

We present some criteria and experiences of different authors in each step. We describe also, some software packages which are designed to analyze the time series or which have tools to detect and correct any heterogeneous trend in the time series.

**Key words:** Time Series, Hydro Climate Data, Hydrology, Climatology, Free Software.

\* Ingeniero Civil. Joven Investigador del Grupo de Investigación en Geomática y Ambiente. Facultad de Ingeniería, Universidad Pedagógica y Tecnológica de Colombia. Correo e.: kianosla@gmail.com

\*\* Ingeniero de Sistemas, Ph.D. en Geografía. Docente asociado de la Escuela de Ingeniería Civil. Coordinador Grupo de Investigación en Geomática y Ambiente. Universidad Pedagógica y Tecnológica de Colombia. Correo e.: sanabriarigaud@yahoo.com

## 1. Introducción

La calidad de las series de tiempo mensuales y anuales de datos hidroclimáticos, como la precipitación y la temperatura, en estudios de índole climático e hidrológico resulta un factor determinante en la confiabilidad de los resultados obtenidos. Esta calidad se puede ver afectada por diferentes factores relacionados con datos ausentes, valores dudosos y heterogeneidades, cada uno de los cuales debe ser analizado con detenimiento antes de la utilización de las series. Hoy en día, se cuenta con gran cantidad de técnicas y metodologías para el tratamiento y corrección de las series de tiempo, que permiten mejorar la confiabilidad de los datos que se van a utilizar [1]. En este artículo se describen las fases generales que se siguen dentro de la aplicación de las metodologías indirectas objetivas para la detección y corrección de heterogeneidades en series de tiempo anuales y mensuales.

Para mejorar la exactitud en análisis climáticos que involucren series históricas, estas deben ser lo más homogéneas posible; infortunadamente, la mayoría de las series disponibles han sido afectadas por factores no climáticos que causan que dichos datos sean poco representativos de las variaciones climáticas que ocurren en el tiempo. Por esto la importancia de corregir las series históricas, o al menos, determinar el posible error que pueden causar las heterogeneidades presentes [1, 2]. Dada la necesidad de contar con series de tiempo homogéneas, en las que las variaciones sean producidas exclusivamente por efectos climáticos [2, 3, citados por 1], se han desarrollado, principalmente en las últimas tres décadas, metodologías para la identificación de variaciones producidas por factores que introducen cambios abruptos en la media o en la tendencia de las series. Estas metodologías están orientadas a identificar heterogeneidades causadas por los cambios en la instrumentación, en los procesos de medición, en la localización de la estación, en los horarios de observación, en la forma de cálculo de valores medios y en el ambiente que rodea la estación [4]. Peterson et ál. [1] presentan dos tipos de metodologías para la detección de heterogeneidades, agrupando como directas aquellas

que utilizan técnicas gráficas y estadísticas para la detección de cambios, y como indirectas aquellas que se basan en la metadata histórica de las estaciones y en el estudio de los efectos de cambios de instrumentación.

Dentro de las metodologías directas se tiene el análisis de la metadata, en el cual la identificación de heterogeneidades parte de la recopilación y revisión del historial de la estación, que puede ser obtenido de documentos en los que se registran cambios asociados a factores que producen heterogeneidades o de la consulta a los operadores encargados del manejo de instrumentos y medición de variables. La ventaja de esta metodología consiste en la identificación de manera precisa del momento de la ruptura y su causa, aunque la ausencia de una metadata completa o la dificultad de convertirla en valores numéricos útiles para la corrección pueden limitar su uso [5]. Otra forma directa de detectar heterogeneidades se basa en la comparación de una serie medida con un instrumento que va a ser reemplazado y otra registrada con un nuevo instrumento de medición [1], en este caso, el limitante se puede presentar cuando el reemplazo de un instrumento sea a causa de que deje de funcionar, por lo cual no se contaría con una de las series para comparar.

Las metodologías indirectas se agrupan en subjetivas y objetivas. Las primeras se basan en la experiencia de quien aborda la identificación de las heterogeneidades; para esto el investigador se apoya en la gráfica de la serie que se va a estudiar, o en la gráfica de la diferencia entre la serie de estudio y una serie vecina, o mediante el análisis de la gráfica de las sumas acumuladas de la serie a ser evaluada contra las sumas acumuladas de una serie vecina [1]. En las objetivas se utiliza el cálculo de estadísticos para identificar las heterogeneidades, los cuales se pueden estimar utilizando solamente los datos de la serie (pruebas absolutas) o recurriendo a una o varias series vecinas para compararlas con la serie de estudio (pruebas relativas); para este último caso, la serie que se va a evaluar se conoce como serie candidata y las vecinas, como series de referencia; esta metodología es la más utilizada y la que se va a abordar con más detalle en el presente artículo.

Los procedimientos para la detección y corrección de heterogeneidades utilizando metodologías indirectas objetivas tienen como pasos: la revisión de la metadata, el control de calidad de las series (identificación de datos faltantes y valores dudosos), la construcción o identificación de series de referencia y la prueba de la homogeneidad y ajuste de las series [2, 6]. A continuación se describe cada uno de los pasos que se han de seguir dentro de estas metodologías aplicadas a series de datos mensuales y anuales.

## 2. Revisión de la metadata

La fuente de información más común para evaluar la homogeneidad de una serie de tiempo es la metadata, ya que en esta se consignan todos aquellos aspectos y cambios que pudieran afectar la serie y, en consecuencia, provee al investigador de información primaria para evaluar la calidad de la serie. Dentro del análisis de la metadata es importante identificar los cambios que pudieran afectar de manera simultánea varias estaciones a la vez, ya que este tipo de heterogeneidades puede no ser detectada durante la evaluación con técnicas estadísticas [1]. Esta primera fase de la metodología busca complementar el análisis estadístico al validar los resultados obtenidos de las pruebas aplicadas. Aunque el uso de la metadata no es un paso obligatorio, sí es recomendable, debido a la importancia de la información primaria que brinda.

## 3. Control de calidad de las series

El control de calidad de las series comprende la identificación y tratamiento de valores dudosos [2, 6] y el llenado de datos faltantes [7], aunque Stepanek [6] aborda este último problema de la serie al final del proceso de homogeneización. Los valores dudosos son aquellos que están bastante distantes de la media y que pueden ser causados por errores o por extremos climáticos [8, 9]; el principal problema de estos datos es que son elementos no representativos de la población, pudiendo distorsionar seriamente los resultados de las pruebas estadísticas empleadas para la detección de heterogeneidades [2, 9]. Los datos faltantes son otro problema común a las series de

datos y pueden ser causados, entre otras razones, por la pérdida de información y por la eliminación de datos clasificados como valores dudosos; algunas técnicas para la determinación de estos valores son: Ajustes por diferencias o razones [10, 11], Método de la razón normal [9, 10, 12], Prueba de Anderson [8], Complementación por razones de distancia [9], Análisis de regresión [10, 11, 13] y correlación de estaciones vecinas [9].

Para la detección de valores dudosos se utiliza generalmente el rango intercuartílico (IQR), definiendo uno o dos límites superiores e inferiores, sumando o restando un número determinado de veces el IQR a los percentiles 75 y 25, respectivamente. González-Rouco et ál. [8] y Saladie et ál. [14] calculan únicamente un límite superior al sumar tres veces el IQR al percentil 75. Para Mekis y Vincent [15] el número de veces que se suma y resta el IQR es de 1,5 para la identificación de valores dudosos en series mensuales de precipitación. Una forma más completa de identificar y clasificar los valores dudosos, descrita por Lobo [9] y Brubaker et ál. [16], define dos límites superiores y dos límites inferiores, correspondientes a sumar y restar 1,5 veces IQR (límites internos) y a sumar y restar el IQR 3,0 veces (límite externo) a los percentiles 75 y 25, respectivamente; en este caso los valores entre los límites interno y externo se clasifican como sospechosos, y los valores superiores al límite externo se clasifican como altamente sospechosos. Posterior a su identificación, se puede optar por eliminarlos y tratarlos como valores faltantes [6] o, como propone González-Rouco et ál. [8], reducir el tamaño de las colas de la distribución, por lo que los reemplaza por los valores calculados como límites; un análisis adicional, propuesto por Saladie et ál. [14], compara los valores dudosos detectados entre estaciones, y si se repiten en más de una, se consideran los valores como correctos.

## 4. Elaboración de la serie de referencia

La forma más común para generar una serie de referencia es calcular, para cada año, un promedio ponderado de los datos de las estaciones homogéneas vecinas, o de secciones de estas estaciones [2]. La

serie de referencia, en lo posible, debe haber sido afectada por todos los eventos climáticos que afectaron a la serie candidata, pero debe estar libre de cambios debidos a efectos no climáticos [1, 2]. La utilización de una serie de referencia para la detección de heterogeneidades puede estar sujeta a problemas como la falta de estaciones lo suficientemente cercanas a la serie candidata, la ausencia de series que tengan la longitud adecuada o posibles heterogeneidades que afecten en el mismo año todas las series de una región. En los caso en los que no sea posible contar con una serie de referencia, algunas técnicas para la detección de heterogeneidades pueden ser aplicadas únicamente a los datos de la serie en estudio, caso en el cual se deben tener precauciones adicionales por lo difícil que resulta distinguir entre los cambios causados por efectos naturales y aquellos producidos por causas no climáticas [2].

Un primer aspecto en la elaboración de una serie de referencia es asegurar de algún modo que las series vecinas empleadas para ello sean lo más homogéneas posible. Una forma de lograr esto es mediante la identificación de series homogéneas a partir del análisis de la metadata; otra forma es por la obtención de series de diferencias sucesivas, las cuales se calculan restando cada valor de la serie por el que lo precede ( $FD_i = T_{i+1} - T_i$ ), con lo que se logra que haya una mejor correlación entre la serie candidata y las vecinas al eliminar las posibles heterogeneidades [1, 17, 18]; otra forma de buscar la homogeneidad en las series de referencia es implementando un proceso iterativo, en el que primero se aplica la prueba estadística a todas las series, tomando como series de referencia el mismo grupo de series, para luego clasificarlas en homogéneas y heterogéneas; en pasos posteriores se corrigen las series heterogéneas y se realiza de nuevo la prueba de homogeneidad, en la que las series de referencia son las series corregidas junto con las clasificadas como homogéneas, y así hasta contar con series homogéneas confiables [19, 9].

En la selección de las series más adecuadas para conformar la serie de referencia, de acuerdo con su semejanza estadística con la serie candidata, usualmente se emplea el coeficiente de correlación

entre las series [2]. Además de la semejanza estadística, se deben tener en cuenta aspectos como la distancia entre estaciones, la elevación, el paisaje, elementos de la vegetación, la cantidad de datos faltantes [20] y barreras topográficas [1]. Stepanek [6] utiliza únicamente las estaciones con un promedio de un valor faltante por cada 1,5 años. En cuanto al número de estaciones por promediar, este puede variar de cuatro a seis [1, 6]. El coeficiente de correlación (o el de determinación), puede ser utilizado también como factor de ponderación al promediar las series escogidas [1, 8], aunque, de acuerdo con Stepanek [6], al utilizar coeficientes de correlación superiores a 0,9, el promedio simple frente al ponderado arroja resultados similares debido a la alta correlación entre las series. De otro lado, la utilización de valores de ponderación basados en la distancia entre estaciones puede ser inconveniente por diferencias climáticas importantes, por ejemplo, para el caso de estaciones cercanas y con una diferencia altitudinal importante [6]; además, Guijarro [22] concluye que la correlación entre estaciones en función de la distancia es bastante difusa.

## 5. Identificación y corrección de las heterogeneidades

Actualmente se cuenta con una gran variedad de técnicas para la detección de heterogeneidades en series temporales de variables hidroclimáticas, las cuales han sido recopiladas, descritas o comparadas por diferentes autores [1, 2, 22, 23, 24, 25]. En la tabla 1 se listan algunas de estas técnicas, junto con una referencia bibliográfica para su consulta, y programas de computador en los que se encuentran implementadas. La aplicación de las diferentes técnicas para la homogeneización de series producirá naturalmente diferentes resultados, por lo que decidir sobre cuál técnica, variante de la técnica o combinación de técnicas es mejor puede ser difícil y depende del objetivo planteado y de la disponibilidad de recursos [1]. Es por esto que diversos autores [6, 22, 26, 27, 28] han propuesto diferentes combinaciones de técnicas, junto con la aplicación de ciertos criterios de acuerdo con la variable de estudio, en algunas de las cuales es más importante

que en otras contar con la metadata de la estación. Una vez que la identificación del punto de cambio ha finalizado, el paso siguiente es decidir cuáles de los puntos identificados se consideran puntos de heterogeneidad, para lo cual inicialmente es recomendable identificar la causa de la ruptura en la metadata,

y cuando esto no es posible, la decisión debe apoyarse en la inspección visual de gráficas de diferencia para el caso de la temperatura, o división, en el caso de la precipitación, de series para evaluar la factibilidad de los puntos de ruptura identificados y para identificar posibles problemas no detectados [2].

**Tabla 1.** Técnicas para la homogeneización de series temporales

	<b>Técnica</b>	<b>Referencia</b>	<b>Programa</b>
1	Prueba ampliada de la Elipse	[29]	—
2	La regresión de dos fases	[18, 24, 30]	AnClim
3	Método de los residuales acumulados	[31]	—
4	Prueba de Craddock	[1, 2, 32]	AnClim
5	Curva de dobles masas	[11, 31, 33]	AnClim
6	Prueba de Caussinus-Mestre	[23, 34, 35]	—
7	Prueba de rangos de Buishand (Desviaciones acumuladas)	[2, 26, 36]	RAINBOW, AnClim
8	Prueba de Pettit	[2, 26]	HidroSIG
9	Prueba de Alexanderson - Standard Normal Homogenization Test (SNHT)	[1, 19, 36]	Climatol, AnClim
10	Von Neumann Ratio	[26, 36]	AnClim
11	Regresiones múltiples	[20]	AnClim
12	Prueba de Craddock	[2, 32]	—
13	Basados en la experiencia de los especialistas	[2, 4]	—
14	Multiple Analysis of Series for Homogenization (MASH)	[2, 37, 38]	MASH
15	Método de Potter	[1, 2, 39]	RHTestV2
16	Prueba máxima penalizada de F	[40]	RHTestV2
17	Prueba máxima penalizada de T	[41]	AnClim
25	Prueba F Modificada	[22]	HidroSIG
27	Prueba de Bartlett	[22]	HidroSIG
28	Prueba de Levene	[22]	HidroSIG
29	Análisis Bayesiano	[42]	—

La combinación de diferentes métodos ha sido la salida escogida por muchos autores para determinar las heterogeneidades en series históricas. Wijngaard et ál. [26] utilizan la combinación de cuatro métodos estadísticos absolutos (SNHT, la prueba de Rango de Buishand, la prueba de Pettit y la prueba de proporción de Von Newmann) para clasificar las series de acuerdo con la confiabilidad de su homogeneidad, para su posterior utilización en análisis de tendencias, de acuerdo con el número de pruebas que rechazan la hipótesis nula en un nivel

del 1% de confiabilidad; clasifican cada serie como “útil” cuando una o ninguna prueba la rechaza, se clasifica como “dudosa” cuando dos pruebas la rechazan, y como “sospechosa” cuando tres o cuatro pruebas la rechazan. Una ampliación de esta metodología se propone en Klok et ál. [43], en la cual se busca encontrar todos los segmentos útiles dentro de las series dudosas y sospechosas y los años de ruptura identificados con las pruebas estadísticas se utilizan para dividir las series en segmentos y estos segmentos son puestos a prueba, procedimiento que

se detiene si se llegan a series con menos de 10 registros. Miranda et ál. [28] combinaron el método SNHT, la prueba de rango de Buishand y la Von Neumann Ratio para detectar heterogeneidades sin utilizar serie de referencia, aceptando las heterogeneidades basados en su experiencia y en qué tanto estaban soportados los cambios por la metadata. González [13] solamente acepta la ruptura cuando dos pruebas realizadas (SNHT y la prueba de Earterling y Peterson) coinciden en señalarla en el mismo año; dicha ruptura es general en los distintos meses de la serie estudiada y no se produce en los 10 años iniciales ni finales (de una serie de 49 años), y además la ruptura se encuentra documentada. Rusticucci y Renom [27] combinan las pruebas SNHT, Regresiones múltiples y Buishand, concluyendo que la metadata resulta importante para aceptar las heterogeneidades, ya que no todos los cambios documentados fueron detectados por las técnicas estadísticas. Stepanek [6] considera que debido a la incertidumbre de los resultados de las pruebas de homogeneidad estadísticas es posible incrementar la confiabilidad de la determinación de heterogeneidades aplicando la mayor cantidad de pruebas a la serie; luego, mediante procesos estadísticos de un gran número de resultados de pruebas, se hace posible calcular la probabilidad de cada heterogeneidad para una serie dada (probabilidad calculada como la porción del número de heterogeneidades detectadas –para cada año, grupo de años o toda la serie– con relación a todas la detecciones teóricas posibles); para decidir sobre la heterogeneidad, un límite de 20% de todas las posibles detecciones ha sido utilizado en casos donde no hay información de metadata relacionada con el cambio; 10-15% fueron suficientes en casos donde las heterogeneidades estuvieron de acuerdo con la metadata. Boroneant y Tomozeiu [44] utilizaron los métodos SNHT y MASH para detectar las homogeneidades en series anuales y estacionales de temperatura media, encontrando que entre estos métodos no hay diferencias significativas en cuanto a la identificación de heterogeneidades.

En otros estudios no se recurre a la combinación de pruebas estadísticas, como la propuesta realizada por Vincent [20], en la que utiliza las regresiones

múltiples como técnica de detección de heterogeneidades y adopta una metodología conservadora, en la que los ajustes se realizaron únicamente si la magnitud del salto identificado en las series anuales de temperatura es mayor a 0,5 °C; los saltos menores a 0,5 °C fueron ajustados solamente cuando pueden ser verificados en la metadata. Gonzáles-Rouco [8], utilizando la prueba SNHT, clasifica como series de precipitación heterogéneas aquellas que: (i) contienen una heterogeneidad significativa en un nivel del 5% localizada a cinco años del inicio o del final de la serie, y (ii) la serie contiene una heterogeneidad significativa en un nivel del 10% cuando es explicada por la metadata; la interacción con la metadata en un nivel del 10% reduce la probabilidad de rechazar series homogéneas, incrementado así la capacidad de la prueba.

Algunos estudios han permitido identificar las fortalezas y debilidades de las técnicas para detección de heterogeneidades, lo que ha llevado al mejoramiento de estas en algunos casos. Dentro de estos estudios, el realizado por Beaulieu et ál. [24] sugiere la utilización de una técnica determinada de acuerdo con el tipo de heterogeneidad: para el caso de la verificación de series sugiere el método de Jaruskova, para la detección de saltos en la serie sugiere el “Approche bivariée”, para la detección de múltiples saltos sugiere el método Bayesiano y para la detección de tendencias sugiere las regresiones múltiples. Para el caso de la prueba SNHT, Hawkins [45, citado por 8] afirma que no se deben aceptar heterogeneidades que no estén sustentadas por la metadata y que se localicen al inicio o al final de la serie, debido al incremento de la probabilidad de tener valores altos de la prueba en estos sectores; esta prueba, que inicialmente fue diseñada para la detección de una ruptura, posteriormente y debido a esta debilidad, fue mejorada y cuenta con una versión que permite la detección de rupturas múltiples y de tendencias [1].

De manera general, ya sea que se utilice o no serie de referencia, o que se combinen diferentes técnicas de detección, es necesario tener en cuenta que si se aplica la prueba a las series mensuales se pueden

detectar diferentes puntos de cambio para cada una, lo cual es comprensible, debido a la aleatoriedad de las series y porque resulta obvio que algunas causas de heterogeneidades pueden tener un mayor impacto en una determinada época del año; por esta razón, resulta recomendable empezar con un análisis de los valores anuales que tienen una variación año tras año menor y generalmente permiten una mejor detección [2]. La aplicación del estadístico para valores anuales resulta necesaria, ya que la aplicación a cada mes por separado puede presentar inconsistencias relacionadas con la variación en el tamaño y localización de la heterogeneidad identificada [46].

En el ajuste de las heterogeneidades identificadas es siempre recomendable corregir los datos para hacerlos coincidir a la sección homogénea más reciente. Haciendo esto se asegura que los datos tomados a futuro sean homogéneos, siempre y cuando no se presenten cambios que causen la heterogeneidad de la serie otra vez [2, 47]. La forma más común para obtener los factores de corrección, para cambios en la media, es calcular la media en la parte inicial de la serie (la serie diferencia o división entre la candidata y la de referencia), antes de la ruptura, y la media con los datos de la parte final de la serie, después de la ruptura. Luego, las dos medias deben ser comparadas calculando su división o diferencia, y el factor de corrección obtenido es aplicado a la parte de la serie heterogénea. En el caso de heterogeneidades graduales, se elimina la tendencia empleando la pendiente calculada de la serie resultante de la comparación entre la serie candidata y la de referencia [2].

## **6. Programas de computador desarrollados para el análisis de series hidroclimáticas**

### **6.1 Hidrosig**

Desarrollado por el posgrado en Aprovechamiento de Recursos Hidráulicos, de la Universidad Nacional de Colombia, sede Medellín; combina funciones de los Sistemas de Información Geográfica con herramientas de visualización y modelación de variables y fenómenos hidrológicos, hidroclimatológicos, geomorfológicos, hidráulicos y

ambientales [48]. Es un programa de distribución gratuita, que se puede descargar, para Windows y Linux, de la página web <http://cancerbero.unalmed.edu.co/~hidrosig/index.php>.

En lo relacionado al análisis de series de tiempo de variables hidroclimáticas, cuenta con el módulo ANSET, que permite la verificación de la homogeneidad y la detección de valores dudosos, a través de diferentes técnicas [49]. Utiliza principalmente pruebas para la detección de cambios en la media y en la varianza, tendencia en la media e independencia. Las pruebas para detección de valores dudosos, con las que cuenta el programa, se dividen en dos grupos: las que requieren definir el número de valores dudosos (Prueba de Rosner, Prueba del estadístico Em, Prueba del estadístico Lm, Prueba de Dixon y Prueba de la distribución normal basada en la asimetría) y las que no requieren definir el número de valores dudosos (Prueba del Rango Normal, Prueba de Shapiro-Wilk, Prueba de Studentized-Deviates, Prueba de Grubbs y Prueba de Kandel y Last).

### **6.2 An Clim**

Para la homogeneización y el análisis de series de tiempo, Pert Stepanek creó el software AnClim [50] en el Departamento de Geografía de la Universidad de Masaryk; en él implementó los métodos más ampliamente utilizados para el análisis de series de datos climáticos (tabla 1). En AnClim están disponibles herramientas para la visualización de información y el cálculo de estadísticas básicas (incluidas pruebas de distribución normal y aleatoriedad), una variedad de pruebas de homogeneización relativa y absoluta, y correlación y análisis espectral [51]. La versión de uso libre de este programa, que puede ser descargada de <http://www.klimahom.com/software>, tiene algunas diferencias con la versión completa, en cuanto a que no permite el manejo de series diarias, el procesamiento automático de series es limitado y no cuenta con soportes en cuanto a su manejo y actualización.

Las pruebas para homogeneidad, con las que dispone

el programa, se presentan divididas en dos menús: uno para pruebas sin serie de referencia y otro para la utilización de serie de referencia. El programa cuenta con una función para la construcción de una serie de referencia, para lo cual se puede escoger entre promedio simple, promedio simple de desviaciones, promedio ponderado para diferencias y promedio ponderado para divisiones. Para el llenado de datos faltantes utiliza la serie de referencia, calculando el promedio de la serie candidata y de la de referencia y reemplazando los datos faltantes por el valor de la serie de referencia sumado al promedio de la serie candidata y restando el promedio de la serie de referencia (para series de temperatura); otra opción es calcular el dato faltante multiplicando el valor de la serie de referencia por el promedio de la serie candidata y dividiendo el resultado por el promedio de la serie de referencia (para series de precipitación). Para la detección de valores dudosos, el programa utiliza el rango intercuartílico.

### 6.3 RHtest

Programa desarrollado por Xiaolan Wang y Feng Yang, en la División de Investigación de Clima y Ambiente de Canadá; permite detectar y ajustar múltiples puntos de cambio de la media en una serie que puede tener errores autorregresivos de primer orden [52]. El programa utiliza la prueba t máxima penalizada y la prueba F máxima penalizada. Este programa puede ser descargado de <http://ccma.seos.uvic.ca/ETCCDMI/software.shtml> y requiere para su funcionamiento el programa estadístico R, este último también gratuito y disponible en <http://www.r-project.org/>.

### 6.4 Climatol

Fue desarrollado en el Centro Meteorológico en Illes Balears, Instituto Nacional de Meteorología de España, y requiere del programa estadístico R. Las herramientas disponibles incluyen la inspección gráfica de los datos y sus anomalías, detección de saltos en la media, detección de tendencias y la construcción de series de referencia [21]. Para la construcción de la serie de referencia utiliza el promedio ponderado de todas las series disponibles,

donde el valor de ponderación depende de la distancia entre estaciones. Para la detección de heterogeneidades cuenta con herramientas de graficación para análisis visuales. Para el llenado de datos faltantes utiliza el método descrito por Paulhus y Kohler [12], el cual se basa en la interpolación espacial de los datos de precipitación expresados en forma de proporciones respecto a la media de las estaciones. El programa es de uso libre, disponible en <http://webs.ono.com/climatol/climatol.html> y en <http://cran.r-project.org/web/packages/climatol/index.html>.

### 6.5 Rainbow

Programa de uso libre (<http://www.iupware.be>), diseñado para estudiar registros meteorológicos e hidrológicos a través de análisis de frecuencias y para evaluar la homogeneidad de los registros. Para detectar heterogeneidades en la serie utiliza la prueba de las desviaciones acumuladas [53], también conocida como prueba de Rangos de Buishand. Este programa no permite el llenado de datos faltantes ni la detección de valores dudosos.

## 7. Conclusiones y comentarios

Las metodologías indirectas objetivas para la homogeneización de series de datos hidrológicos y meteorológicos se han convertido en herramientas indispensable en estudios que involucren el análisis de series históricas. En estas se cuenta con una gran cantidad de experiencias, en las cuales se aplican las técnicas estadísticas a series de diferentes lugares del mundo, que permiten contar con una metodología bien estructurada para la verificación y mejoramiento de la calidad de las series. Las pruebas estadísticas han sido implementadas en programas de computador de distribución gratuita, lo cual abre las puertas a la determinación de las técnicas más apropiadas para regiones con características meteorológicas específicas. Estas metodologías permiten la obtención de series de datos históricos con la calidad adecuada para estudios hídricos y climáticos, teniendo siempre en mente que la gran variedad de técnicas hace necesario considerar como parte de cualquier estudio la identificación de la más adecuada



para las particularidades de la zona de interés; aspecto que es de gran importancia si se tiene en cuenta la diversidad de climas que se presentan en Colombia.

La flexibilidad con la que cuentan las metodologías descritas tiene como inconveniente que dificulta la comparación detallada entre estudios realizados en diferentes países, y es por esto que a través de la COST (European Cooperation in the Field of Scientific and Technical Research) se ha emprendido la acción titulada “Action. COST-ES0601: Advances in homogenization methods of climate series: an integrated approach”, que busca producir métodos estandarizados que faciliten la comparación y la divulgación de experiencias de homogeneización. Dado que en este proyecto tal vez se realicen estudios de caso con series de datos de otras regiones diferentes a Colombia, sus resultados deben ser vistos con cuidado, y se deben, al igual que con las metodologías descritas en este artículo, poner a prueba, considerando las condiciones particulares propias de nuestras regiones.

## Referencias

- [1] T. Peterson, et ál. «Homogeneity adjustments or in situ atmospheric climate data: A review». *International Journal of Climatology*. Vol. 181, 493-1517, 1998.
- [2] E. Aguilar, I. Auger, M. Brunet, T. Peterson, J. Wieringa. *Guidelines on climate metadata and homogenization*. Geneva: World Meteorological Organization, 2003. 52p.
- [3] V. Conrad and C. Pollak. *Methods in Climatology*. Cambridge: Harvard University Press MA, 1950. 459 pp.
- [4] P. D. Jones, S. C. Raper, R. S. Bradley et ál. «Northern Hemisphere Surface Air Temperature Variations: 1851-1984». *Journal of Climate and Applied Meteorology*. Vol. 25, p. 161-179, 1986.
- [5] M. Brunetti, M. Maugeri, F. Monti and T. Nanni. «Temperature and Precipitation Variability in Italy in the Last Two Centuries from Homogenized Instrumental Time Series». *International Journal of Climatology*. Vol. 26, p. 345-381, 2006.
- [6] P. Stepanek. «Homogenization of Air Temperature en the Czech Republic During a Period of Instrumental Measurements». In: *Proceedings of the 4rd Seminar for Homogenization*. WCDMP No. 56, WMO-TD No. 1236. Geneve, sept. 2004.
- [7] S. Feng, Q. Hu and W. Qian. «Quality Control of Daily Meteorological Data in China, 1951-2000: A New Dataset». *International Journal of Climatology*. Vol. 24, p. 853-870, 2004.
- [8] F. González-Rouco, L. Jiménez, V. Quesada y F. Valero. «Quality Control and Homogeneity of Precipitation Data in the Southwest of Europe». *Journal of Climate*. Vol. 14, p. 964-978, 2001.
- [9] D. Lobo. *Guía metodológica para la elaboración del mapa de zonas áridas, semiáridas y subhúmedas secas de América Latina y el Caribe*. 2005. 66 p. [En línea, disponible en [http://www.cazalac.org/documentos/Guia\\_Mapas\\_ZA\\_ALC.pdf](http://www.cazalac.org/documentos/Guia_Mapas_ZA_ALC.pdf)].
- [10] R. Alfaro and R. Pacheco. «Aplicación de algunos métodos de relleno de series anuales de lluvia de diferentes regiones de Costa Rica». *Top. Meteor. Oceanogr*. Vol. 7, p. 1-20, 2000.
- [11] World Meteorological Organization. *Guide for Climatological Practices*. WMO-No. 100. Geneva: WMO: 1983. 2 ed. 182 p.
- [12] J. L. Paulhus and M. A. Kohler. «Interpolation of missing precipitation records». *Monthly Weather Review*. Vol. 80, p. 129-133, 1952.
- [13] C. Gonzáles, M. De Luis Martín, P. Stepanek, J. Raventos y M. Cuadrat. *Reconstrucción, estabilidad y proceso de homogeneizado de series de precipitación en ambientes de elevada variabilidad pluvial*. VII Reunión de Clima de la Asociación de Geógrafos españoles. Albarracín, 2002.
- [14] O. Saladie, E. Brunet, E. Aguilar et ál. «Evolución de la precipitación en el sector suroriental de la depresión del Ebro durante la segunda mitad del siglo XX». En: III Congreso de la Asociación Española de Climatología. Palma de Mallorca, 16 jun. 2002.
- [15] E. Mekis, L. Vincent. «Homogeneity Assessment of Canadian Precipitation Data for Joined Stations». In: 13th Conference on Applied Climatology. Portland, 13 may. 2002.

- [16] K. Brubaker, P. Kirmeyer, A. Sudradjat et ál. «A 36-yr Climatological Description of the Evaporative Source of Warm-Season Precipitation in the Mississippi River Basin». *Journal of Hydrometeorology*. Vol. 2, p. 537-557, 2001.
- [17] R. Tomozeiu, S. Stefan and A. Busuioc. «Winter Precipitation Variability and Large-Scale Circulation Patterns in Romania». *Theoretical and Applied Climatology*. Vol. 81, p. 193-201, 2005.
- [18] D. Easterling, T. Peterson and T. Karl. «On the Development and use of Homogenized Climate Dataset». *Journal of Climate*. Vol. 9, p. 1429-1434, 1995.
- [19] I. Hansen-Buer and E. Forland. «Homogenizing Long Norwegian Precipitation Series». *Journal of Climate and Applied Meteorology*. Vol. 7, p. 1001-1013, 1994.
- [20] L. Vincent. «A Technique for the Identification of Inhomogeneities in Canadian Temperature Series». *Journal of Climate and Applied Meteorology*. Vol. 11, p. 1094-1104, 1998.
- [21] J. Guijarro. «CLIMATOL: software libre para la depuración y homogeneización de datos climáticos». En: *El clima entre el mar y la montaña*. Serie A N.º 4, pp. 493-502, 2004.
- [22] J. C. Correa, R. Iral y L. Rojas. «Estudio de potencia de pruebas de homogeneidad de varianza». *Revista Colombiana de Estadística*. Vol. 29, pp. 57-76, 2006.
- [23] O. Mestre. *Méthodes Statistiques pour L'homogénéisation de Longues Séries Climatiques*. Thèse pour obtenir le grade de Doctorat Mathématiques appliqués - statistiques, Université Paul Sabatier, 2000. 230 p.
- [24] C. Beaulieu, et ál. *Étude Comparative de Méthodes D'homogénéisation des Précipitations*. Quebec: OURANOS: 2006. 96 p.
- [25] J. Reeves, J. Chen, X. Wang, R. Lund and Q. Lu. «A Review and Comparison of Change-point Detection Techniques for Climate Data». *Journal of Applied Meteorology and Climatology*. Vol. 46, p. 900-916, 2007.
- [26] J. Wijnngaard, A. Klein and G. Können. «Homogeneity of 20th Century European Daily Temperature and Precipitation Series». *International Journal of Climatology*. Vol. 23, p. 679-692, 2003.
- [27] M. Rusticucci and M. Renom, Madeleine. «Homogeneity and Quality Control of Long Time Series of Daily Temperature in Uruguay». In: *The 15th Conference on Applied Climatology*. Savannah, 25 jun. 2005.
- [28] P. Miranda, F. Coelho, A. Tome, M. Valente et ál. «20th Century Portuguese Climate and Climate Scenarios». In: *Climate Change in Portugal: Scenarios, Impacts and Applications Measures*. p. 23-83, 2002.
- [29] C. Costa and A. Soares. «Identification of Inhomogeneities in Precipitation Time Series Using SUR Models and the Ellipse Test». In: *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*. Lisbon, 2006.
- [30] A. Solow. «Testing for Climate Change: An Application of the Two-Phase Regression Model». *Journal of Climate and Applied Meteorology*. Vol. 26, p. 1401-1405, 1987.
- [31] R. Allen et ál. *Evotranspiración del cultivo: Guía para la determinación de requerimientos de agua de los cultivos*. Roma: Organización de las Naciones Unidas para la agricultura y la alimentación: 2006. 298 p.
- [32] M. Maugeri. «Our Approach to homogenization». In: *Training Session on Homogenization Method*. Bologna, 17-18/05/2005.
- [33] J. Searcy y C. Hardison. *Curva de dobles masas*. Santa Fe de Bogotá: Himat: 1983. 36 p.
- [34] V. Alexandrov, M. Schneider, E. Koleva and J. M. Moisselin. «Climate Variability and Change in Bulgaria During the 20th Century». *Theoretical and Applied Climatology*. Vol. 79, p. 133-149, 2004.
- [35] V. Alexandrov, B. Dubuisson, J. M. Moisselin and E. Koleva. «A Case Study on Utilization of Precipitation Indices in Bulgaria». In: *Proceedings of the Conference on Water Observation and Information System for Decision Support*. Ohrid, 23 may. 2006.

- [36] J. Wijngaard and A. Klein. «Homogeneity of the ECA Temperature Data». In: *Proceedings of the 3rd Seminar for Homogenization*. Budapest, 25 sept. 2000.
- [37] T. Szentimrey. «Multiple Analysis of Series for Homogenization (MASH). Seasonal Application of MASH (SAM), Automating use of Metadata». In: *Proceedings of the 3rd Seminar for Homogenization*, 25 sept. 2000.
- [38] T. Szentimrey. «Multiple Analysis of Series for Homogenization (MASH): Verification Procedure for Homogenized Time Series». In: *Proceedings of the 4rd Seminar for Homogenization*. WCDMP No. 56, WMO-TD No. 1236. Geneve, sept. 2004.
- [39] K. Potter. «Illustration of a New Test for Detecting a Shift in Mean in Precipitation Series». *Monthly Weather Review*. Vol. 109, p. 2040-2045, 1981.
- [40] X. Wang. «Penalized Maximal F Test for Detecting Undocumented Mean Shift without Trend Change». *Journal of Atmospheric and Oceanic Technology*. Vol. 25, p. 368-382, 2008.
- [41] X. Wang, Q. Wen and Y. Wu. «Penalized Maximal t Test for Detecting Undocumented Mean Change in Climate Data Series». *Journal of Applied Meteorology and Climatology*. Vol. 46, p. 916-930, 2007.
- [42] P. S. Lucio, F. C. Conde and A. M. Ramos. «A Bayesian Approach for Recovering and Homogenizing Meteorological Time Series». In: *Proceedings of the 8 ICSHMO*, 24 April 2006.
- [43] L. Klok, A. Klein, M. Begert and E. Zenclusenn. «Assessment of the Available Station Density for the Gridding and Daily Data Quality/Homogeneity». *ENSEMBLES project report*, 2006. 36 p. Available in: <http://eca.knmi.nl/publications/index.php>.
- [44] C. Boroneant and R. Tomozeiu. «On the Homogenization of Long Series of Annual and Seasonal Temperature Means in Romania». In: *Proceedings of the 3rd Seminar for Homogenization*. Budapest, 25 sept. 2000.
- [45] P. M. Hawkins. «Testing a Sequence of Observations for a Shift in Location». *J. Amer. Stat. Assoc.* Vol. 72, p. 180-186, 1977.
- [46] G. Müller-Westermier. «Statistical Analysis of Results of Homogeneity Testing and Homogenization of Long Climatological Time Series in Germany». In: *Proceedings of the 4rd Seminar for Homogenization*. WCDMP No. 56, WMO-TD No. 1236. Geneve, sept. 2004.
- [47] R. Heinkelmann, J. Boehm and H. Schuh. «Homogenization of Surface Pressure Recordings and its Impact on Long-Term Series of VLBI Tropospheric Parameters». In: *Proceedings of the 17th Working Meeting on European VLBI for Geodesy and Astrometry*. Noto, Italy, 22 April, 2005.
- [48] J. Vélez, G. Poveda, O. Mesa et ál. «HidroSig: Un sistema de información geográfica para la gestión de recursos naturales y modelamiento del medio ambiente». *I Jornadas de SIG libre*. Girona, 5 mar. 2007.
- [49] Escuela de Geociencias y Medio Ambiente. *HidroSig, Versión 3.0 Beta: Manual de Usuario*. Medellín: Universidad Nacional de Colombia, Sede Medellín: 2003. 109 p.
- [50] P. Stepanek. *AnClim: Software for Time Series Analysis*. Dept. of Geography, Fac. of Sciences, Masaryk University: 2006.
- [51] P. Stepanek, R. Brazdil and V. Kveton. «Homogenization of Air Temperature Series in the Czech Republic». In: *Proceedings of the 3rd Seminar for Homogenization*. Budapest, 25 sept. 2000.
- [52] X. Wang and Y. Feng. *RHtestV2: User Manual*. Ontario: Science and Technology Branch, Environment Canada: 2007. 19 p.
- [53] D. Raes, P. Willems and F. Gbaguidi. «RAINBOW: A Software Package for Analyzing Data and Testing Homogeneity of Historical Data Sets». In: *Proceedings of the 4th International Workshop on "Sustainable Management of Marginal Drylands"*. Islamabad, 27 jan. 2006.

Fecha de recepción: 6 de octubre de 2007  
Fecha de aprobación: 30 de mayo de 2008