

## CONTROL DE CALIDAD Y PROCESO DE HOMOGENEIZACIÓN DE SERIES TÉRMICAS CATALANAS

E. AGUILAR\*, J. M. LÓPEZ\*\*, M. BRUNET\*,  
O. SALADIÉ\*, X. SIGRÓ\* y D. LÓPEZ\*

\* *Grupo de Cambio Climático. Unidad de Geografía. U.R.V. Tarragona*

\*\* *Dept. d'Enginyeria Informàtica i Matemàtiques. U.R.V. Tarragona*

### RESUMEN

La falta de homogeneidad de las series climáticas dificulta la detección y caracterización del cambio climático a diferentes escalas espaciales, por lo que previamente a su utilización con esta finalidad, deben ser sometidas a un proceso riguroso de control de calidad y de homogeneización. En esta comunicación se aborda la aplicación de la prueba SNHT de Alexandersson y Moberg (1997) a los registros de la temperatura del aire de la región catalana. La finalidad de esta aplicación es obtener una base de datos térmicos homogeneizados.

**Palabras clave:** Cambio climático, temperaturas, homogeneización, SNHT, Cataluña.

### ABSTRACT

Lack of homogeneity in climatic data series is an obstacle to detecting and characterising climatic change in different space scales; especially if inhomogeneities and actual change are expected to be figures of the same order of magnitude. Hence, before data can be used to determine the existence of this climatic change and its magnitude, a rigorous process of quality control and homogenisation must be followed. In this communication, Alexandersson and Moberg's test (SNHT) is used, with some modifications, with the object of obtaining a set of adjusted data for Catalonia.

**Key words:** Climate change, temperature, homogenisation, SNHT, Catalonia.

### INTRODUCCIÓN

La creación de bases de datos climáticos, sometidos a un proceso objetivo de control de calidad y homogeneización de los registros, constituye el paso previo para el estudio del cambio climático a cualquier escala espacial. Para que una serie temporal represente las variaciones de un elemento climático y pueda ser calificada como homogénea es necesario establecer que dichas variaciones respondan sólo a causas climáticas y no a otras de carácter artificial o "no climático" (CONRAD y POLLACK, 1962).

La falta de homogeneidad de los registros obedece a causas diversas. Entre otras, hay que destacar las que introducen discontinuidades abruptas en las series, como las asociadas a las relocalizaciones de los observatorios, cambios en las prácticas observacionales y en el cálculo de las medias, etc. o bien graduales, generalmente relacionadas con cambios en el entorno (efecto urbano, cambios de uso del suelo, deforestación o reforestación, etc.) (MITCHEL, 1953).

En la bibliografía especializada se recoge un número abundante de pruebas para llevar a cabo el test de homogeneización de una serie temporal, mediante aproximaciones absolutas o relativas, según requieran o no de la existencia previa de series de probada homogeneidad.

ALEXANDERSSON (1986) puso a punto un método aplicado a las precipitaciones, mediante el cual se corregían los saltos detectados en las series candidatas. Posteriormente fue rediseñado para corregir no sólo discontinuidades abruptas, sino también para eliminar las tendencias lineales incorporadas en los registros (ALEXANDERSSON y MOBERG, 1997; MOBERG y ALEXANDERSSON, 1997; MOBERG y BERGSTROM 1997).

Esta última metodología se ha aplicado al banco de datos históricos de Cataluña, con el objetivo de obtener una base de datos homogeneizada de la temperatura del aire, que sirva posteriormente para establecer la estructura temporal y espacial de la deriva térmica a esta escala.

En este trabajo se expone el método utilizado en el control de calidad de los registros térmicos y se discute la aplicación del test de homogeneidad de Alexandersson, con las modificaciones introducidas por los autores, tendentes a conseguir una mayor eficacia en el binomio coste/tiempo, con vistas al objetivo final: la creación de una base de datos ajustada. En consecuencia, primero se exponen los procedimientos de control de calidad de los datos observacionales, para abordar a continuación la aplicación del test utilizado y finalmente dar cuenta de las modificaciones establecidas en el mismo.

## **1. PROCESO DE CONTROL DE CALIDAD Y SELECCIÓN DE LA RED TÉRMICA CATALANA**

En Cataluña existen registros desde mediados del siglo XIX, aunque la red catalana presenta grandes variaciones temporales en el nivel de cobertura espacial de la región y de continuidad temporal de los registros, como sucede por otra parte en el conjunto del territorio español. Otros hechos han intervenido en la representatividad de la red catalana: cabría citar, por ejemplo, el crecimiento urbano a partir de los años sesenta, la desaparición de estaciones rurales relacionada con embalses, la reforestación de antiguos espacios agrícolas o la ampliación de la red de observatorios de los años setenta. Éstos y otros aspectos han quedado incorporados en la metadata de los observatorios utilizados en el proceso de homogeneización y control de calidad que se resumen a continuación.

### **1.1. Selección de estaciones**

El trabajo se inició con la obtención de los datos de temperaturas máximas y mínimas disponibles en formato digital en los archivos del Instituto Nacional de Meteorología. A partir de esta información se han calculado las temperaturas medias, con el fin de evitar el sesgo introducido por cambios en el procedimiento de cálculo, así como la amplitud térmica diaria.

Una vez examinada la información disponible, se procedió a la selección de las estaciones a utilizar siguiendo dos criterios: a) Inicio en fecha anterior a 1946. b) Existencia de un 5% como máximo de datos mensuales perdidos o ausentes en el periodo 1961-80.

Se contempló también la posibilidad de componer series entre observatorios cercanos, para maximizar la información antigua disponible, siempre que cumpliesen los criterios exigidos de proximidad geográfica y homogeneidad climática. Se dio un solo caso, el observatorio de Manresa que se compuso con el de Sallent (Barcelona).

Este proceso de selección proporcionó inicialmente un total de 17 estaciones, que fueron sometidas a un control de calidad similar al planteado por los científicos del NCDC (National Climate Data Center, del NOAA) para la base de datos GHCN (Global Historical Climate Network) (PETERSON *et al.*, 1997).

### 1.2. Control de calidad

El control de calidad se llevó a cabo en tres etapas sucesivas:

a) Etiquetado de datos sospechosos o ausentes. Se detectaron todas las lagunas existentes en la información de base, a escala mensual y una vez almacenada la información correspondiente a cada mes y observatorio, se llevó a cabo una aproximación climatográfica, con el fin de contrastar visualmente la información de cada uno de ellos con los dos más cercanos espacialmente. Todos los datos sospechosos fueron marcados en los ficheros originales. A continuación se procedió a un nuevo proceso de test estadístico mediante el cual se identificaron los *outliers*, marcando como tales todos los datos situados fuera del intervalo  $\pm 4\sigma$  respecto de la media.

b) Recuperación de datos etiquetados. Todos los datos etiquetados -ausentes o discrepantes-, fueron comprobados en los archivos originales, para recuperar al máximo posible la información. Pudieron ser sustituidos un 50% de los datos, al presentar problemas derivados básicamente del proceso de digitalización. Completada esta fase, se volvió a aplicar el proceso de etiquetado, hasta que la última información marcada correspondía a datos irrecuperables.

c) Relleno de lagunas originales o generadas en el etiquetado. Dado que el método de homogeneización a aplicar requiere la existencia de un registro continuo, se procedió al relleno de lagunas. Se calculó la correlación entre todas las estaciones existentes en base mensual y al encontrar un dato ausente, se le asignó un valor estimado a partir de las restantes estaciones, obtenido mediante el mismo procedimiento que proporciona los valores de referencia para el test. Se adoptaron como referencia todas las estaciones que presentaban una correlación  $> 0.5$ . En la práctica, la correlación utilizada alcanzó cifras superiores al límite inferior fijado, puesto que sólo el 10% de los valores se situó entre el 0.5 y el 0.6.

Este proceso permitió crear el Banco de Datos Sin Ajustar, que constituye el punto de partida para la fase de homogeneización.

## 2. APLICACIÓN DEL TEST AL PROCESO DE HOMOGENEIZACIÓN DE LAS SERIES TÉRMICAS CATALANAS

Para la creación del Banco de Datos Ajustado de las temperaturas máximas, mínimas, medias y amplitud térmica diaria, se ha aplicado el *Standard Normal Homogeneity Test for Shift Detection (SNHTs)* (ALEXANDERSSON y MOBERG, 1997) a la información mensual, estacional y anual. El método fue implementado con ciertas modificaciones menores, que se comprobó no alteran el resultado final, pero agilizan su consecución.

### 2.1. Consideraciones generales

El método se basa en la aplicación de un proceso iterativo; en cada paso se distingue entre una serie candidata (de la que se buscan inhomogeneidades) y un conjunto de series que actúan como referencias. Se parte de la base de que no se conoce de antemano la calidad de ninguna de las series, por lo que no se presupone -como ocurre con otras técnicas- la homogeneidad de ningún observatorio; así, todas las series tienen la consideración de posibles referencias y todas son en algún paso candidatas a ser homogeneizadas. Alexandersson y Moberg parten de una premisa simple, pero efectiva: aunque ninguna serie fuera homogénea, es muy poco probable que todas lo fueran en el mismo punto. En consecuencia, si la comparación de cualquier candidata con un número suficiente de referencias se lleva a cabo de una forma estadísticamente correcta, debe poner de manifiesto los cambios súbitos o graduales que pueden aparecer en la misma.

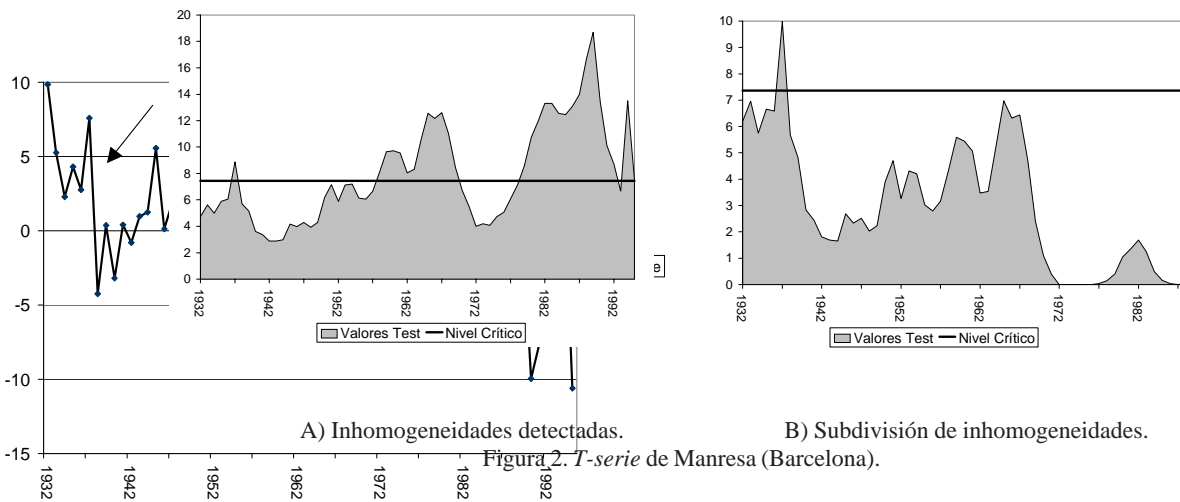
### 2.2. Implementación del test

El método de homogeneización de temperaturas se basa en el análisis estadístico de la llamada *Q-serie*, o serie de los valores diferencia entre valor observado y valor de referencia, obtenido éste mediante ponderación de datos correspondientes a las estaciones de referencia elegidas. La representación gráfica de esta serie suele dar una idea cualitativa de la existencia de uno o varios puntos de cambio.

Dada una *Q-serie*, se normaliza para obtener una *Z-serie*. Ésta servirá para la determinación estadística del cambio o cambios existentes, así como su magnitud y significación, mediante el contraste de hipótesis entre la hipótesis nula ( $Z$  es una variable aleatoria de distribución normal) y la alternativa. Se realiza mediante un proceso de subdivisión sucesiva, similar al descrito en trabajos anteriores (PETERSON y EASTERLING, 1994). La Fig. 1 ejemplifica esta fase del proceso.

Para cada valor temporal intermedio, se plantea la hipótesis alternativa siguiente: la variable aleatoria  $Z$  es normal de varianza 1 y media  $m_1$  hasta ese momento, y varianza 1 y media  $m_2$  a partir del dato siguiente ( $m_1$  y  $m_2$  son los promedios respectivos); se calcula el estadístico de prueba ( $T$ ) correspondiente. Con ello se crea una nueva serie temporal (*T-serie*), que responde a la evaluación de la diferencia entre cada uno de los conjuntos de semiseries generados, que se representan gráficamente utilizando como abscisa el punto de cambio. Si el valor máximo en la *T-serie* supera el nivel crítico en algún punto se concluye que la serie es inhomogénea (Fig. 2A).

Figura 1. Inhomogeneidades detectadas en la Z-serie de Manresa (Barcelona).



Dada la naturaleza de la hipótesis alternativa, se aprecia que la existencia de más de una inhomogeneidad puede provocar que el estadístico  $T$  no supere en ningún punto los valores críticos. De ahí la importancia de la serie temporal  $T$ , por cuanto la existencia de varias inhomogeneidades se refleja en la aparición de varios picos. Mediante la observación de esta serie, puede repetirse el test para intervalos temporales más cortos utilizando como límite aquel punto en el que previamente se ha detectado una inhomogeneidad. Si en ellos se cumple la hipótesis alternativa (un único salto), ello quedará reflejado en un alto valor de  $T$ , y en una mejor estimación de la magnitud de ese salto (Fig. 2B).

### 2.3. Modificaciones al proceso original

El método se ha implementado en un programa, a partir del desarrollo de un conjunto de aplicaciones elaboradas en Fortran-77, que reproducen el proceso iterativo de Alexandersson y Moberg (1997) con las siguientes modificaciones para hacerlo más operativo:

a) Los coeficientes de correlación fueron obtenidos utilizando la serie primera diferencia de las medias anuales, obtenida de los datos mensuales disponibles, sin alteración de ningún valor durante el proceso para todos los parámetros.

El método original recalcula a cada nueva corrección el coeficiente de correlación y utiliza, para cada parámetro, sus propios coeficientes.

b) Para cada candidata fueron seleccionadas ocho estaciones de referencia, según los criterios expuestos por los autores, en función de las más altas correlaciones, similitud climática y proximidad geográfica. Las estaciones preseleccionadas fueron incorporadas al conjunto de referencias sólo en el caso de que el coeficiente de correlación entre cada una de ellas y la candidata fuera  $\cong 0.5$ . En la práctica, las correlaciones obtenidas fueron muy superiores a este umbral: menos del 10 % de las estaciones utilizadas figuraban con valores próximos a este valor límite.

c) Se realizó un orden aleatorio de corrección y cada estación corregida se incorporó inmediatamente al conjunto de estaciones de referencia, sustituyendo la versión anterior del mismo observatorio. El concepto original de sesión se reemplazó por un procedimiento continuo y más rápido. Tras comprobar que los resultados obtenidos a través de ambas aproximaciones eran plenamente homologables, continúa el proceso, que finaliza cuando se consigue la homogeneidad, según el SNHTs.

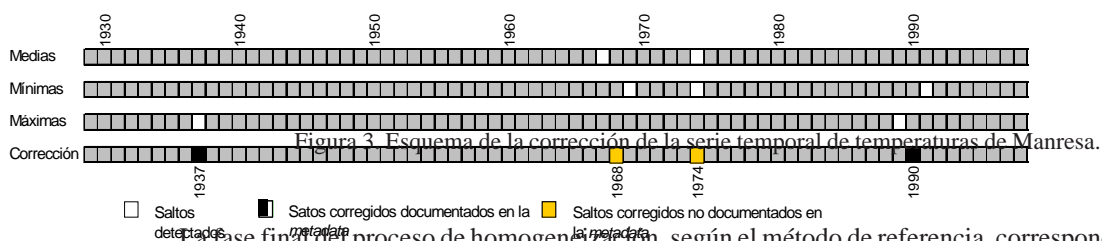
d) Se distinguen dos fases: Preliminar o preparación de las series y final o corrección. Las series obtenidas como resultado de la fase preliminar serán las que servirán como referencias en la final. En la ronda final se tomaron como candidatas las series del Banco de Datos sin Ajustar (es decir, datos originales más rellenos). Esta ronda se realizó una vez, y en ella, a diferencia de la fase preliminar, las series corregidas en esta fase no se incorporan al conjunto de referencias. El objetivo de esta etapa corresponde al que se indica en el método original: asegurar la bondad de la corrección utilizando como referencia series cuya cantidad de errores sea mínima. En este punto no es necesaria la iteración, puesto que las series de referencia no se modifican en ningún momento.

e) El proceso de homogeneización se ha dividido en dos partes: detección y corrección. Los posibles saltos han sido detectados a partir de la aplicación del test al promedio anual de las temperaturas máximas, mínimas y medias diarias. A partir de estos tres parámetros se diseñó un esquema de corrección, que se aplicó sistemáticamente a todos los valores estacionales y mensuales.

f) Se incorporaron a la base de datos cinco estaciones que inician su información en los primeros años del siglo XX y que corresponden a regiones vecinas, como referencias auxiliares, con la finalidad de completar el número insuficiente de referencias en la homogeneización de las cuatro estaciones catalanas cuyas series se remontan a principios de siglo. Estas cinco estaciones sólo se utilizaron en los períodos para los que se disponía de menos de cuatro referencias para cualquiera de las estaciones a homogeneizar. Naturalmente, la información correspondiente a estas cinco estaciones, aunque no se incorporó por razones obvias al banco de datos, participó en el proceso como candidatas y como referencias.

Todas estas modificaciones introducidas al método original perseguían como objetivo conseguir un proceso menos costoso e igualmente eficiente, aplicable a otros trabajos de homogeneización. Se realizaron numerosas pruebas previamente a su definición y pudo comprobarse la escasa o nula influencia en los resultados obtenidos con estos cambios.

El proceso de detección de inhomogeneidades fue diseñado de forma que permitía aislar las simples, es decir, las que presentaban una sola incidencia. Se aplicó el SNHTs a cada observatorio y a cada uno de los parámetros citados, comparándolo con las estaciones de referencia preseleccionadas. Cuando se detectaba un punto de cambio, se dividía la serie temporal en dos partes y se aplicaba de nuevo el test a cada una de ellas. El proceso de partición continuaba hasta que no aparecían nuevas inhomogeneidades o se llegaba a una serie parcial de una longitud inferior a 11 años, señalizando los puntos de cambio detectados. La detección de los tres parámetros anuales permite dibujar el esquema final de la corrección (Fig. 3).



La fase final del proceso de homogeneización, según el método de referencia, corresponde a la fase de corrección. Se aplicó de manera sistemática a todas las series estacionales y mensuales, mediante el procedimiento de corrección “de adelante hacia atrás” (EASTERLING y PETERSON, 1995). Ello permite estimar cada salto, calibrando su magnitud con datos que no presentan ninguna otra inhomogeneidad.

De esta forma, se corrigió en primer lugar la irregularidad más cercana al presente, estimado su efecto con los datos comprendidos entre el año siguiente al penúltimo punto de cambio y el último valor disponible. Se procede de idéntica forma hasta corregir la totalidad de la serie, que de esta forma queda homogénea.

Se aplicó el doble proceso detección-corrección para cada uno de los observatorios tantas veces como fue necesario hasta conseguir completar la fase final y una vez adicional para cumplir con la fase final. Una vez finalizado este punto, se consideró que las series eran homogéneas y en consecuencia, se había creado el Banco de Datos Ajustado. La Fig. 4A ilustra el proceso de corrección. En la figura 4B puede observarse la diferencia entre la homogeneizada, respecto de la que no lo está.

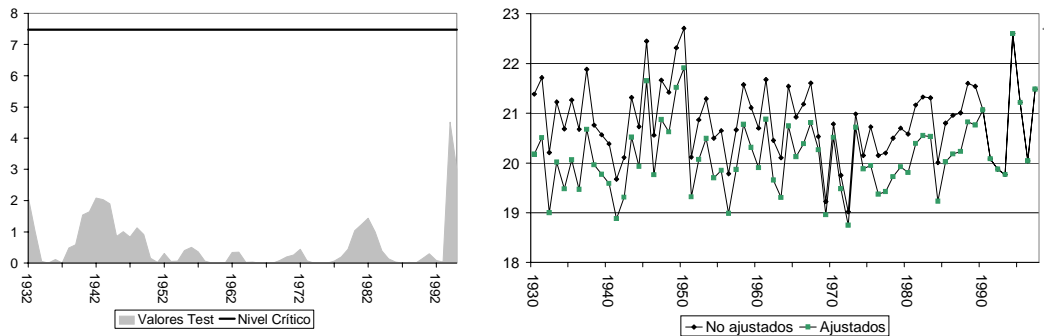


Figura 4. A) Serie térmica corregida.  
B) Diferencias entre las series homogeneizada y sin homogeneizar. Manresa (Barcelona).

## CONCLUSIONES

La aplicación del método modificado de Alexanderson y Moberg mediante la implementación de un programa adecuado permite realizar la corrección de las inhomogeneidades y crear una base de datos ajustados, que permiten obtener resultados en la detección de la evolución del clima a cualquier escala. El método, modificado para lograr una mayor agilidad y rapidez en el proceso de corrección de saltos, ha demostrado su bondad reiteradamente y permite llevar a cabo de manera práctica la corrección de inhomogeneidades atribuibles a causas ajenas a la propia dinámica climática.

## BIBLIOGRAFÍA

- ALEXANDERSSON, H (1986): "A homogeneity test applied to precipitation data", *Journal of Climate*, **6**, 661-675.
- ALEXANDERSSON, H. y A. MOBERG, (1997): "Homogenization of Swedish Temperature Data. Part I: Homogeneity Test for Linear Trends", *Int. Journal of Climate*, **17**, 25-34.
- CONRAD, V. y L. D. POLLACK (1962): *Methods in Climatology*, Cambridge: Harvard University Press, 459 pp.
- EASTERLING, D.R. y T.C. PETERSON (1995): "A new method for detecting undocumented discontinuities in climatological time series", *Int. Journal of Climatology*, **15**, 369-377.
- MITCHELL, J. M. (1953): "On the causes of instrumentally observed secular temperature trends", *Journal of Meteorology*, **10**, 244-261.
- MOBERG, A. y H. ALEXANDERSSON (1997): "Homogenization of Swedish Temperature Data. Part II: Homogenized Gridded Air Temperature compared with a subset of global gridded air temperature since 1861", *Int. Journal of Climatology*, **17**, 35-54.



MOBERG, A. y H. BERGSTROM (1997): "Homogeneization of Swedish Temperature Data. Part III: The long temperature records from Upsala and Stockholm", *Int. Journal of Climatology*, **17**, 667-699.

PETERSON, T.C. y D. R. EASTERLING (1994): "Creation of homogeneous composite climatological reference series", *Int. Journal of Climatology*, **14**, 671-679.

PETERSON, Th. C., R. VOSE, R. SCHMOYER, y V. RAZUVAEV (1997): *Quality control of monthly temperature data: The GHCN experience*, <http://www.ncdc.noaa.gov/ol/climate/research/ghcn/ghcnqc.html>, 14 pp.

#### **AGRADECIMIENTOS**

Este trabajo ha sido financiado por CICYT, Proyecto de Investigación de I+D CLI96-1842-C05-01. Agradecemos al I.N.M. por la cesión de la base de datos y especialmente al personal del Centro Meteorológico de Barcelona, Sr. Lara (Director) y Sara Satué (responsable Sección de Climatología), por su colaboración en el proceso de recuperación de datos. También, agradecemos la participación de P. Chana durante el tratamiento de los datos.