

# **CLIMATOL: SOFTWARE LIBRE PARA LA DEPURACIÓN Y HOMOGENEIZACIÓN DE DATOS CLIMATOLÓGICOS**

José A. GUIJARRO PASTOR

*Centro Meteorológico en Illes Balears, Instituto Nacional de Meteorología*

## **RESUMEN**

Se presenta un conjunto de rutinas de depuración y homogeneización de datos climatológicos denominado *CLIMATOL*, desarrollado para su uso dentro del paquete estadístico R que, al ser libre y multiplataforma, permite su uso en una gran variedad de entornos de trabajo. Las funcionalidades implementadas incluyen la comparación de cada serie original con una de referencia, estimada a partir de un promedio de las demás (modulado mediante una función inversa de la distancia y previa tipificación de los datos de cada estación), así como la inspección gráfica de los datos y sus anomalías. Los errores puntuales, saltos en las medias y tendencias se pueden detectar en las series de diferencias entre los datos originales y los estimados mediante las pruebas paramétricas habituales.

**Palabras clave:** Homogeneización, series climáticas, errores, software.

## **ABSTRACT**

*A set of computer routines for error correction and homogenization of climatological data (named CLIMATOL) has been developed to be used in the frame of the free and multi-platform statistical package R, thereby allowing its use in a wide variety of working environments. Implemented functionalities include the comparison of each original series with a reference one (computed as a weighted average of normalized values of the other series), and the visual inspection of graphs of the data and their anomalies. Point errors, average shifts and trends may be detected in the series of differences between the original and reference data through standard parametric tests.*

**Key words:** Homogenization, climatic series, errors, software.

## **1. INTRODUCCIÓN**

Los errores son inherentes a toda actividad de medida, y la Meteorología no es una excepción. Por ello, la preocupación por la detección y corrección de errores e inhomogeneidades en las series climatológicas es tan antigua como la Climatología misma.

Antes del advenimiento de las tecnologías de la información, el estudio climatológico de las series de medidas meteorológicas se realizaba manualmente, resultando muy útiles para la detección de inhomogeneidades los métodos gráficos, como el de las dobles masas (JANSÀ, 1969), y muchas veces el objetivo se limitaba a la obtención de promedios referidos a un periodo común de observaciones, para lo que bastaban procedimientos como los de las diferencias o las proporciones (CONRAD y POLLACK, 1962).

Actualmente los ordenadores permiten un tratamiento masivo de grandes volúmenes de datos, con lo que es posible abordar la reconstrucción de series climatológicas completas mediante estima de los datos ausentes en los registros. Pero los ordenadores precisan ser instruidos por los

investigadores sobre la manera en que han de realizar dichas tareas, mediante la codificación de los programas adecuados. Esta labor suele llevarse a cabo de forma independiente por los diferentes equipos de trabajo, a pesar de utilizar frecuentemente algoritmos similares, siendo muy limitada la difusión de paquetes de programas (*software*) pre-elaborados para estos menesteres.

De hecho, en el momento de escribir estas líneas, este autor únicamente conoce los casos de los paquetes *MASH* (SZENTIMREY, 1998) y *AnClim* (STÌPÁNEK *et al.*, 2000; STÌPÁNEK, 2003), que presentan el inconveniente de tener restringido su funcionamiento a plataformas (ordenadores personales -PC's-) y sistemas operativos concretos (DOS y Windows respectivamente).

En cambio, las técnicas discutidas en este trabajo se han desarrollado mediante el paquete estadístico R, que es multiplataforma (hay versiones para otras máquinas además de para los PC's) y funciona con distintos sistemas operativos (GNU-Linux, Solaris, Windows, etc), lo que permite su uso en una amplia gama de entornos de trabajo. Además, y no menos importante, su código es abierto y modificable (bajo la licencia GPL), con lo que otros investigadores pueden adaptarlo a sus preferencias y, si lo desean, poner sus contribuciones a disposición de los demás. Cuentan para ello con la enorme variedad de pruebas estadísticas y representaciones gráficas del propio paquete R, y de un sencillo pero potente lenguaje de programación para implementar nuevas funcionalidades.

En los apartados siguientes se pasará revista a los tipos de errores y otras inhomogeneidades de las series climatológicas, y se discutirán las técnicas implementadas en las rutinas objeto de este trabajo para su detección y eventual corrección.

## 2. METODOLOGÍA

### 2.1. Planteamiento

Cuando hablemos de homogeneidad de una serie climatológica entenderemos, a la manera de CONRAD y POLLACK (1962), que nos referimos a homogeneidad relativa, esto es, en relación a las variaciones del elemento climático en la zona donde se ubica el observatorio, puesto que en este tipo de series no cabe esperar homogeneidad absoluta en sentido puramente estadístico (o dejarían de tener sentido los trabajos sobre el cambio climático).

El estudio de la homogeneidad de una serie implica por tanto la comparación de la misma con una serie de referencia. Para ello se podría emplear una serie próxima de reconocida calidad y homogeneidad pero, como a priori es muy aventurado hacer esas suposiciones, resulta más conveniente usar una serie artificial, combinando varias series reales, con objeto de diluir sus propias posibles inhomogeneidades.

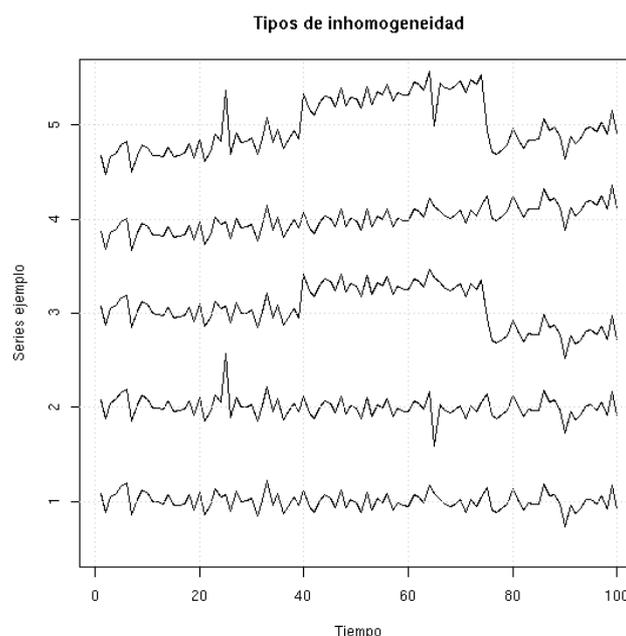
Una vez construida la serie de referencia, la comparación con el problema se puede realizar estudiando la serie de cocientes o diferencias entre ambas. Los cocientes presentan problemas cuando la serie *divisora* tiene valores nulos o próximos a cero, como sucede al aplicar el test de Alexandersson en su forma primitiva (ALEXANDERSSON, 1986), de modo que aquí se usarán las diferencias entre las series a comparar, como hacen otros autores (AGUILAR *et al.*, 1999, por ejemplo).

Esta serie de diferencias debe estar ya libre de las oscilaciones que el elemento climático estudiado haya podido sufrir en la zona de estudio, así como de las características de persistencia y estacionalidad propias de las series climatológicas. Por tanto, en el caso ideal, esta serie debería comportarse como un *ruido blanco*, pudiendo considerarse como inhomogeneidad a cualquier anomalía que se aparte del mismo. Estas anomalías pueden agruparse en tres tipos:

1. Errores aislados, debidos a errores de lectura o transcripción, o a errores puntuales en el sensor (estaciones automáticas). Se pueden detectar como datos anómalos aislados (*outliers* en la terminología inglesa).
2. Saltos en las series, atribuibles a cambios de instrumentación o de las condiciones de instalación de los aparatos. Se pueden detectar por cambios en los promedios anteriores y posteriores a los saltos (mediante pruebas de la  $t$ , con el método de Alexandersson ya citado, etc.).
3. Derivas graduales en las series, debidos a cambios progresivos en la respuesta de los sensores (descalibración) o en el entorno (crecimiento urbano, cambios en los usos del suelo, etc.). Detectables mediante análisis de tendencias o de desviaciones acumuladas.

En la figura 1 se ilustran estos tres tipos de inhomogeneidades. La serie 1 simula una serie aleatoria con distribución normal (ruido blanco), como sería el caso de una serie de diferencias entre dos series climáticas homogéneas de observatorios próximos. En la serie 2 se han simulado dos errores puntuales: +5s en el término 25, y -5s en el 65. La serie 3 presenta dos cambios bruscos de la media: +3s en el término 40, y -2s en el 75 (-5s respecto a los términos anteriores, 40 a 74). La serie 4 simula una tendencia desde -2s hasta +2s a lo largo de los 100 términos de la serie. Por último, en la serie 5 se han incluido todas las inhomogeneidades anteriores, observándose un notable enmascaramiento mutuo que dificulta su diagnóstico y corrección, como lamentablemente cabe esperar en las series climatológicas reales.

Fig. 1. Tipos de inhomogeneidad esperables en las series climatológicas:  
 1) Serie homogénea;  
 2) Errores puntuales;  
 3) Saltos en la media;  
 4) Tendencias;  
 5) Todas las anteriores



## 2.2. Construcción de las series de referencia

Las series de referencia suelen obtenerse a partir de una o, más frecuentemente, varias series de estaciones próximas o que estén bien correlacionadas con la serie del observatorio problema. En este caso, la serie de referencia se obtiene como promedio de los valores de las series de los observatorios escogidos, utilizando algún método de ponderación para dar mayor peso a los más próximos y/o mejor correlacionados.

Es lógico pensar que la proximidad geográfica vaya ligada con una mayor correlación, pero la figura 2 nos indica que esa relación es bastante difusa. Así, mientras que es cierto (al menos en este ejemplo) que las correlaciones más altas corresponden siempre a estaciones próximas, para una distancia de unos 25 km encontramos coeficientes de correlación muy variados, que oscilan entre 0,65 y 0,90. El coeficiente de correlación permite entonces discriminar observatorios que, aún estando a distancias similares, responden de forma más o menos afín a las variaciones del elemento climático en el observatorio estudiado por estar situado en un terreno fisiográficamente más similar, y es por eso por lo que su uso como criterio de ponderación goza de una gran popularidad.

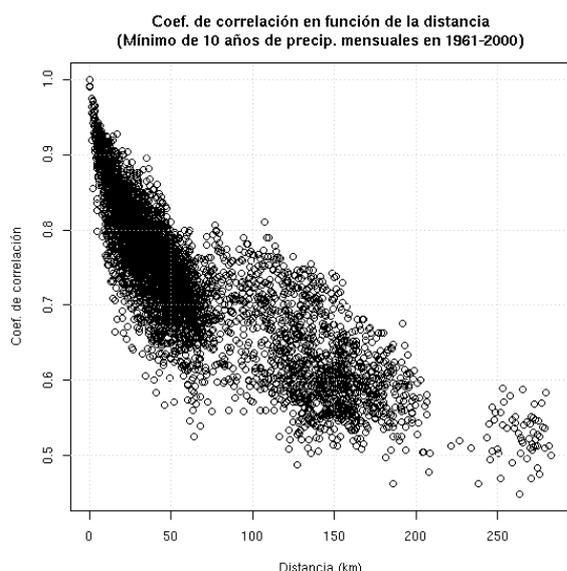


Fig. 2. Coeficientes de correlación en función de la distancia entre las series de precipitación mensual de las estaciones de Baleares que han funcionado un mínimo de 10 años en el periodo 1961-2000

Sin embargo, además de la conocida falta de robustez frente a datos anómalos o con series de valores que no se distribuyan normalmente (como sucede con los datos de precipitación), el coeficiente de correlación presenta el inconveniente de necesitar para su aplicación que las series hayan tenido un periodo común de funcionamiento, cuanto más largo mejor. Esto impide el uso de series cortas, así como el de observatorios muy próximos pero con escasa o nula simultaneidad, como por ejemplo sucede cuando un observatorio cambia su emplazamiento a una ubicación muy próxima (alrededor de 1 km), cosa que es frecuente encontrar en el historial de los observatorios con series más largas. En estos casos, si consideramos una serie única, encontraremos inhomogeneidades en los momentos de cambio de emplazamiento, y si consideramos las series como pertenecientes a distintos observatorios, no podremos usar sus datos como referencia entre ellos mismos.

Por estos motivos se ha preferido en este trabajo utilizar una metodología más sencilla, utilizada hace ya más de 50 años por PAULHUS y KOHLER (1952) para el relleno de datos de precipitación diaria, basada en la interpolación espacial de los datos de precipitación, pero expresados en forma de proporciones respecto a la media de la estación.

A pesar de su simplicidad, esta técnica ofrece buenos resultados, como puso de manifiesto YOUNG (1992) al comparar este método con el de interpolación óptima de GANDIN (1963) y con el de análisis discriminante múltiple (técnica multivariante similar a las funciones empíricas ortogonales, componentes principales, etc.). Este último método fue el que arrojó menores errores en las estimas de los datos mensuales de precipitación, pero el método de las proporciones, aunque produjo errores algo mayores (pero menores que la interpolación óptima), fue el único que obtuvo estimas insesgadas, y también el que menos redujo la varianza de los datos estimados respecto a los originales.

Minimizar el sesgo de las estimas nos permite reducir los errores de los promedios, mientras que una menor reducción de la varianza resulta de gran importancia si se van a utilizar las series reconstruidas para el cálculo de periodos de retorno u otras aplicaciones altamente sensibles a este estadístico.

Un inconveniente de este método es que, sobre todo en el caso de las series más cortas, la precipitación media del observatorio, calculada con los datos disponibles, podrá diferir considerablemente de la del periodo de referencia (normalmente todo el periodo de la base de datos a homogeneizar). Esto se puede obviar mediante un proceso iterativo en el que, una vez obtenidas las series de referencia de cada observatorio, se substituyen los datos ausentes de las series originales por los datos estimados (deshaciendo la tipificación que se haya llevado a cabo previamente), se recalculan los promedios y, caso de que alguno de ellos haya variado más que una cantidad dada, se repite la estima de las series (a partir de las originales), procedimiento que se utilizó en el relleno de datos ausentes en las series de precipitación mensual de Baleares en un trabajo anterior (GUIJARRO, 2002).

Para la obtención de series de referencia termométricas el método de las proporciones no parece el más indicado, y se ofrece entonces la posibilidad de centrar los datos (restándoles su respectiva media mensual) o estandarizarlos (dividiendo los datos centrados por su desviación típica).

En cuanto a la selección de observatorios de referencia, se optó por usarlos todos, ponderando su influencia por una función inversa de la distancia. Esta elección supone mayor carga de cálculo para el ordenador que si se fija un radio máximo de influencia, pero nos libera de tener que elegir ese radio, necesariamente variable según la distribución y densidad de la red de observación. No obstante, es necesario decidir hasta qué punto una estación próxima debe pesar más que otras un poco (o mucho) más alejadas. En este punto hay que tener presente que las estaciones próximas pueden tener errores, por lo que, al menos en una primera fase del estudio de homogeneidad, conviene que en la serie a estimar intervengan datos de varias estaciones con un peso similar, en lugar de sobreponderar mucho la estación más próxima. Como quiera que de este modo estaremos disminuyendo la varianza de las estimas, en función de la finalidad del estudio puede ser conveniente, como se ha comentado antes, minimizar esta disminución repitiendo posteriormente el proceso con una función de ponderación que prime mucho más la influencia de las estaciones próximas respecto a las demás.

Para conseguir estos objetivos los pesos de cada observatorio se han calculado mediante la función  $a/(a+d^2)$ , donde  $d$  es la distancia entre la estación problema y la de referencia, y ' $a$ ' es un parámetro de forma. La distancia aparece elevada al cuadrado porque, dado que se calcula aquí mediante el teorema de Pitágoras (el efecto de la curvatura terrestre no influirá mucho en los resultados) aplicado a las coordenadas UTM de las estaciones, nos ahorraremos el cálculo de la raíz cuadrada de la hipotenusa. La influencia del parámetro de forma se puede ver en la figura 3, en la que se representa la función de ponderación en su forma *canónica*  $1/(1+d^2/a)$ , obtenida al dividir por ' $a$ ', con lo que sus valores estarán acotados en el intervalo  $(0,1]$  en lugar de en el  $(0,a]$ . Valores de  $a=1.000$  o superiores diluirán los errores de las estaciones de referencia al considerar gran número de ellas con pesos del mismo orden, mientras que con  $a=10$  o inferior obtendremos estimas con mayor ajuste a la Climatología local de la estación problema y con poca disminución de varianza.

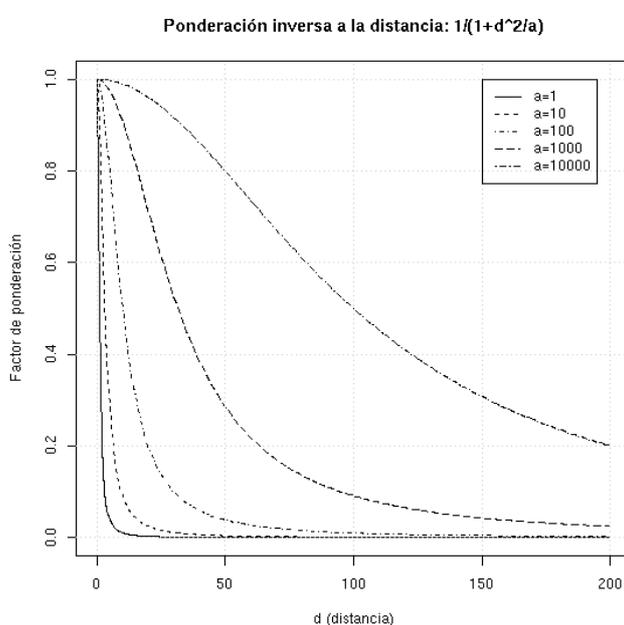


Fig. 3. Ponderación en función de la distancia, para distintos valores del parámetro de forma ' $a$ '

### 2.3. Estudio de la homogeneidad de las series

Una vez construidas las series de referencia (series de datos estimados para cada observatorio) por el procedimiento anterior, la homogeneidad de cada serie original se puede contrastar mediante comparación con su serie estimada. Para ello se estudian las diferencias entre ambas series que, en el caso ideal, deberían semejar a una serie aleatoria de distribución normal.

Esto puede hacerse de modo interactivo, visualizando gráficamente estas diferencias en forma de anomalías tipificadas, como se observa en la figura 4, en la que se muestran los gráficos de temperaturas máximas medias del mes de enero de: a) una estación homogénea; b) una estación con una anomalía sospechosa y una clara tendencia, tal vez debida al crecimiento urbano; c) una estación con cambios bruscos en la media y elevada varianza; d) visualización conjunta de los datos originales y los estimados.

Decidir a partir de qué umbral una anomalía puede considerarse errónea y debe corregirse no es fácil. La hipótesis de que las series de diferencias deben tener una distribución normal permite utilizar diversas pruebas paramétricas como herramientas de decisión, pero siempre existirán casos en que tomemos como erróneos datos ciertos, mientras que no corregiremos los errores de pequeña magnitud. En el caso de los errores puntuales, lo ideal es estudiar las series originales de datos: temperaturas extremas diarias, por ejemplo, en lugar de temperaturas medias calculadas a partir de ellas. (Piénsese que un error de lectura de 5 o incluso 10 grados en una temperatura extrema sólo alterará el promedio mensual en unas pocas décimas. Los demás tipos de inhomogeneidad se detectarán mejor con las series mensuales (o incluso estacionales o anuales), libres de la varianza de los datos diarios. En el caso del error puntual de la figura 4B se comprobó que los datos diarios eran correctos, y sin embargo el promedio mensual grabado era erróneo: 13,7°C en lugar de 15,8. Como el programa de depuración estimaba como dato más probable 15,75, en este caso su funcionamiento demostró ser más que aceptable.

Una vez determinadas las inhomogeneidades, su corrección puede ser bastante laboriosa. Lo ideal sería poder recurrir a los datos originales de cada observatorio y a su historial (metadatos), de forma que se pudiera contrastar si ha habido un error de transcripción (en el caso de errores puntuales) o si hubo algún cambio de instrumentación o de emplazamiento que justifique un salto en la media. Pero muchas veces el acceso a los datos originales es complicado, por no hablar de la falta de exhaustividad de los historiales de las estaciones, cuando existen. Si, a pesar de todo, se tiene la posibilidad y la voluntad de abordar esta tarea, hay que ser conscientes de la enorme carga de trabajo que implica, aún en bases de datos de moderado tamaño.

Por eso, sin desdeñar la conveniencia de realizar esas tareas cuando sea posible, muchas veces la necesidad de obtener resultados en un plazo relativamente corto nos obligará a adoptar una estrategia más drástica, como es el suprimir los datos considerados erróneos y substituirlos por los datos estimados. En el caso de saltos en la media, se procederá a fragmentar las series como si pertenecieran a estaciones diferentes. Sobre este punto hay que señalar que muchos autores se inclinan por considerar la subserie más moderna como la correcta, y ajustarán a ella los promedios de las subseries anteriores. Pero, en ausencia de un historial de la estación que documente prácticas de observación erróneas o no ajustadas a las normas actuales (como la observación en terrazas y/o con garitas termométricas metálicas de principios del siglo XX), parece más prudente el considerar las subseries como distintos muestreos climatológicos de la misma zona; su cartografía pondrá de manifiesto cuál de ellas parece más representativa y cuáles estarían afectadas de defectos de observación o instalación.

En cualquier caso, el trabajo de homogeneizar una base de datos climatológicos debería abordarse en varias etapas, filtrando primero únicamente los errores más grandes, y ajustando después progresivamente los umbrales de rechazo para detectar los demás. De este modo evitaremos eliminar datos correctos que aparecían como erróneos por influencia de errores grandes en alguna estación vecina.

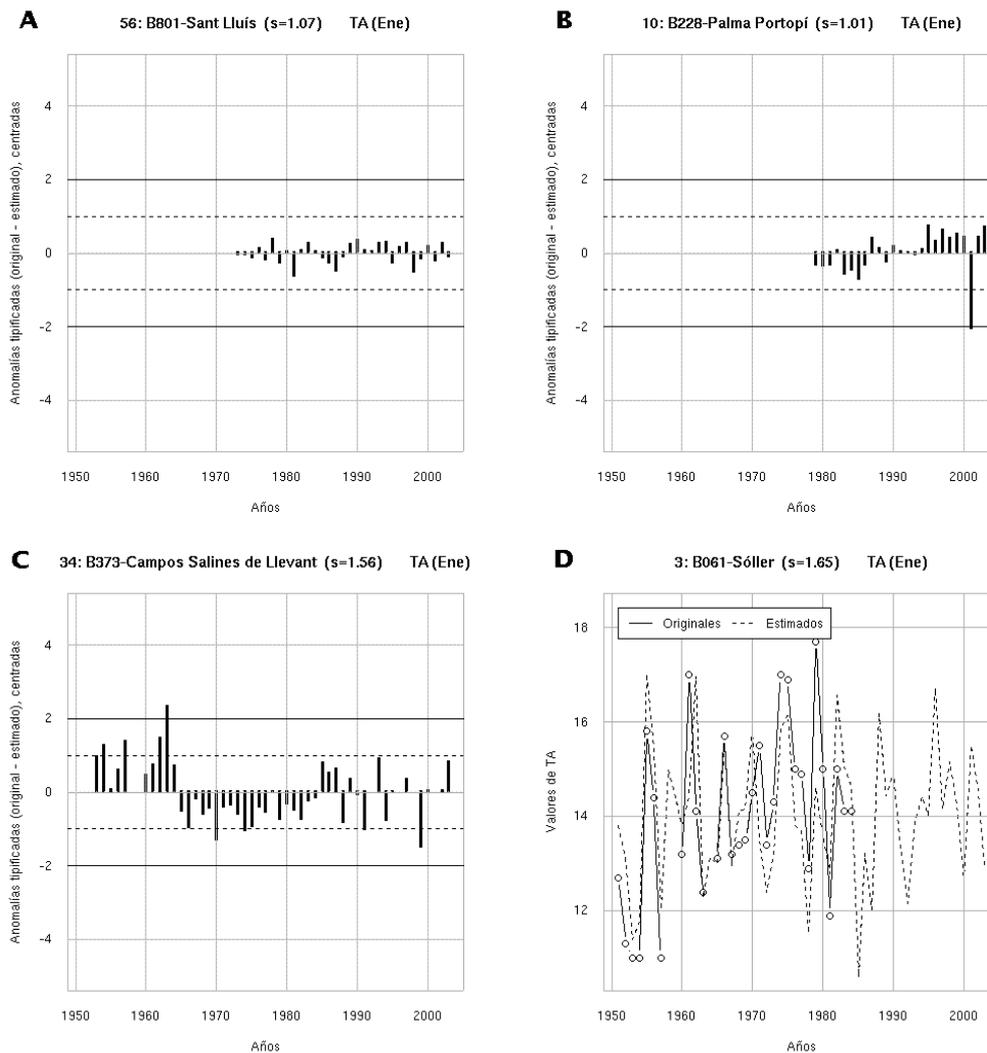


Fig. 4. Estudio de homogeneidad de las temperaturas máximas medias del mes de enero en Baleares. a) Estación homogénea; b) Error puntual y tendencia; c) Cambios bruscos en la media; d) Visualización de los datos originales y los estimados

### 3. USO DE *CLIMATOL*

Si no somos ya usuarios de R, deberemos empezar por obtener el paquete de instalación de <http://cran.r-project.org/> o de alguna de las réplicas repartidas por diferentes servidores. Descargada e instalada la versión adecuada para nuestra plataforma y sistema operativo, conviene introducirse en el programa leyendo al menos el primer capítulo de la introducción de VENABLES *et al.* (2004; existen traducciones al español). Para consultar las funciones disponibles, así como su sintaxis y opciones, se dispone de una excelente ayuda en formato HTML. Para usar las rutinas de *CLIMATOL* descargaremos la última versión de <http://home.worldonline.es/jaguijar/> (el nombre del archivo comprimido de la versión 1.0 es *climatol\_1.0-es.tgz*) y lo instalaremos con la orden 'R CMD INSTALL climatol\_1.0-es.tgz' si trabajamos en Unix, o 'Rcmd INSTALL climatol\_1.0-es.tgz' si lo hacemos en Windows.

Los datos mensuales a estudiar deben introducirse en un archivo de texto de nombre VAR\_AI-AF.dat, y las coordenadas y nombres de las estaciones en VAR\_AI-AF.est, siendo VAR cualquier acrónimo de la variable climatológica que contenga, y AI y AF las dos últimas cifras del año inicial y final abarcado por los datos. Dentro del archivo los datos se ordenan cronológicamente, estación por estación, especificando los datos ausentes con "NA".

En el archivo de estaciones figurarán los datos (cada estación en una línea) 'X Y Z ID NOMBRE', siendo X e Y las coordenadas UTM en km, Z la altitud en m, ID un identificador de la estación, y por último su NOMBRE, que deberá ir entrecorriado si contiene caracteres en blanco. Las coordenadas se pueden introducir en otras unidades, pero habrá que tener en cuenta su posible influencia en el cálculo de la distancia (si se dan en grados de latitud y longitud, por ejemplo). La coordenada vertical también se tiene en cuenta para el cálculo de las distancias entre observatorios, previa transformación interna a km; si las unidades no son las especificadas, habrá que ajustar su ponderación respecto a las coordenadas horizontales. También se puede sobreponderar su influencia intencionadamente, lo que puede ser de interés en el tratamiento de temperaturas mínimas, por el distinto comportamiento térmico de las estaciones de distintas altitudes si son frecuentes las inversiones térmicas nocturnas en la zona.

Arrancado el programa R, comenzaremos por cargar las rutinas de *CLIMATOL* con:  
> library(climatol)

A continuación podemos iniciar el estudio de homogeneidad de los datos mensuales de precipitación (hemos llamado PTOT a la variable) entre 1951 y 2003, tipificados por el método de las proporciones (ttip=2), e indicando que deseamos ver las representaciones gráficas de los datos (graf=TRUE):  
> depudm("PTOT", 1951, 2003, ttip=2, graf=TRUE)

Cuando la precipitación es muy irregular, puede ser conveniente aplicar una transformación raíz cuadrada para que la distribución de las anomalías se aproxime más a la normal. La función de depuración tiene prevista esa opción (sqrtrans=TRUE). Si alguno de los meses es habitualmente muy seco, aunque ocasionalmente pueda presentar valores altos (caso del mes de julio en muchas zonas mediterráneas), el promedio mensual será muy inestable en las estaciones con registros cortos. Es preferible entonces usar el promedio anual de precipitación para la tipificación por el método de las proporciones (opción refglob=TRUE).

Otra posibilidad sería realizar una depuración automática (auto=TRUE) de datos de temperaturas máximas medias mensuales (variable TA) del mismo periodo 1951-2003, eliminando las anomalías superiores a 2,5 desviaciones típicas (dz.max=2.5):  
> depudm("TA", 1951, 2003, auto=TRUE, dz.max=2.5)

Esto generará un archivo de datos depurados y con sus lagunas rellenadas, de nombre TA\_51-03.dep, del que podemos obtener los datos medios del treintenio internacional 1961-90 con la orden:  
> depstat("TA", 1951, 2003, 1961, 1990)

Los resultados se escribirán en el archivo TA\_51-03.med. Durante el proceso se sobre-escribe cualquier fichero que exista previamente con el mismo nombre en el directorio de trabajo, por lo que habrá que renombrarlos si queremos conservar los resultados de procesos anteriores (si queremos probar distintos umbrales de rechazo o diferentes tipificaciones, por ejemplo).

El proceso automático únicamente rellenará las lagunas de datos y corregirá los errores puntuales. Por el momento no se contempla la corrección automática de saltos en la media y tendencias, pero los resultados de una prueba móvil de diferencia de medias y un test de tendencias, para cada serie, quedarán reflejados en un "cuaderno de bitácora" de nombre `climatol.log`.

En estos ejemplos no se ha especificado el valor del parámetro de forma de la función de ponderación, con lo que se usa un valor por defecto de  $w_a=100$ . Para mayor información sobre las opciones de las funciones, consultar las páginas de ayuda del propio paquete.

#### 4. CONSIDERACIÓN FINAL

Seguramente los potenciales usuarios de *CLIMATOL* echarán de menos muchas otras funcionalidades, algunas de las cuales pueden ir incorporándose con el tiempo. No obstante, las series generadas por estos programas quedan residentes en el espacio de trabajo de R mientras no se cierra la sesión (o no se borran explícitamente), por lo que es posible seguir trabajando con ellas y aplicarles la amplia variedad de tratamientos estadísticos que proporciona R, así como generar diversas representaciones gráficas.

#### 5. REFERENCIAS

- AGUILAR, E. *et al.* (1999). Control de calidad y proceso de homogeneización de series térmicas catalanas. En RASO, J.M. y MARTÍN-VIDE, J. (Eds.). *La Climatología española en los albores del siglo XXI*. AEC, Serie A, 1, pp. 15-23.
- ALEXANDERSSON, H. (1986). "A homogeneity test applied to precipitation data". *J. Climatol.*, 6, pp. 661-675.
- CONRAD, V. y POLLACK, L.W. (1962). *Methods in Climatology*. Harvard Univ. Press, Cambridge-Massachusetts.
- GANDIN, L.S. (1983). *Objective analysis of meteorological fields*. Gidrometeorologicheskoe Izdatel'stvo, Leningrado, 242 pp.
- GUIJARRO, J.A. (2002). Tendencias de la precipitación en el litoral mediterráneo español. En GUIJARRO, J.A. *et al.* (Eds.). *El agua y el clima*. AEC, Serie A, 3, pp. 237-246.
- JANSÀ, J.M. (1969). *Curso de climatología*. SMN B-19, 445 pp, Madrid.
- PAULHUS, J.L.H. y KOHLER, M.A. (1952). "Interpolation of missing precipitation records". *Month. Weath. Rev.*, 80, pp. 129-133.
- STĪPÁNEK, P.; BRÁZDIL, R. y KVETON, V. (2000). Homogenization of air temperature series in the Czech republic. In: *Third Seminar for Homogenization of Surface Climatological Data*, Budapest, Hungary, 25-29 September, 2000.
- STĪPÁNEK, P. (2003). *AnClim - software for time series analysis (for Windows)*. Dept. of Geography, Fac. of Natural Sciences, MU, Brno. 1.47 MB.
- SZENTIMREY, T. (1998). Multiple Analysis of Series for Homogenization (MASH). In: *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data*, Budapest, Hungary, 9-13 November 1998, WMO-TD No. 962, pp. 27-46.
- VENABLES, W.N.; SMITH, D.M. y el *R Development Core Team* (2004). *An Introduction to R*, 108 pp. (<http://cran.r-project.org/>).
- YOUNG, K.C. (1992). "A three-way model for interpolating monthly precipitation values". *Monthly Weather Review*, 120, pp. 2561-2569.