

Homogeneización de series climáticas con Climatol 3.0

José A. Guijarro

Agencia Estatal de Meteorología (AEMET), Delegación en Islas Baleares, España

Resumen

La homogeneización de las series climáticas, para eliminar las perturbaciones que frecuentemente contienen por cambios en las condiciones de observación, es necesaria para aumentar la fiabilidad de los estudios de variabilidad climática derivados de su análisis. Pero esta es una tarea muy laboriosa si no se realiza con la ayuda de paquetes de software que liberen al investigador de la repetitiva manipulación y comparación de largas series de datos. El paquete *Climatol*, programado en *R*, es una de las herramientas disponibles para facilitar estas tareas, y en su nueva versión 3.0, a la función de homogeneización automática *homogen* ya disponible en la anterior versión 2.2, añade nuevas funcionalidades para facilitar tanto el proceso de la homogeneización como la explotación de las series homogeneizadas. En este artículo se detalla la aplicabilidad de estas funciones y se presentan ejemplos sobre su uso.

Palabras clave: software, homogeneización de series climáticas, *Climatol*.

1. Introducción

Como es sabido, cambios de emplazamiento de los observatorios meteorológicos (aunque sean pequeños), cambios en la instrumentación o en los abrigo destinados a proteger los sensores de la radiación solar, o incluso cambios en el entorno (usos del suelo, nuevas construcciones, etc) introducen en las series climatológicas alteraciones que no cabe achacar a cambios en el clima. Por tanto es necesario proceder a eliminar estas alteraciones antes de caracterizar la variabilidad climática a partir de esas series, proceso denominado *homogeneización*.

Hay una extensa literatura científica dedicada a los distintos métodos que pueden usarse para ese fin, como recopilan Peterson et al. (1998) y Aguilar et al. (2003). En cualquier caso, la homogeneización de series se convierte en un procedimiento laborioso y tedioso cuando su número supera un par de docenas, por lo que resulta muy conveniente usar algunas de las implementaciones existentes en forma de paquetes de software públicamente disponibles, cuya lista, junto con unas tablas comparativas que resumen sus características, se pueden encontrar en la página web <http://www.climatol.eu/tt-hom/index.html>

Uno de ellos, desarrollado por este autor en lenguaje *R* (R Development Core Team, 2013), es el paquete *Climatol* (Guijarro, 2016), cuya anterior versión (2.2) ya contaba con la función *homogen* para automatizar la homogeneización de un

conjunto de series limitado únicamente por la capacidad del ordenador, así como con la función *dahstat* para obtener diversos estadísticos (medias, tendencias, percentiles, etc) de las series mensuales homogeneizadas.

La nueva versión 3.0 (<https://cran.r-project.org/web/packages/climatol/index.html>) añade nuevas funcionalidades, que se detallan en los siguientes apartados, finalizando con algunos ejemplos de aplicación. No obstante, para mayor información sobre las opciones de las distintas funciones del paquete debe consultarse la documentación del mismo, de la que este artículo es solamente un complemento.

2. Nuevas funciones

2.1. db2dat

Una parte importante del tiempo necesario para aplicar *Climatol* u otros paquetes de homogeneización en un caso real hay que dedicarlo a preparar los datos de entrada en el formato requerido por el programa a utilizar. La función *db2dat* viene a facilitar este proceso, pues puede acceder a bases de datos a través de un controlador ODBC, y generar los ficheros de entrada para *Climatol*, *.dat y *.est

2.2. dd2m

La detección de inhomogeneidades en series diarias es poco sensible debido a la alta variabilidad inherente a esta resolución temporal. La función *dd2m* permite generar series mensuales a partir de las diarias, sobre las que el análisis de homogeneidad será más efectivo.

2.3. homogen

Esta función ya existía en la versión anterior, pero se han añadido nuevos parámetros, como poder especificar independientemente los umbrales superior e inferior de rechazo de datos anómalos, así como los valores máximo y mínimo admisibles (100 en humedad relativa, 0 en muchas variables, etc). Hay que tener en cuenta que algunos parámetros de esta función han sufrido cambios de nombre para que resulten más fáciles de recordar, como los antiguos umbrales SNHT para cortar las series en las fases 1 (por ventanas solapadas) y 2 (sobre las series completas), que antes se llamaban *tVt* y *snhtt*, y ahora pasan a denominarse *snht1* y *snht2* respectivamente.

Esta función genera ahora un nuevo fichero binario (conteniendo las series originales, homogeneizadas, y otros parámetros utilizados en el proceso), así como listas de valores anómalos y saltos en la media detectados. Este último fichero puede editarse para ajustar las fechas a eventos del historial de los observatorios susceptibles de producir inhomogeneidades, para luego volver a aplicar la función de manera que tenga en cuenta los metadatos disponibles.

2.4. homospplit

Cuando el número de series a tratar es muy elevado (varios miles), puede suceder que lleguemos a superar las capacidades de memoria de nuestro ordenador, o bien que el tiempo de proceso se alargue durante muchos días. Esto se puede solventar mediante esta nueva función, que subdivide el dominio espacial a tratar en áreas rectangulares solapadas, de modo que el número de series se reduzca a unos cientos al tiempo que el solapamiento actúe como zona tampón para minimizar el riesgo de que aparezcan fronteras espurias en las divisorias de cada área.

2.5. dahstat

Esta función de postproceso añade a las funcionalidades anteriores el cómputo de tendencias por regresión lineal con el tiempo (OLS) y sus p-valores, así como la posibilidad de extraer todas las series homogeneizadas a ficheros CSV individuales.

2.6. dahgrid

Esta es otra función de postproceso pensada para generar datos homogeneizados y normalizados interpolados sobre una rejilla definida por el usuario, que se graban en un fichero con formato NetCDF. La normalización es la que se haya elegido al realizar la homogeneización mediante el parámetro *std*, que puede adoptar los valores 1 (restar a cada término de cada serie su valor medio), 2 (dividir por su valor medio) o 3 (restar su valor medio y dividir por su desviación típica, que es la opción por defecto).

El fichero generado contiene también las interpolaciones de los valores medios (y de las desviaciones típicas, si *std*=3), de modo que pasar de capas de valores normalizados a valores absolutos es inmediato. No obstante, si se desea generar mapas de calidad a partir de este fichero, lo más conveniente es suministrar rejillas de alta resolución de las medias (y desviaciones típicas en su caso) obtenidas con algún método geoestadístico que incluya el efecto de factores externos como la orografía, puesto que la interpolación que realiza esta función no puede tener en cuenta esos factores.

3. Ejemplo de aplicación

Este ejemplo se va a basar en un supuesto práctico en el que vamos a partir de un conjunto de series de datos diarios observados, para homogeneizarlos y obtener distintos índices y mapas.

3.1. Datos de partida

Supongamos que queremos hacer un estudio de las características de la precipitación diaria de una determinada zona, y nuestros datos están en una base de datos

accesible mediante el protocolo ODBC, para lo cual tendremos que tener instaladas y configuradas las aplicaciones necesarias en nuestro sistema, y también instalar el paquete de R 'RODBC'.

Para preparar los datos de entrada mediante la función *db2dat* necesitaremos saber tanto el nombre de la base de datos como los de las tablas y campos en los que están almacenados los datos diarios de precipitación y sus fechas correspondientes, así como los de las coordenadas, códigos y nombres de las estaciones pluviométricas, que también vamos a necesitar. Estos nombres podrían ser:

clima: Nombre de la base de datos (con usuario *USER* y clave *PASS*).

Temperatura: Nombre de la tabla que almacena las temperaturas diarias.

Fecha: Nombre del campo que contiene las fechas.

Valor: Nombre del campo que contiene las precipitaciones diarias.

Estaciones: Nombre de la tabla que almacena los datos de las estaciones.

Codigo: Nombre del campo con los códigos de las estaciones. (Supondremos que este campo también está en la tabla de las precipitaciones).

Nombre: Nombre del campo con los nombres de las estaciones.

LatGra: Nombre del campo con las latitudes (en grados, con decimales).

LonGra: Nombre del campo con las longitudes (en grados, con decimales).

Altitud: Nombre del campo con las altitudes (en metros).

Conociendo estos datos, elegimos *Ttest* como nombre de nuestra variable a estudiar, durante el periodo 1981-2000, y procedemos del siguiente modo (*minny*=3 es para rechazar series con menos de 3 años de datos, *dformat* especifica el formato de las fechas, y todo lo que sigue a # son comentarios):

```
R #arrancamos R (versión 3 o mayor)
library(RODBC)
library(climatol)
ch <- odbcConnect("clima",uid="USER",pwd="PASS")
db2dat('Ttest',1981,2000,minny=3,ch,dformat='%Y-%m-%d',
'Temperatura','Codigo','Fecha','Valor','Estaciones',
'Codigo','Nombre','LonGra','LatGra','Altitud')
odbcClose(ch) #cerrar la conexión con la base de datos
```

Esto generaría los ficheros *Ttest_1981-2000.est* y *Ttest_1981-2000.dat* conteniendo el primero las coordenadas, códigos y nombres de las estaciones, y el segundo los datos diarios de precipitación de los diez años en la primera de las estaciones, seguidos por los demás, estación por estación. En los días en los que no se disponga de datos figurará NA (*Not Available*), que es el código de ausencia de dato estándar en R.

Pero esta función requiere que solo haya un dato de la variable a estudiar en cada registro de la base. Es decir, en el registro de una fecha determinada puede haber datos de varias variables (temperatura, humedad, etc), pero no varios datos de la

misma variable (como cuando por ejemplo todos los datos diarios están en un solo registro mensual).

En este caso tendríamos que preparar manualmente los ficheros de entrada, que son ficheros de texto plano con el formato indicado en el párrafo anterior. Así el fichero Ttest_1981-2000.est podría comenzar por:

```
-108.035 44.38 1169.5 "WY003" "Small Horn"  
-108.9006 44.4139 1599.6 "WY018" "Narrow Canyon"  
-108.5931 44.8919 1251.2 "WY020" "Wide Meadows"  
... (etc)
```

Y el fichero Ttest_1981-2000.dat estaría compuesto por líneas como:

```
NA NA NA NA NA -1.5 -5 -1.6 6.7 6.6  
3.6 4.3 5 -0.4 -3 -0.2 -3.4 NA -12.5 -8.6  
... (etc)
```

Estos datos van a ser usados en los ejemplos de los apartados siguientes. Para reproducirlos, el lector interesado puede generar estos ficheros del siguiente modo:

```
library(climatol)  
data(Ttest)  
write(dat, 'Ttest_1981-2000.dat')  
write.table(est.c, 'Ttest_1981-2000.est', row.names=FALSE,  
           col.names=FALSE)  
rm(dat, est.c) #eliminamos los datos de la memoria
```

3.2. Control de calidad de los datos diarios.

La función *homogen* nos va a permitir constatar de cuántos datos disponemos y si contienen anomalías atribuibles a errores. Para ello ejecutaremos:

```
homogen('Ttest', 1981, 2000, snht1=0)
```

Con el parámetro *snht1=0* indicamos que en este momento no nos interesa realizar un análisis de saltos en la media mediante el test SNHT (Alexandersson, 1986), puesto que las series diarias son muy ruidosas y no permiten una adecuada detección de cambios en la media. Pero los primeros gráficos del fichero Ttest_1981-2000.pdf nos informarán de la disponibilidad y distribución de frecuencias de los datos, y también encontraremos, tres páginas antes de la última, un histograma general de anomalías, que nos permitirá detectar datos erróneos y proceder a su corrección antes de continuar. Por defecto se considerarán datos anómalos aquellos que se separen más de 5 desviaciones típicas del valor esperado a partir de los datos vecinos, pero la inspección del histograma puede sugerirnos umbrales de rechazo diferentes. Además, el fichero Ttest_1981-2000_out.csv aporta una lista de los datos sospechosos y los valores esperables a partir de los datos vecinos.

3.3. Agregación de los datos diarios en series mensuales.

Si hubiéramos recopilado datos mensuales podríamos proceder a su homogeneización directamente, pero como se ha mencionado antes, los datos diarios son mucho más variables, y esto hace difícil la detección de inhomogeneidades en sus series. Por este motivo es mejor construir los ficheros de datos mensuales a partir de los diarios, lo que podemos hacer automáticamente con:

```
dd2m('Ttest', 1981, 2000)
```

Esto habrá generado los ficheros `Ttest-m_1981-2000.est` y `Ttest-m_1981-2000.dat`, que se distinguen de los anteriores por el sufijo *m* añadido al nombre de la variable.

3.4. Homogeneización de las series mensuales

La efectuaremos mediante la función *homogen*, que podemos aplicar primeramente de modo exploratorio con:

```
homogen('Ttest-m', 1981, 2000, expl=TRUE)
```

Observando los histogramas de máximos valores de SNHT encontrados en las dos primeras fases del proceso en el fichero `Ttest-m_1981-2000.pdf` podemos ajustar los umbrales mediante los parámetros `snht1` y `snht2`. Vemos en la figura 1 que el valor por defecto, establecido en 25, es adecuado en la aplicación del SNHT a ventanas solapadas (para soslayar problemas de detección debidos a múltiples inhomogeneidades en la misma serie), mientras que al aplicarlo a las series completas podemos aumentarlo a 30, que es un umbral que parece separar adecuadamente las series homogéneas de las que no lo son. Entonces aplicamos de nuevo la función *homogen* ajustando este segundo umbral:

```
homogen('Ttest-m', 1981, 2000, snht2=30)
```

Esto habrá generado los siguientes ficheros:

- `Ttest-m_1981-2000_out.csv` : Lista de datos anómalos corregidos.
- `Ttest-m_1981-2000_brk.csv` : Lista de saltos en la media corregidos.
- `Ttest-m_1981-2000.pdf` : Gráficos de diagnóstico.
- `Ttest-m_1981-2000.txt` : Bitácora con los mensajes generados durante el proceso.
- `Ttest-m_1981-2000.rda` : Datos de la homogeneización (fichero binario de R).

El fichero que contiene la lista de saltos en la media corregidos se usará a continuación para obtener las series diarias homogeneizadas, pero este fichero puede editarse para ajustar las fechas detectadas a los historiales de las estaciones (metadatos). No obstante, hay que tener en cuenta que no todos los eventos sucedidos durante el periodo de funcionamiento de una estación tienen que producir inhomogeneidades en su serie, y que probablemente no todos los eventos que alteran

las serie habrán sido registrados en su historial. Por tanto, si editamos este fichero, únicamente procederemos a ajustar aquellas fechas suficientemente próximas a los cambios en las condiciones de observación que consten en los metadatos.

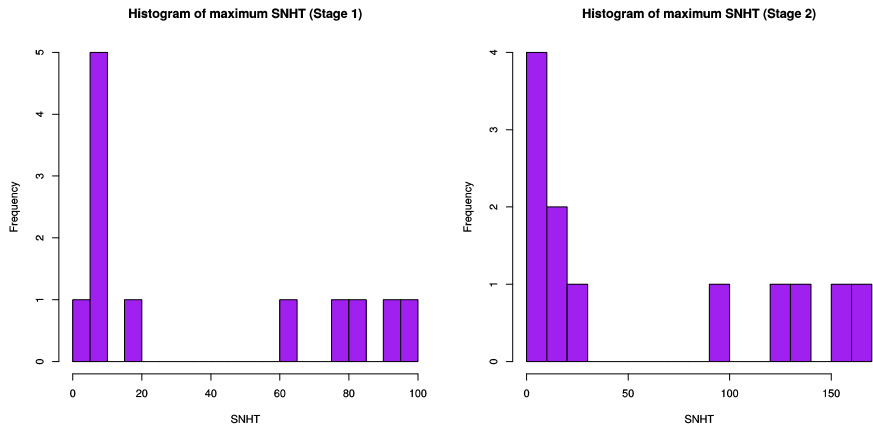


Fig. 1: Histogramas de máximos SNHT encontrados en la aplicación por ventanas solapadas (izquierda) y sobre las series completas (derecha).

3.5. Obtención de las series diarias homogeneizadas

Una vez generado el fichero de saltos detectados a escala mensual, podemos proceder a usarlo para cortar las series diarias por estos puntos y reconstruirlas. (Si no hemos editado ese fichero para ajustar las fechas de las inhomogeneidades de acuerdo a los metadatos de los historiales de las estaciones, el corte se efectuará el día 1 del mes del salto). Para ello volveremos a usar la función general de homogeneización, esta vez con el parámetro `metad=TRUE`. Además, si el histograma de anomalías que obtuvimos en el control de calidad de los datos diarios (apartado 3.2) sugiere que el umbral de rechazo de datos por defecto, establecido en 5 desviaciones típicas, es demasiado riguroso, podemos relajarlo a, por ejemplo, 6 desviaciones típicas:

```
homogen('Ttest', 1981, 2000, dz.max=6, metad=TRUE)
```

Las series diarias homogeneizadas y otros datos del proceso quedan almacenados en el fichero binario `Ttest_1981-2000.rda`, que será usado por las funciones `dahstat` y `dahgrid` para generar los productos deseados. (Sus datos también pueden ser cargados en memoria mediante `load('Ttest_1981-2000.rda')` para su posterior utilización por el usuario).

3.6. Obtención de productos con dahstat

Esta función permite generar ficheros conteniendo las series homogeneizadas:

```
dahstat('Ttest',1981,2000,stat='series')
```

Se obtendrán así dos ficheros CSV por cada serie reconstruida a partir de cada fragmento homogéneo, uno con los datos homogeneizados, y otro indicando si cada dato es el observado originalmente (0), ha sido rellenado (1), o se ha corregido (2).

Pero la principal utilidad de esta función reside en el cálculo de tablas de distintos estadísticos a partir de las series homogeneizadas. El ejemplo más trivial de aplicación a nuestras series diarias sería:

```
dahstat('Ttest',1981,2000)
```

Con lo que habremos obtenido los valores medios de cada serie en el fichero Ttest_1981-2000_me.csv. Pero es más probable que estemos interesados en obtener las medias (y otros estadísticos) a nivel mensual. Entonces debemos volver a agregar los datos diarios en series mensuales de modo similar a como hicimos anteriormente (subapartado 3.3), pero indicando que queremos agregar los datos homogeneizados (con `homog=TRUE`):

```
dd2m('Ttest',1981,2000,homog=TRUE)
```

Los ficheros generados tendrán el sufijo mh para indicar que se trata de datos mensuales obtenidos a partir de series diarias homogeneizadas, de forma que no se sobre-escribirán los ficheros de la agregación mensual anterior. Llegados a este punto, ya podemos proceder a obtener listas de parámetros estadísticos mensuales a partir de estos ficheros, indicando `mh=TRUE`. Ejemplos:

```
dahstat('Ttest',1981,2000,mh=TRUE) #medias
dahstat('Ttest',1981,2000,mh=TRUE,stat='std',ndc=2)#desv. típ.
dahstat('Ttest',1981,2000,mh=TRUE,stat='q',prob=.2)#quintil 1
dahstat('Ttest',1981,2000,mh=TRUE,stat='tnd')#tendencias
```

Esta última orden genera dos ficheros: el Ttest_1981-2000_tnd.csv que contiene las tendencias (calculadas por regresión lineal con el tiempo), y otro con los correspondientes p-valores (Ttest_1981-2000_pval.csv).

Por defecto estos productos se generan para todas las series reconstruidas, pero se pueden especificar los códigos de las estaciones que nos interesen, o seleccionarlas de acuerdo a diferentes criterios (funcionamiento al final del periodo estudiado, fragmento más largo, SNHT superior a un umbral, etc).

3.7. Series homogeneizadas interpoladas en una rejilla

Muchas bases de datos climatológicas suministran la información en forma de rejilla, lo que implica una estructura espacio-temporal regular que facilita el estudio de la variabilidad climática. La función *dahgrid* se ha programado para poder obtener automáticamente datos homogeneizados en rejilla, procedentes del fichero *.rda que genera la función *homogen* vista anteriormente.

Para ello el usuario debe definir previamente la rejilla espacial sobre la que desee que se haga la interpolación, como por ejemplo:

```
rejilla=expand.grid(x=seq(-109, -107.7, .02), y=seq(44, 45, .02))
library(sp)
coordinates(rejilla) = ~x+y
```

(Si el área contiene zonas marítimas o de países vecinos para los que no se dispone de datos, puede ser conveniente eliminar de la rejilla los puntos correspondientes).

Y ahora ya se pueden interpolar los datos homogeneizados sobre la rejilla definida, para cada paso de tiempo, bien a escala mensual (240 grids, uno por cada mes), con:

```
dahgrid('Ttest', 1981, 2000, grid=rejilla, mh=TRUE)
```

bien a escala diaria (7305 grids), eliminando `mh=TRUE` de la orden anterior. Los grids quedarán grabados en los ficheros `Ttest-mh_1981-2000.nc` o `Ttest_1981-2000.nc` respectivamente, en formato NetCDF.

Estos grids se calculan por el sencillo método de ponderación inversa a la distancia, pero con datos normalizados, minimizando así la influencia de la altitud en los valores de la variable estudiada. Se suministran también los grids de medias y desviaciones típicas con los que poder calcular grids deshaciendo la normalización, lo que puede llevarse a cabo fácilmente usando herramientas externas, como por ejemplo las CDO (*Climate Data Operators*):

```
cdo add -mul Ttest-mh_1981-2000.nc Ttest-mh_1981-2000_s.nc
Ttest-mh_1981-2000_m.nc Ttest-mu_1981-2000.nc
```

Pero para que los resultados tengan en cuenta efectos orográficos es preferible que el usuario genere estos grids mediante métodos geoestadísticos, para lo que también se facilita el fichero `Ttest_1981-2000_means.csv`, que contiene las medias y desviaciones típicas de cada serie junto con sus coordenadas. (Más detalles en la documentación estándar del paquete).

3.8. Homogeneización de miles de series mediante áreas solapadas

Si el número de series excede la capacidad del ordenador, o alarga el tiempo de proceso demasiado (muchos días), se puede utilizar la función *homogsplit* para realizar la homogeneización dividiendo el dominio espacial en áreas rectangulares. Estas áreas han de solaparse en sus fronteras, para que se puedan utilizar series próximas de las áreas vecinas, pues de lo contrario aparecerían discontinuidades al analizar los resultados globalmente.

Como ejemplo vamos a suponer que queremos analizar las 3005 series españolas de temperatura máxima media que hemos recopilado en el fichero de datos TA_1951-2015.dat (con las coordenadas de las estaciones en TA_1951-2015.est). Entonces podríamos realizar la homogeneización subdividiendo nuestra área por los meridianos -12° y 1° , y por los paralelos 31° y $40,4^\circ$, con un solapamiento de 1° en longitud y $0,76^\circ$ en latitud (figura 2), mediante la orden (no reproducible sin disponer de los citados ficheros de entrada):

```
homogsplit('TA', 1951, 2015, xc=c(-12,-3), yc=c(31,40.4), xo=1, yo=.76)
```

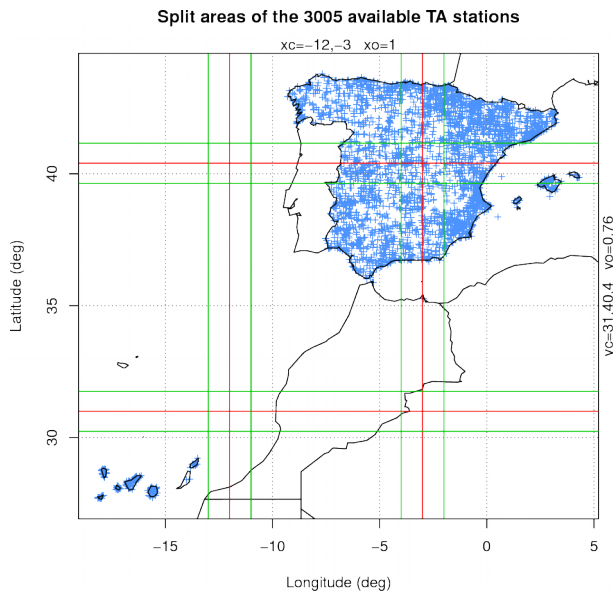


Fig. 2: Mapa mostrando las áreas a homogeneizar. Para cada rectángulo delimitado por las líneas rojas se usarán también las estaciones de alrededor (hasta las líneas de color verde).

Se realizarían así 5 procesos independientes de homogeneización (uno para cada subárea conteniendo un mínimo de series), cuyos gráficos aparecerían también en sendos documentos en PDF, pero los datos homogeneizados quedarían finalmente agrupados en un solo fichero llamado TA_1951-2015.rda

4. conclusiones

La nueva versión del paquete de R *Climatol* permite automatizar de una manera cómoda todo el proceso de un estudio climatológico, desde la recopilación de datos de observación hasta la obtención de productos a partir de las series homogeneizadas.

El trabajo mediante órdenes en una terminal de texto puede resultar un poco árido para el usuario no acostumbrado, pero tiene la enorme ventaja de permitir la automatización de los diferentes procesos incluyendo las órdenes en un *script* ejecutable.

Agradecimientos

Las mejoras introducidas en *Climatol* desde su primera versión (Guijarro, 2004) se han beneficiado de las discusiones mantenidas con otros colegas durante y después de la Acción COST ES0601 (Venema *et al.*, 2011).

Este artículo es una actualización del publicado en las actas del X Congreso de la Asociación Española de Climatología (Guijarro, 2016).

Bibliografía

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J (2003): *Guidelines on climate metadata and homogenization*. WCDMP-No. 53, WMO-TD No. 1186. World Meteorological Organization, Geneve.

Alexandersson H (1986): A homogeneity test applied to precipitation data. *Jour. of Climatol.*, 6:661-675.

Guijarro JA (2004): Software libre para la depuración y homogeneización de datos climatológicos. En García-Codron et al. (Eds.), *El clima, entre el Mar y la Montaña*, Asociación Española de Climatología, A-4:493-502.

Guijarro JA (2016): <https://cran.r-project.org/web/packages/climatol/index.html> y <http://www.climatol.eu/index.html>

Guijarro JA (2016): Automatización de la homogeneización de series climáticas: nuevas funciones del paquete *Climatol* 3.0. X Congreso Internacional de la Asociación Española de Climatología (Alicante, 5 a 8 de octubre), en prensa.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Böhm R, Gullett D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salinger J, Førland E, Hanssen-Bauer I, Alexandersson H, Jones P, Parker D (1998): Homogeneity Adjustments of 'In Situ' Atmospheric Climate Data: A Review. *Int. J. Climatol.*, 18, pp:1493-1518.

R Development Core Team (2013): *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>

Venema V, Mestre O, Aguilar E, Auer I, Guijarro JA, Domonkos P, Vertacnik G, Szentimrey T, Stepanek P, Zahradnicek P, Viarre J, Müller-Westermeier G, Lakatos M, Williams CN, Menne M, Lindau R, Rasol D, Rustemeier E, Kolokythas K, Marinova T, Andresen L, Acquavotta F, Fratianni S, Cheval S, Klancar M, Brunetti M, Gruber C, Prohom Duran M, Likso T, Esteban P And Brandsma T (2012):| Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, 8:89-115.