

EJERCICIO DE HOMOGENEIZACIÓN Y RELLENO DE SERIES DIARIAS DE TEMPERATURA MÁXIMA, MEDIANTE EL USO DE CLIMATOL

Elisa M^a HERNÁNDEZ GARCÍA, Juan A. GARCÍA VALERO, Juan E. PALENZUELA CRUZ, Fernando BELDA ESPLUGUES
Delegación Territorial de AEMET en Murcia (jugarciav@aemet.es)

RESUMEN

Disponer de series diarias de variables climáticas con un periodo temporal suficientemente largo es necesario para un análisis profundo de la variabilidad natural, y fundamental en el análisis de extremos. No obstante, las inhomogeneidades presentes en las series son un gran inconveniente requiriéndose para ello el uso de métodos de homogeneización y relleno de lagunas, la mayoría de ellos pensados para la escala mensual. En este estudio, se presentan los resultados de diferentes pruebas de homogeneización y relleno de 83 series diarias de temperatura máxima ubicadas en la Región de Murcia. Las pruebas fueron efectuadas con el paquete estadístico Climatol (R software) que genera series de referencia con la información de las estaciones disponibles y evalúa la homogeneidad de las series de anomalías entre cada estación y su referencia mediante el test SNHT. Los resultados evidencian la necesidad de emplear valores de los estadísticos para el rechazo de la homogeneidad muy superiores a los propuestos por otros autores en la escala mensual.

BASE DE DATOS

La base de datos está constituida por series diarias de temperaturas máximas y mínimas de 83 estaciones, relativas a la red termoplumiométrica que AEMET tiene distribuidas en la Región de Murcia. La distribución regular de las estaciones en la Región puede verse sobre la Fig. 1, representándose en color azul las estaciones más completas del periodo de estudio (1950-2010). Dichas estaciones presentan más del 90% de los datos de dicho periodo, y serán referidas de aquí en adelante como estaciones PD90.

En la Fig. 2, se representa la variación del número de estaciones a lo largo del periodo de estudio. La distancia media entre las estaciones más cercanas es de 11 km.



Figura 1: Localización de las 83 estaciones utilizadas (estaciones PD90 en azul).

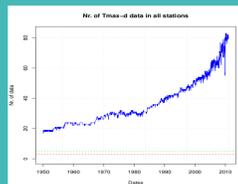


Figura 2: Variación del número de estaciones a lo largo del periodo de estudio (1950-2010).

CONTROL DE CALIDAD

Durante el periodo previo al año 2000, el control de calidad (QC) aplicado a los datos en la D.T. AEMET en la Región de Murcia, consistía en la detección de valores imposibles ($t_{max} < t_{min}$, $t_{max} > 50^{\circ}C$, etc.). A partir de ese año, además de este QC, los datos empezaron a ser sometidos a un chequeo de consistencia espacial, en el que se comparaban los valores de una estación con los de las estaciones vecinas que compartían una variabilidad similar. El QC aplicado a los datos durante el periodo de estudio probablemente no haya filtrado otro tipo de errores presentes en las series y, además, el periodo temporal en el que se está realizando el análisis de la coherencia espacial parece insuficiente. Por ello, se han planteado dos nuevos filtros para el QC de las series y que se detallan a continuación:

1. **Persistencia de las series:** Eliminación de datos repetidos de una misma estación, tanto en temperaturas máximas como en mínimas, durante un determinado número de días consecutivos. Este valor se estimó previamente para cada estación con el fin de tener en cuenta la diferente variabilidad, que pudiera existir entre las distintas estaciones. Para la determinación de ese umbral se escogió el valor medio del Tercer Cuartil y Máximo de la población de persistencia de cada serie. En la Fig. 3, aparecen los distintos umbrales empleados para cada una de las estaciones.

2. **Coherencia espacial:** Eliminación de datos poco coherentes con las estaciones vecinas. Diariamente se obtuvo un valor de temperatura media regional con las estaciones disponibles, obteniéndose para cada estación una serie de anomalías normalizadas, respecto a la media regional. De este modo se eliminaron los datos de aquellas estaciones con desviaciones estándar superiores a 3.



Figura 3: Umbrales empleados para el QC de persistencia de las series, localizados por estaciones.

PROCESO DE HOMOGENEIZACIÓN

CLIMATOL

- El paquete de R, Climatol, ha sido empleado aquí para el control de outliers, y para el proceso de homogeneización y relleno de lagunas de las series.
- La función Climatol debe inicializarse a partir de un conjunto de parámetros, descritos en la Tabla 1. Los valores de dichos parámetros dependerán de la BBDD empleada, así como de la aplicación que vaya a dársele a los datos resultantes del proceso de homogeneización y relleno.
- La mayoría de ejercicios de homogeneización de series que emplearon Climatol, se llevaron a cabo con series mensuales. Por ello, los parámetros de ajuste definidos por defecto en la función, no son los idóneos para trabajar con series diarias.
- Climatol lleva a cabo un proceso de homogeneización relativa, que emplea el test SNHT (Alexandersson-Moberg) para el análisis de la homogeneidad de las series de anomalías, obtenidas éstas, a partir de la estación a testear y otra de referencia, calculada esta última a partir de un promedio ponderado de estaciones vecinas, teniendo en cuenta en la ponderación, la distancia entre las estaciones a la estación de referencia.

Tabla 1: Parámetros necesarios para la inicialización de Climatol.

PARÁMETRO	SIGNIFICADO
<i>wd</i>	Peso otorgado a las estaciones vecinas para su ponderación en la creación de series de referencia
<i>dz.max</i>	Umbral de desviaciones estándar en el análisis de outliers
<i>snht</i>	Umbral de rechazo del test de homogeneidad cuando se aplica a todo el periodo
<i>tv</i>	Umbral de rechazo del test de homogeneidad cuando se realiza por ventanas
<i>swa</i>	Tamaño de la ventana cuando se realiza el test de homogeneidad por ventanas

AJUSTE DE PARÁMETROS CLAVE EN EL PROCESO DE HOMOGENEIZACIÓN

- Los parámetros *tv* y *snht*, umbrales de inhomogeneidad de los estadísticos del test SNHT cuando se aplica por ventanas o a toda la serie, respectivamente, fueron determinados a partir de un ejercicio inicial, en el que sólo se aplicó relleno de lagunas. En este ejercicio, Climatol ofrece como resultado el histograma con los *tv* máximos, obtenidos para cada una de las series.
- El análisis exploratorio del histograma (Fig. 4) permite inferir los valores de *tv* que se utilizarán en las pruebas de homogeneización. De este modo, se toma como valor del parámetro, el correspondiente al segundo máximo relativo del histograma.
- La subjetividad de este método para la elección de *tv*, sugiere la posibilidad de elección de nuevos umbrales. En nuestro caso, se consideró también el percentil 95 de los *tv* obtenidos en el proceso exploratorio, pero correspondientes sólo a las estaciones PD90.
- De este modo, se seleccionaron dos umbrales para el parámetro *tv*: *tv*=300 y *tv*=500, obtenidos de los métodos que emplean el histograma y percentil95, respectivamente. En cuanto al parámetro *snht*, se utilizaron los mismos umbrales que en el parámetro *tv*.

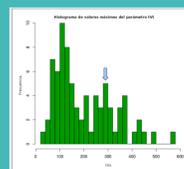


Figura 4: Histograma de valores máximos tv, obtenido con Climatol.

EJERCICIOS DE HOMOGENEIZACIÓN REALIZADOS

- Con las series diarias de temperatura máxima se llevaron a cabo tres ejercicios:
 - Sólo relleno de lagunas (*tv*=0). Para la obtención de los parámetros *tv* y *snht*, según los métodos explicados anteriormente.
 - Dos ejercicios de homogeneización, considerando los parámetros *tv*=300 y *tv*=500.
 En los tres ejercicios, se mantuvieron fijos otros parámetros requeridos por Climatol (*wd*=c(0,0,15), *swa*=365,*dz.max*=5), descritos en la Tabla 1.
- Además, se efectuaron ejercicios similares, considerando las series mensuales. En este caso se tomaron como parámetros, aquellos que vienen definidos por defecto en la función Climatol (*wd*=c(0,0,15), *swa*=60,*dz.max*=5,*tv*=25,*snht*=50).



Figura 5: Esquema de los ejercicios de homogeneización realizados con Climatol.

RESULTADOS

EJERCICIOS A ESCALA DIARIA

- En la Tabla 2 se muestran los resultados de las pruebas de relleno de lagunas (*tv*=0) y homogeneización (*tv*=300 y *tv*=500) obtenidas con las series diarias.
- Para la comparación entre las distintas pruebas se han utilizado tres índices: ECM (error cuadrático medio de las series rellenas u homogeneizadas), SDM (desviación estándar o variabilidad media de las series) y PDOM (Porcentaje medio de datos originales conservados en las series rellenas u homogeneizadas). Los índices fueron calculados considerando las 83 series iniciales ("todas las series") y para las 19 estaciones más completas ("estaciones PD90").
- En general, cuando se consideran las estaciones PD90 las pruebas efectuadas mantienen la varianza inicial. En el caso de todas las series, el elevado número de lagunas presente hace que no se conserve toda la varianza, pero sí una parte importante.
- En cuanto al ECM, éste disminuye en torno a un 20% cuando se homogeneiza, respecto a cuando sólo se rellena.
- Cuando el nivel de exigencia de la homogeneización es mayor (menores valores de *tv*), el ECM disminuye ligeramente, pero disminuye el PDOM de forma más significativa.

EJERCICIOS A ESCALA MENSUAL

- En la Tabla 3, se muestran los índices obtenidos en las pruebas con las series mensuales. En este caso, se obtienen resultados similares a las series diarias: mantenimiento de la varianza y disminución del ECM tras el proceso de homogeneización, pero a costa de perder datos originales.

Tabla 2: Resultados de los ejercicios de homogeneización a escala diaria

Pruebas diarias	tv	Todas las series			Estaciones PD90		
		ECM	PDOM	SDM	ECM	PDOM	SDM
Series originales	-	-	-	7,46	-	-	7,51
Relleno de lagunas	0	1,72	42,80	6,82	2,22	96,21	7,51
Homogeneización	300	1,48	26,30	6,94	1,83	45,16	7,49
	500	1,53	31,30	6,93	1,95	57,05	7,55

Tabla 3: Resultados de los ejercicios de homogeneización a escala mensual

Pruebas mensuales	tv	Todas las series			Estaciones PD90		
		ECM	PDOM	SDM	ECM	PDOM	SDM
Series originales	-	-	-	6,75	-	-	6,78
Relleno de lagunas	0	0,93	43,63	6,66	1,37	97,11	6,79
Homogeneización	25	0,66	21,61	6,56	0,77	34,00	6,74

COMPARACIÓN ENTRE LA ESCALA DIARIA Y LA MENSUAL

- Se ha analizado la capacidad de Climatol de detectar saltos similares en las series, obtenidos tanto con las series diarias como en las mensuales. Los resultados evidencian un mayor número de saltos obtenidos en la escala mensual, de los cuáles un 38% y 34% también se detectaron en la escala diaria usando los umbrales *tv*=300 y *tv*=500, respectivamente. Por otro lado, el 62% y 68% de los saltos detectados en la escala diaria (*tv*=300 y *tv*=500, respectivamente) también fueron detectados con las series mensuales.
- Un ejemplo de detección simultánea puede verse en la Fig. 6, correspondientes a las pruebas de homogeneización efectuadas sobre la serie de Alcantarilla (una de las más completas de toda la BBDD).

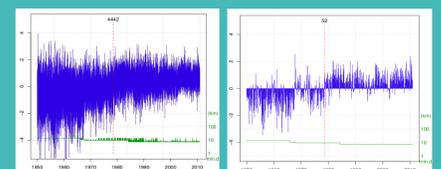


Figura 6: Fechas coincidentes de saltos similares en la serie de Murcia/Alcantarilla. A la izquierda se muestra la serie diaria de anomalías (estación-referencia) y a la derecha la serie mensual.

CONCLUSIONES

- La exploración de inhomogeneidades en las series diarias es un reto actual, que requiere el uso de herramientas versátiles. En este estudio, se ha probado el funcionamiento de Climatol (R software), pensado más, en un principio, para su aplicación en series mensuales. Su uso, relativamente sencillo, así como su capacidad de conservar gran parte de la varianza original, y su poder de detección de los mismos saltos, tanto en la escala diaria como en la mensual, convierten a Climatol en una herramienta interesante para seguir investigando en la homogeneización de las series diarias.
- La dificultad de homogeneización a escala diaria con Climatol, radica fundamentalmente en la elección del estadístico del test SNHT. Nuevos métodos de selección, mediante simulaciones de Montecarlo deberían ser explorados, aunque la idea que se extrae de los resultados aquí obtenidos es que éstos deberían utilizar, fundamentalmente, las series de estaciones más completas.
- A falta de un criterio claro para la selección de los umbrales, lo más recomendable sería optar por umbrales más exigentes, cuando se pretenda realizar un análisis a posteriori de la variabilidad de los valores medios de las series homogeneizadas, y más conservadores, cuando el objetivo sea el análisis de la variabilidad de ciertos índices extremos.

AGRADECIMIENTOS

Nuestro agradecimiento a AEMET por la concesión de la beca "Estudio climatológico de golpes y olas de calor", sin la cual no hubiera sido posible realizar este estudio. También nos gustaría agradecer a Jose Antonio Guijarro el gran esfuerzo dedicado en el desarrollo del paquete Climatol, y a todos los colaboradores de la D.T. AEMET en la Región de Murcia.