

Selección de estaciones termométricas mediante un filtro de comparación de curvas

José Antonio Sosa Cardo

Instituto Nacional de Meteorología, Centro Meteorológico Territorial en Extremadura. (jasosa@inm.es)

(Recibido: 5-Dic-2001. Publicado: 27-Dic-2001)

Resumen

A la hora de realizar cualquier estudio climatológico o interpolación espacial es fundamental contar con un conjunto de estaciones fiables, y desechar aquellas que lo único que nos van a aportar es ruido. Se han desarrollado dos métodos para seleccionar estaciones de forma objetiva: 1) Mediante análisis de las formas de las curvas de las series, siendo necesaria otra estación con una curva similar para que se de por válida una estación. 2) Mediante análisis del valor diario. Cuando existe una diferencia importante entre el valor diario transportado y el de las demás estaciones se rechaza la estación en cuestión. Estos métodos permiten eliminar ruido indeseado y tener cierta confianza en la calidad de las estaciones seleccionadas.

Palabras clave: selección, estaciones termométricas, temperatura, depuración de datos.

1. Introducción

Todos los que en algún momento han trabajado con estaciones termométricas de la red secundaria saben que la calidad de los datos en muchas ocasiones no es todo lo buena que se desearía. Por eso cuando se pretenden utilizar este tipo de datos se hace casi imprescindible aplicar alguna depuración o filtrado, ya sea de tipo objetivo o subjetivo, de las estaciones que se van a utilizar.

Si se trabaja con promedios de estaciones que han funcionado durante periodos de tiempo diferentes resultan muy útiles métodos sencillos como el de las diferencias (Conrad y Pollack, 1962; Jansá, 1969). Otros autores han utilizado métodos de relleno de lagunas o detección de errores basados en las correlaciones entre estaciones vecinas (De Ruffray *et al.*, 1981; Brisse y Granjean, 1981; Felicísimo y Álvarez, 1982; Guijarro y Morey, 1998; Aguilar *et al.*, 1999), generalmente aplicados a series de datos mensuales.

En este trabajo se describirán un par de métodos objetivos complementarios que, aplicados a series de datos diarios, tienen por objeto realizar una selección de estaciones termométricas, rechazando las que no superen ciertos umbrales de calidad.

2. Filtro de comparación de curvas

La idea es simple: Si dos observadores llegan a unas series de observaciones en un intervalo temporal determinado común, que presentan formas parecidas al ser representadas gráficamente, ambas series de observación, en dicho periodo temporal común, deben ser buenas. Diremos entonces que las dos estaciones se avalan una a la otra. Como lo que se intenta es que los datos que se van a utilizar sean buenos, no se aceptarán por lo tanto datos que provengan de estaciones que no hayan sido avalados por los datos de otra estación.

En la figura 1 se presenta un ejemplo de una pareja de estaciones que se avalan mutuamente porque tienen curvas mensuales parecidas, mientras que otro par de estaciones no se avalan entre sí al tener curvas mensuales distintas.

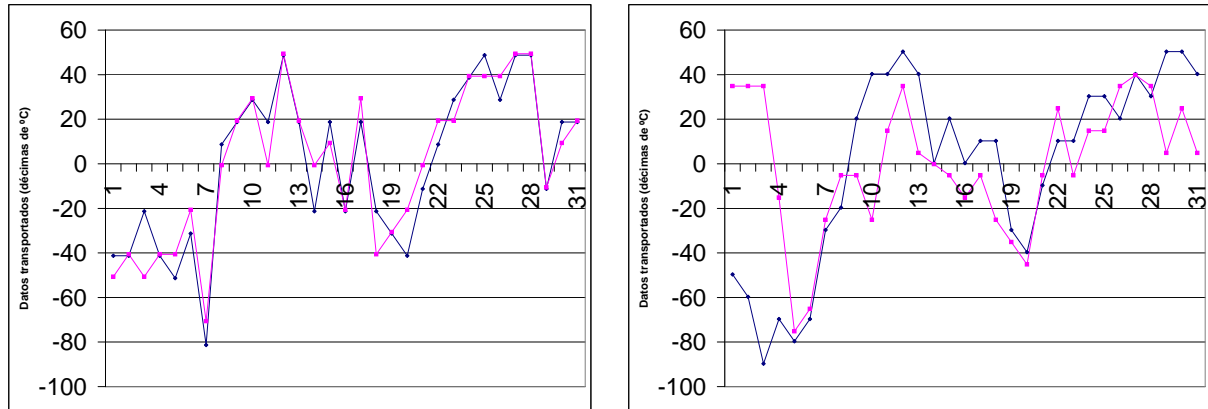


Fig. 1.- Ejemplo de parejas de estaciones con datos termométricos diarios concordantes (izquierda) y discordantes (derecha).

2.1. Filtro de comparación de curvas

Para la puesta en práctica de esta idea, lo primero que se ha hecho ha sido sustraer, a cada valor diario, la media mensual de la serie de datos diarios en cuestión; así llevamos a todas las estaciones a un mismo nivel en que podemos compararlas. En adelante, llamaremos a los datos diarios así generados *datos diarios transportados*.

En segundo lugar se ha definido la distancia entre curvas mensuales como una distancia geométrica, suponiendo que cada curva mensual es un punto en un espacio n -dimensional, siendo n el número de días de la ventana temporal utilizada, y que los datos diarios transportados son las coordenadas geométricas, es decir:

$$d_{ij} = \sqrt{\sum_{k=1}^n [(x_{ik} - \bar{x}_i) - (x_{jk} - \bar{x}_j)]^2}$$

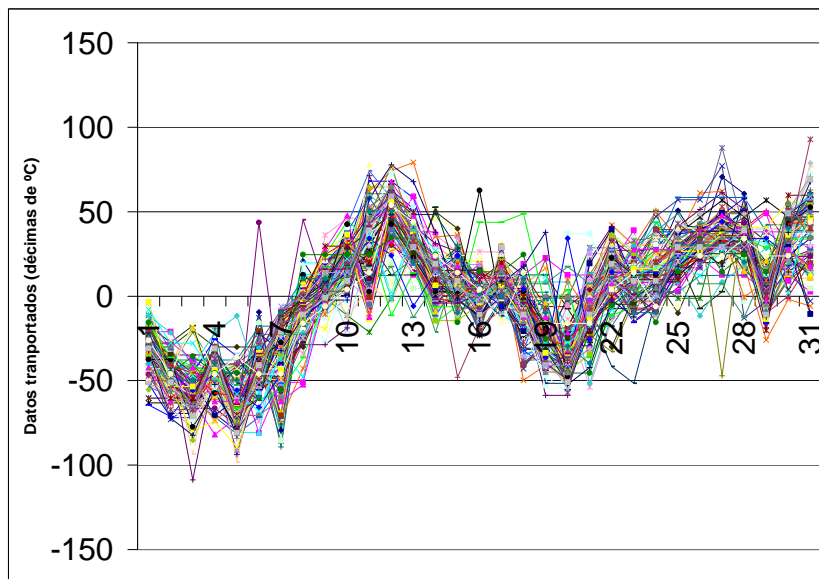
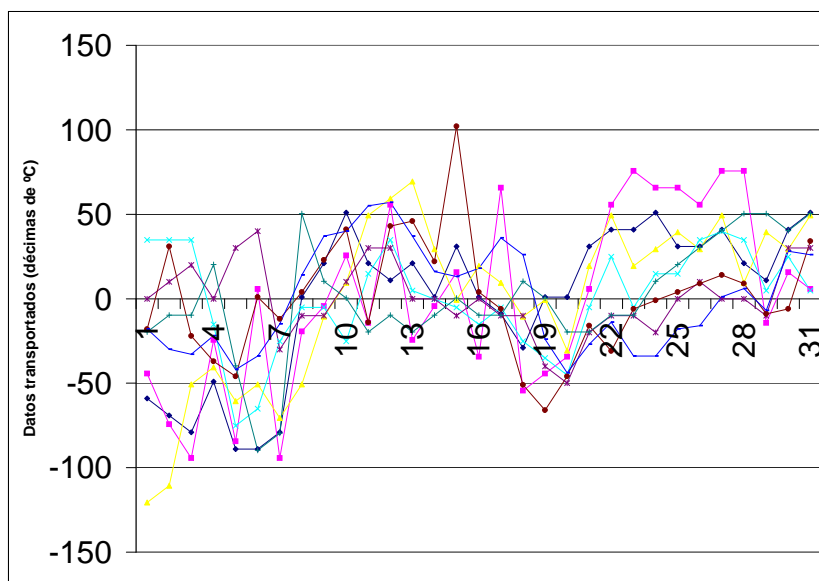
donde d_{ij} es la distancia entre las curvas mensuales de las estaciones i y j , x_{ik} es el valor diario de la estación i y del día k , \bar{x}_i es la media mensual de los valores diarios de la estación i , y n es el número de días de la ventana temporal utilizada (p.e. un mes). Una vez se tienen las distancias entre estaciones, se comprueba para cada una de las estaciones si existe otra a una distancia menor a un umbral dado, y si es así se validan ambas estaciones.

Uno de los problemas del método es determinar cuál es el umbral óptimo. El umbral empleado para el filtrado se ha tomado en unidades de la suma de las desviaciones estándar de los datos diarios:

$$d_{umbral} = f \sqrt{\sum_{k=1}^n \sigma_k^2}$$

donde n es el número de días del mes en cuestión, σ_k son las desviaciones estándar de los datos diarios de todas las estaciones del día k , n es el número de días de la ventana temporal empleada, y f es un parámetro de ajuste. Este parámetro nos dará por lo tanto una medida de la tolerancia del método de filtrado en función de las desviaciones diarias: Mayor (menor) f implica mayor (menor) tolerancia. De este modo un par de estaciones i y j se aceptarán como válidas si $d_{ij} < d_{umbral}$

Como ejemplo de aplicación de este primer filtro podemos ver las figuras 2 y 3, donde representamos los datos diarios de temperatura máxima del mes de enero de 1997, transportados mediante sustracción de sus valores medios, del conjunto de estaciones de la red secundaria del INM disponibles para Extremadura, utilizando un valor de $f = 1, 2$.

Fig. 2.- Datos de las estaciones que han pasado el filtro con $f = 1,2$ Fig. 3.- Datos de las estaciones que no han pasado el filtro con $f = 1,2$

2.2. Filtro de Picos diarios

Un problema que se ha encontrado a la hora de establecer un umbral para el filtrado es que, cuando hay picos diarios pronunciados y sospechosos, el valor que hay que asignar a f para que filtre estas estaciones es tal que también se filtran estaciones de las que en principio no hay por qué desconfiar. Por eso se propone un segundo filtro complementario a éste, que esté especializado sólo en la búsqueda de picos diarios.

Este segundo filtro actúa de la siguiente forma: si el valor diario transportado (valor diario menos el valor medio de la serie en cuestión), es mayor que p veces el valor de la desviación estándar de los valores diarios transportados de todas las estaciones para ese día, entonces se rechaza la serie; si es menor se aceptará.

Es decir la serie mensual se acepta si:

$$|x_i - \bar{x}_i| < p\sigma_i$$

para cualquier valor de i comprendido entre 1 y n , donde x_i es el valor diario de la variable transportada del día i de la ventana temporal elegida, \bar{x}_i es el valor medio de la variable transportada del día i , σ_i es el valor de la desviación estándar de los valores diarios de la variable transportada de todas las series (también para el día i), n es el número de días de la ventana temporal elegida, y p es el parámetro que determina la tolerancia de este filtro.

Para determinar el umbral de p se pueden utilizar gráficos como el de la figura 4, donde se representa la dispersión de los datos diarios transportados en unidades de desviación estándar diaria.

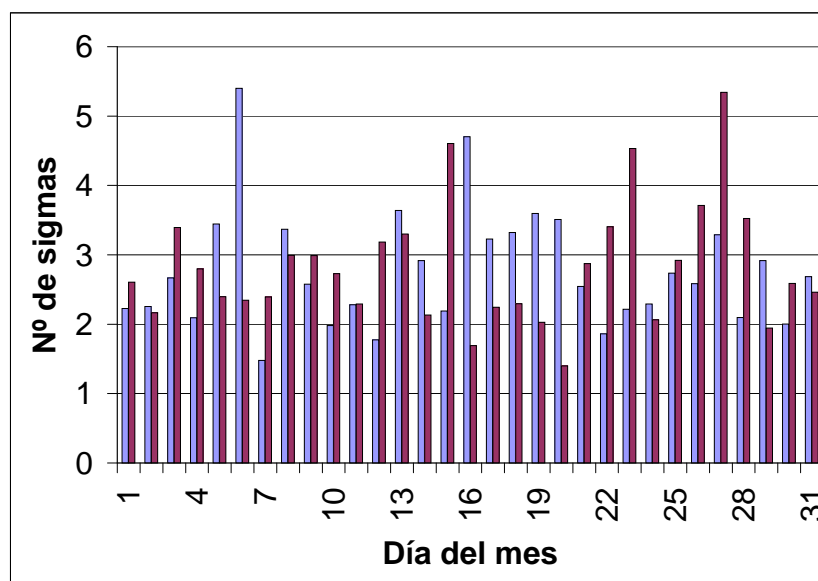


Fig. 4.- Dispersión de los datos transportados en unidades de desviación estándar diaria.

Puede verse que, para el caso presentado, un buen valor para el umbral puede ser $p = 4$, ya que se eliminarían los valores más extremos.

Combinando ambos filtros, obtenemos una selección más lógica desde el punto de vista subjetivo (figuras 5 y 6).

3. Aplicación a la cartografía de temperaturas máximas diarias

Tras aplicar ambos filtros, con los umbrales optimizados en el apartado anterior ($f=1,2$ y $p=4$), a las temperaturas máximas diarias de enero de 1997 de nuestra zona, se seleccionó un conjunto fiable de estaciones termométricas. En la figura 7 se puede ver la diferencia entre el mapa obtenido con la selección efectuada y el que se obtendría si se utilizasen todas las estaciones disponibles. Ambos se han construido a partir de las temperaturas máximas, interpoladas mediante *kriging*, correspondientes al día 2 de enero de 1997.

Como puede observarse el mapa obtenido tras el filtrado es mucho más creíble, ya que es mucho más acorde con la geografía subyacente (se pueden entrever el Valle del Guadiana, Gredos, Villuercas, y las Sierras del Sur) que el obtenido utilizando todo el conjunto de estaciones, aportando éstas un ruido innecesario (aparecen algunos dipolos sospechosos) y molesto a la hora de realizar cualquier tipo de estudio.

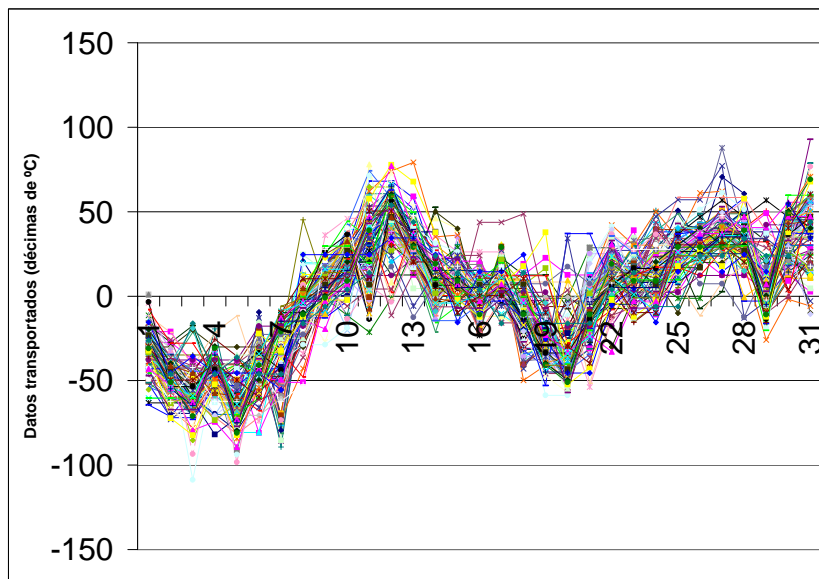


Fig. 5.- Datos de las estaciones que han pasado ambos filtros $f = 1, 2$ y $p = 4$

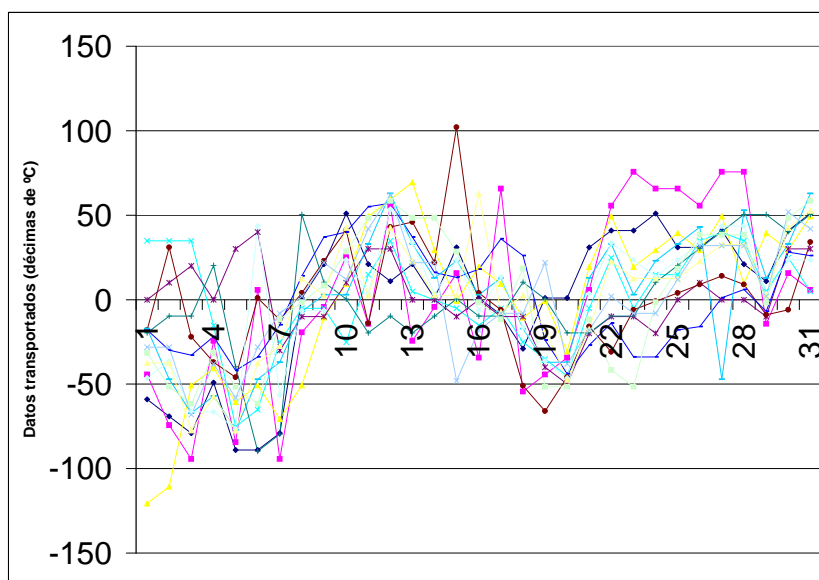


Fig. 6.- Datos de las estaciones que no han pasado alguno de los dos filtros $f = 1, 2$ y $p = 4$

4. Discusión

La limitación más evidente de estos filtros propuestos es que, por la forma en que están contruidos, no son capaces de identificar aquellas estaciones que a pesar de medir *bien* tienen errores sistemáticos de medida, mala instrumentación o mal emplazamiento de la estación. La única manera de identificar esos errores sería, en principio, mediante un modelo espacial lo suficientemente bueno como para que dejase en evidencia ese tipo de datos.

Otra de las limitaciones es que la determinación de los umbrales ha de hacerse, de momento, de forma subjetiva, aunque todo el proceso posterior es objetivo. Si la selección de las estaciones se hace de forma manual, las subjetividades son mucho mayores y se suelen basar en unas presunciones que no siempre son ciertas. Queda pendiente para el futuro desarrollar un método de establecimiento de umbrales óptimos de forma objetiva.

Aunque el método se ha desarrollado para temperaturas máximas, cabe aplicarlo también a las mínimas. Sin embargo habrá que repetir la optimización de los umbrales, puesto que la topografía debe imponer una variabilidad espacial mayor en las temperaturas mínimas que en las máximas.

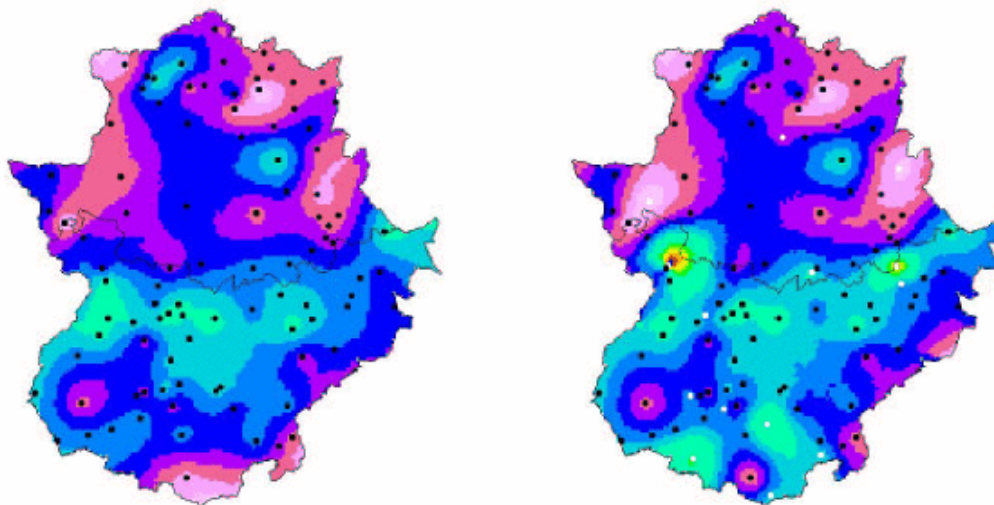


Fig. 7.- Temperatura media de las máximas del día 2-1-1997 de Extremadura (España). El mapa de la izquierda se ha obtenido a partir del conjunto de estaciones que han pasado los dos filtros propuestos ($f=1,2$ y $p=4$), y utilizando como método de interpolación el *kriging*. El de la derecha también se ha interpolado mediante *kriging*, pero usando todo el conjunto de estaciones disponibles. (En blanco aparecen las estaciones no utilizadas en el mapa de la izquierda, por haber sido rechazadas por el método de filtrado).

5. Conclusiones

La utilización de los filtros propuestos permite eliminar estaciones termométricas con medidas anómalas, mejorando la calidad de las cartografías u otros productos elaborados a partir de temperaturas diarias.

No obstante, el proceso de determinación de los umbrales óptimos todavía exige un laborioso proceso manual, que se tratará de automatizar en el futuro.

Bibliografía

Aguilar E, López JM, Brunet M, Saladié O, Sigró X, López D, 1999: Control de calidad y proceso de homogeneización de series térmicas catalanas. *La Climatología española en los albores del siglo XXI* (Raso, Martín-Vide, Eds.), Asociación Española de Climatología, A-1:15-23.

Conrad V, Pollack LW, 1962: *Methods in climatology*. Harvard Univ. Press, Cambridge-Massachussets.

De Ruffray P, Brisse H, Granjean G, 1981: Un procédé d'estimation des données manquantes fondé sur les correlations climatiques entre postes voisins. *La Météorologie*, 24:47-60.

Felicísimo AM, Álvarez MA, 1982: Aplicación del análisis de componentes principales al régimen termopluviométrico de Asturias. *Avances Inv. Bioclim.* (CSIC, Salamanca):565-581.

Guijarro JA, Morey M, 1998: Aplicación de una recta de Teissier a la depuración de una base de datos termopluviométricos mensuales. *Territoris*, 1:157-170.

Jansá JM, 1969: *Curso de climatología*. Servicio Meteorológico Nacional, B-19, 445 pp, Madrid.