

Selección de estaciones de la Red Secundaria del INM: Temperatura.

José Antonio Sosa Cardo
Sección de Estudios y Desarrollos
Centro Meteorológico Territorial en Extremadura. INM.

Introducción:

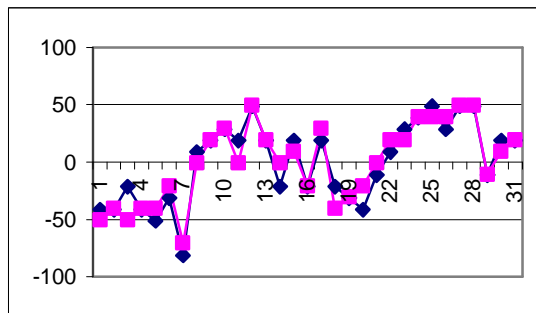
Todos los que en algún momento han trabajado con estaciones de la red secundaria saben que la calidad de los datos en muchas ocasiones no es todo lo buena que se desearía, por eso cuando se pretenden utilizar este tipo de datos se hace casi imprescindible alguna depuración o filtrado, ya sea de tipo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

En este documento se describirán un par de métodos objetivos complementarios de selección de estaciones de la red secundaria.

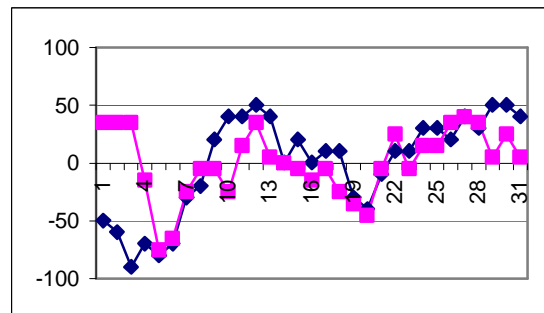
Filtro de comparación de curvas.

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.

Veamos un ejemplo



Una estación avala a la otra porque tienen curvas mensuales parecidas



Las dos estaciones no se avalan entre sí al tener curvas mensuales distintas

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 estos 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:

$$d_{ij} = \sqrt{\sum_{k=1}^n ((x_i^k - \bar{x}_i) - (x_j^k - \bar{x}_j))^2}$$

donde d_{ij} sería la distancia entre las curvas mensuales de las estaciones i y j , x_i^k 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 comprobaría para cada una de las estaciones si existe otra a una distancia menor a un umbral dado, si es así se validan ambas estaciones. Uno de los problemas del método es determinar cual 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 y 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 nuestro parámetro que nos determinará lo tolerante que es la selección. El parámetro f 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 tal forma que un par de estaciones i y j se aceptan como válidas si $d_{ij} < d_{umbral}$

Veamos a continuación unos ejemplos gráficos donde representamos los valores de los datos diarios transportados de la temperatura media de las máximas del mes de enero de 1997 del conjunto de estaciones de la red secundaria del INM disponibles para Extremadura.

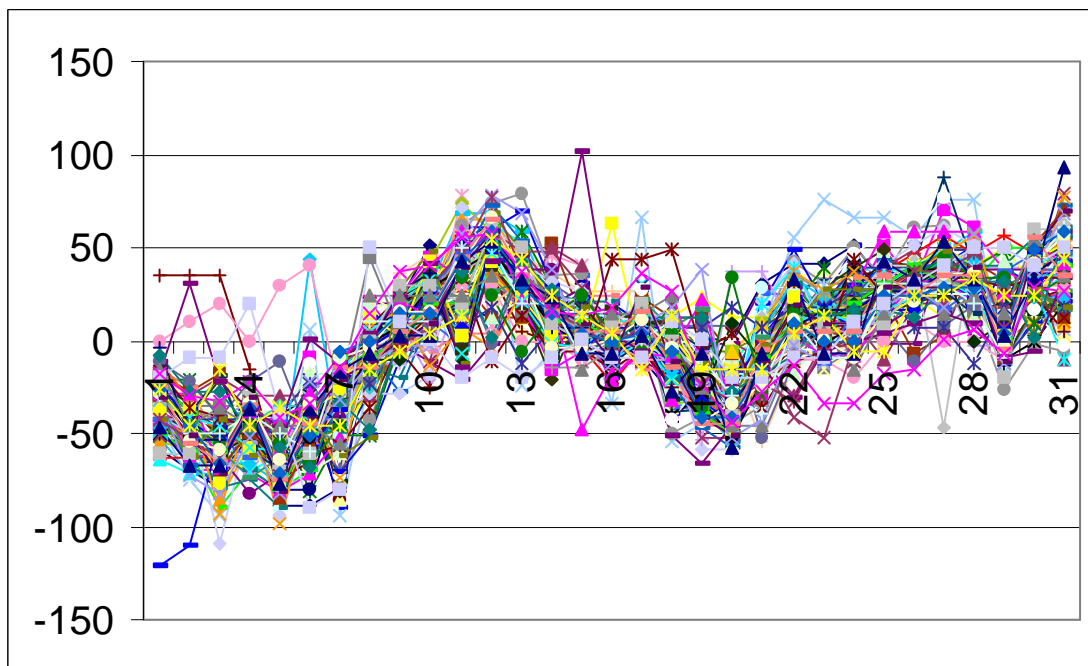


Gráfico 1.- Todas las estaciones del mes (enero 1997)

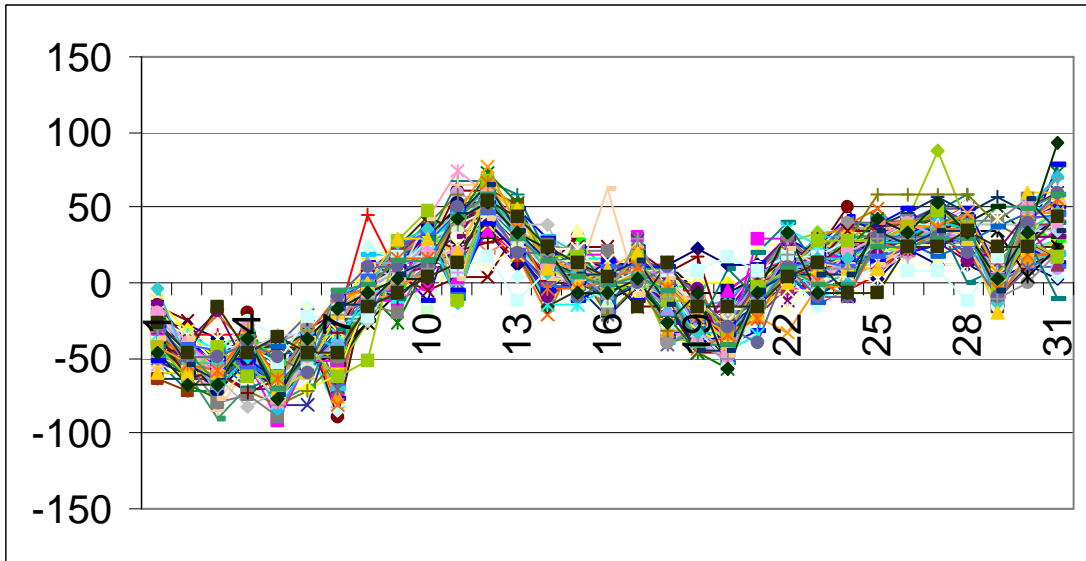


Gráfico 2.- Estaciones que han pasado el filtro con $f=0.9$

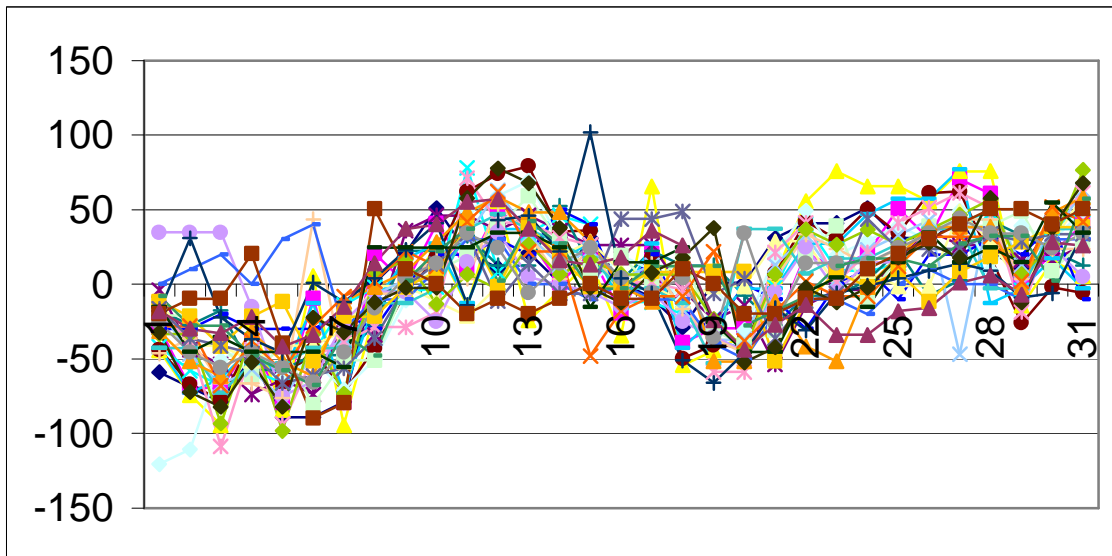


Gráfico 3.- Estaciones que no han pasado el filtro con $f=0.9$

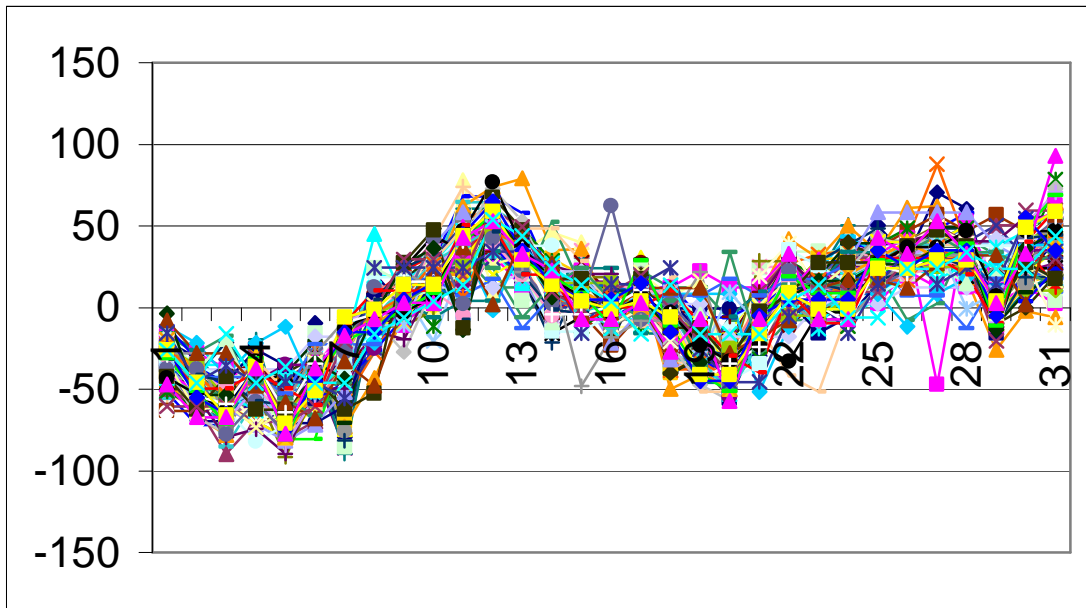


Gráfico 4.- Estaciones que han pasado el filtro con $f=1$

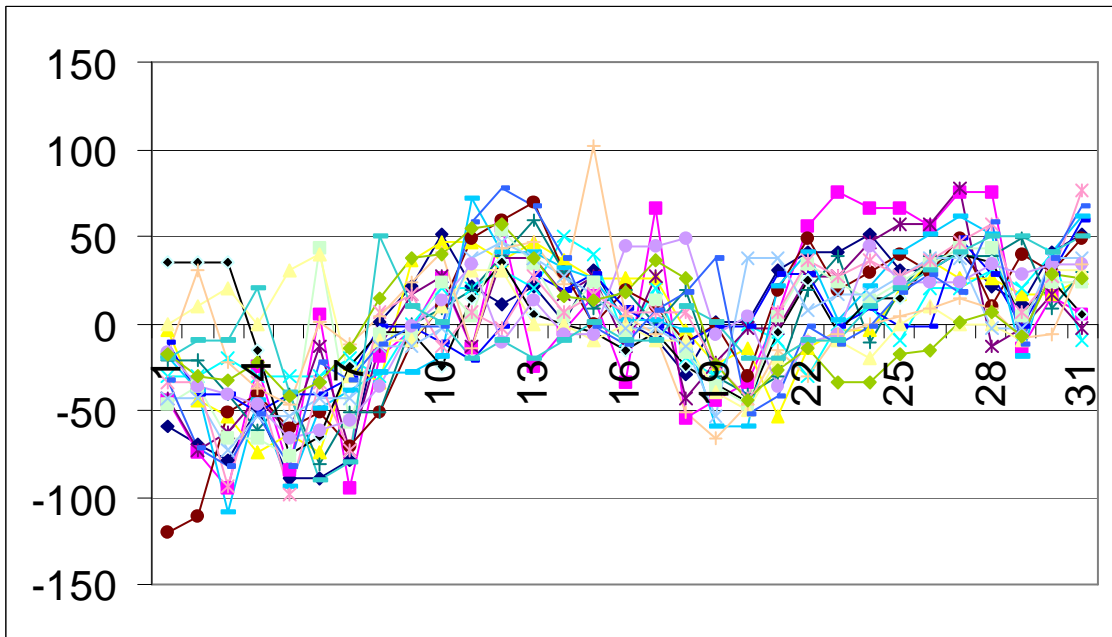


Gráfico 5.- Estaciones que no han pasado el filtro con $f=1$

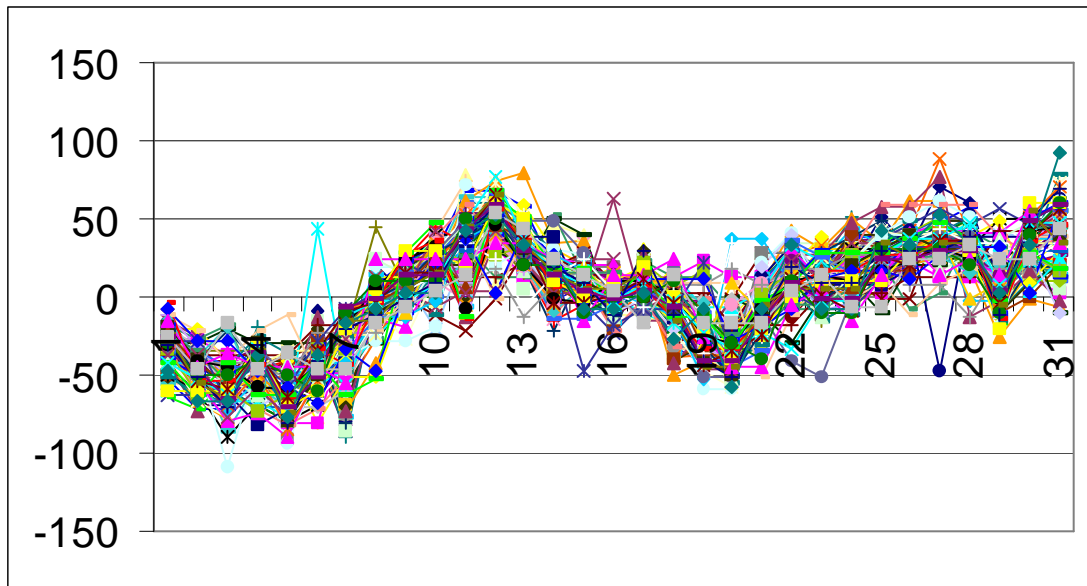


Gráfico 6.- Estaciones que han pasado el filtro con $f=1.1$

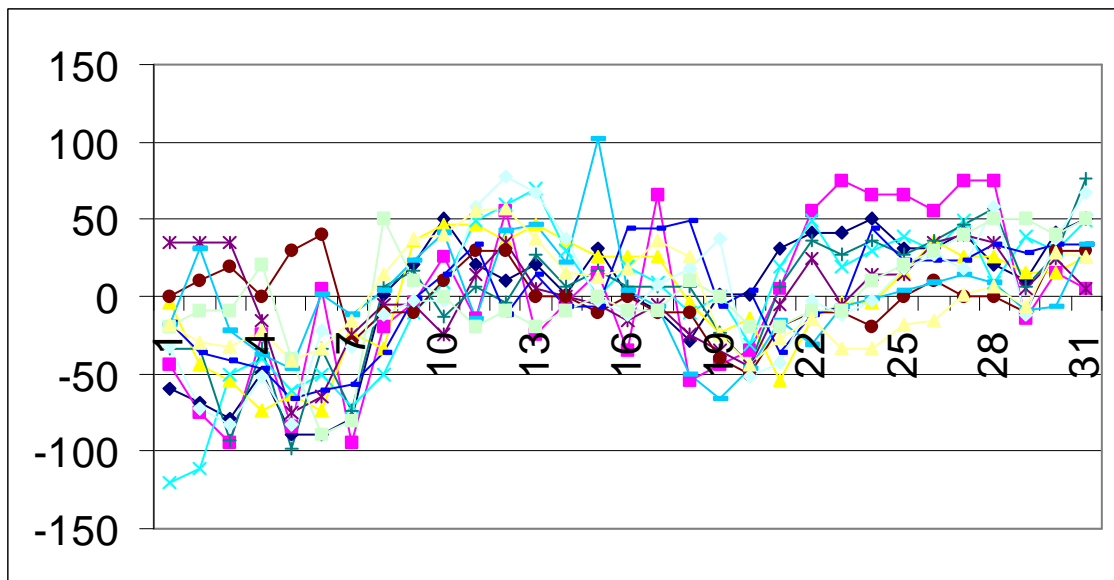


Gráfico 7.- Estaciones que no han pasado el filtro con $f=1.1$

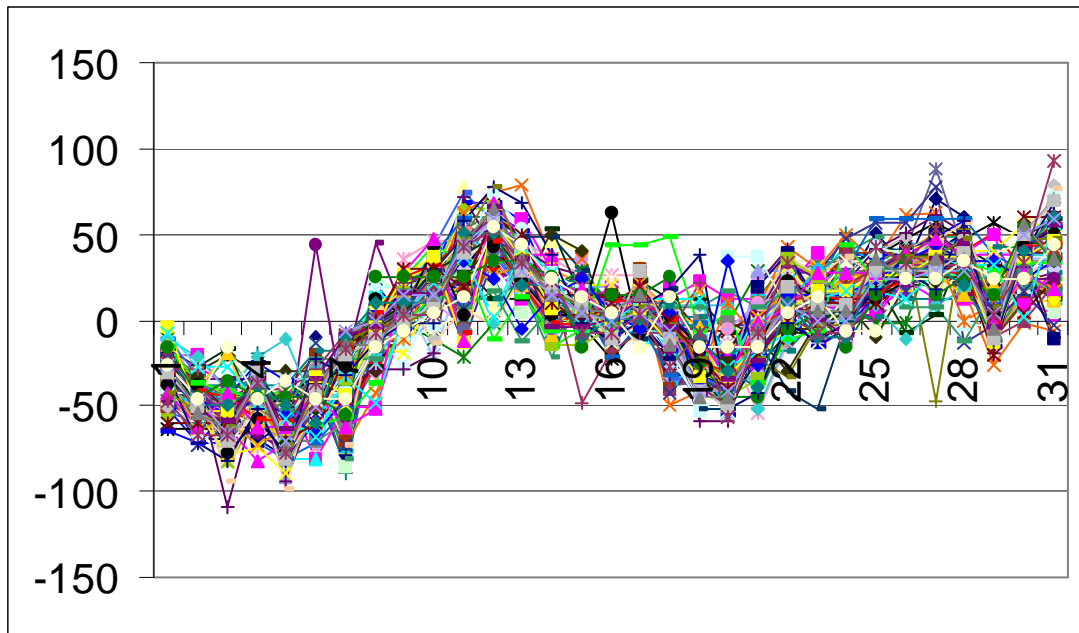


Gráfico 8.- Estaciones que han pasado el filtro con $f=1.2$

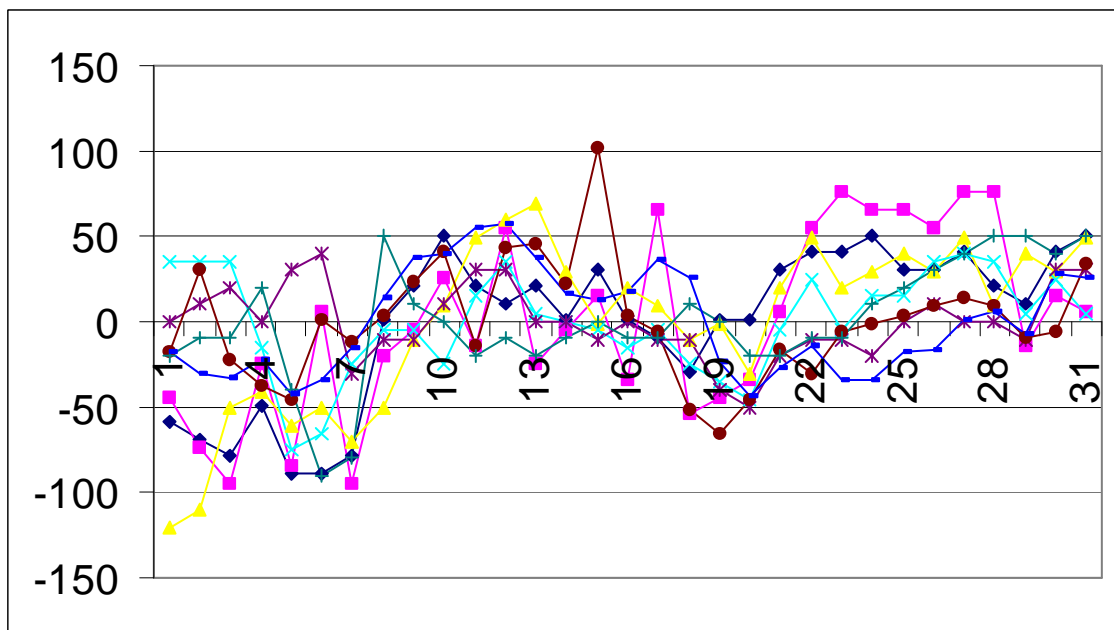


Gráfico 9.- Estaciones que no han pasado el filtro con $f=1.2$

Como puede observarse de los gráficos anteriores, otro de los problemas 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 ponerle al umbral para que filtre este tipo de estaciones es tal que se filtran estaciones que en principio no hay por qué desconfiar de ellas, por eso se propone un segundo filtro complementario a este, que esté especializado sólo en la búsqueda de picos diarios.

Filtro 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 'n' 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:

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

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 y σ_i es el valor de la desviación estándar de los valores diarios de la variable transportada de todas las series y n es el número de días de la ventana temporal elegida.

Para determinar el umbral de este segundo filtro se pueden utilizar gráficos como el siguiente:

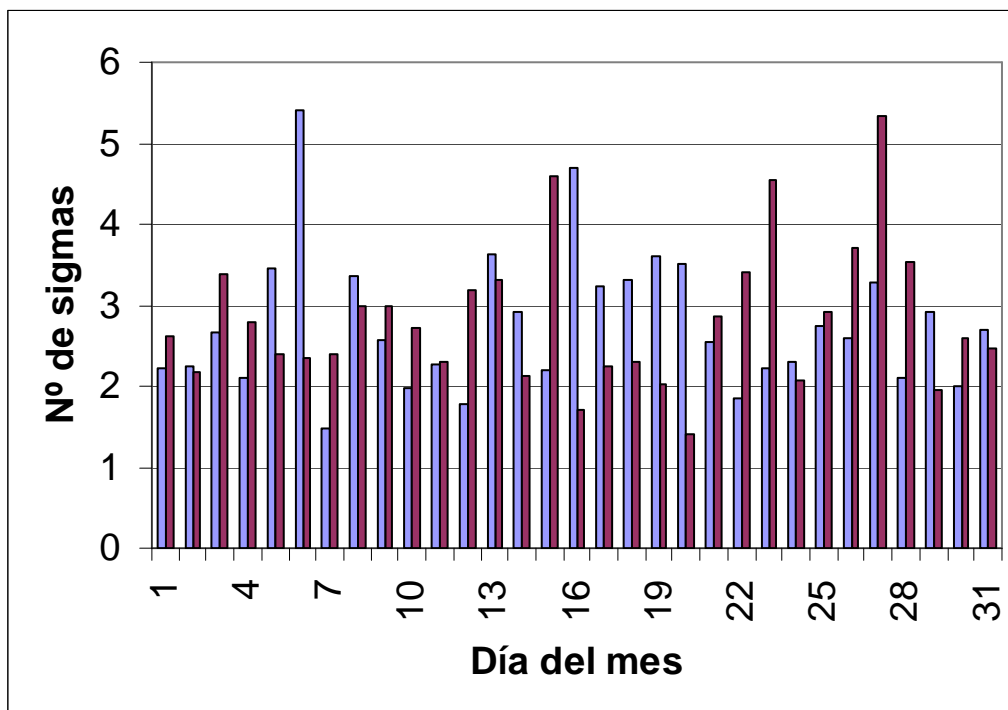


Gráfico 10.- Dispersión de los datos transportados en unidades de desviación estándar diaria

Donde se representa la dispersión de los datos diarios transportados en unidades de desviación estándar diaria.

Puede verse, que para el caso presentado, un buen valor para n en este caso puede ser 4, ya que se eliminarían los valores más extremos.

Combinando ambos filtros, obtenemos la siguiente selección:

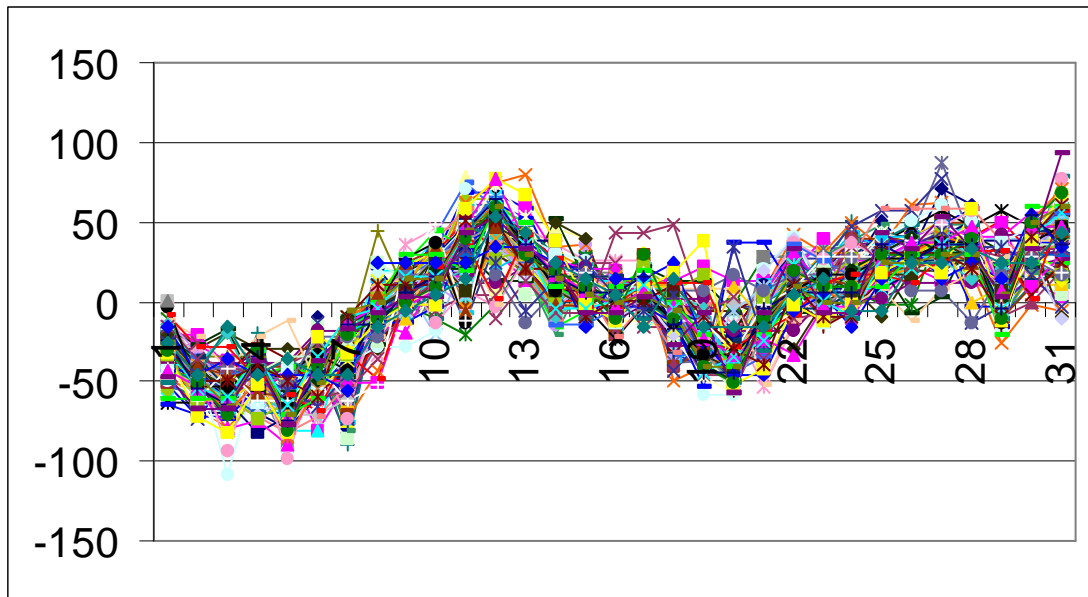


Gráfico 11.- Estaciones que han pasado ambos filtros $f=1.2$ y $n=4$

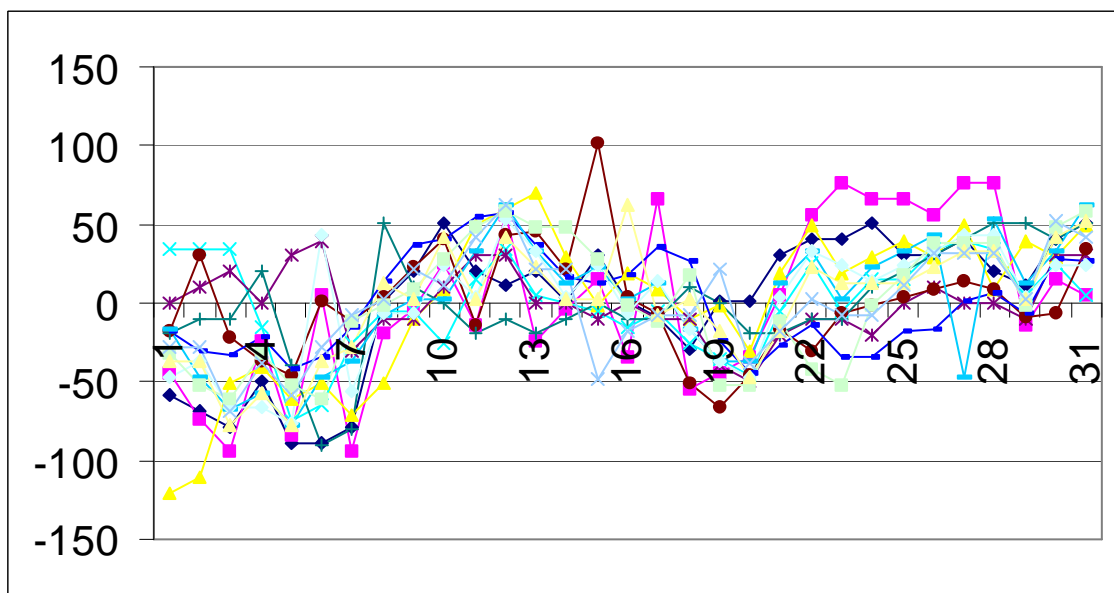
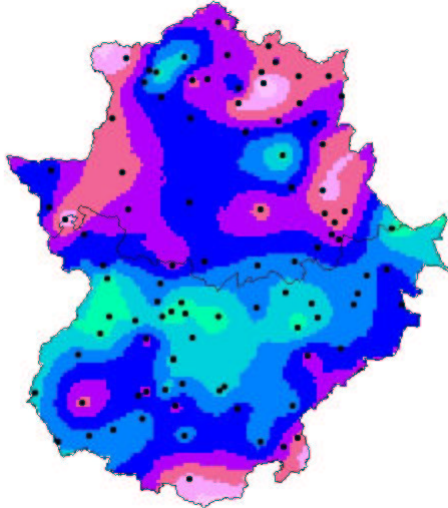


Gráfico 12.- Estaciones que no han pasado alguno de los dos filtros $f=1.2$ y $n=4$

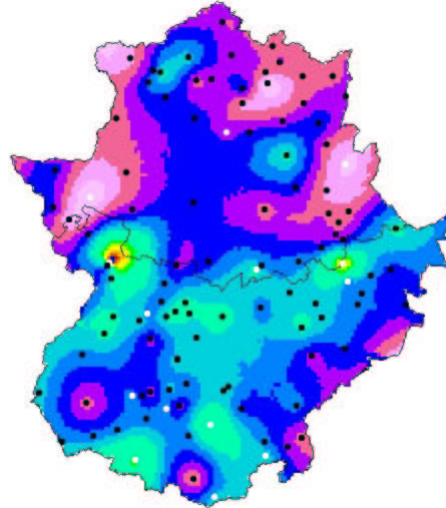
Combinando ambos tipos de filtros obtenemos una selección más lógica desde el punto de vista subjetivo. El principal problema de estos métodos como puede verse es que su calibración ha de hacerse de forma subjetiva. Aunque todo el proceso posterior es objetivo. Si la selección de las estaciones se hace mediante la forma tradicional, las subjetividades son mucho mayores y se suele basar en unas presunciones que no siempre son ciertas.

Resultados Obtenidos

A continuación se presentan dos mapas, interpolados mediante kriging, correspondientes al día 2 de enero de 1997, obtenidos a partir de las estaciones que han pasado los dos filtros propuestos ($f=1.2$ y $n=4$) y a partir de todo el conjunto de estaciones disponible respectivamente:



Temperatura media de las máximas del día 2-1-1997, obtenida a partir del conjunto de estaciones que han pasado los dos filtros propuestos ($f=1.2$ y $n=4$) y utilizando como método de interpolación el kriging.

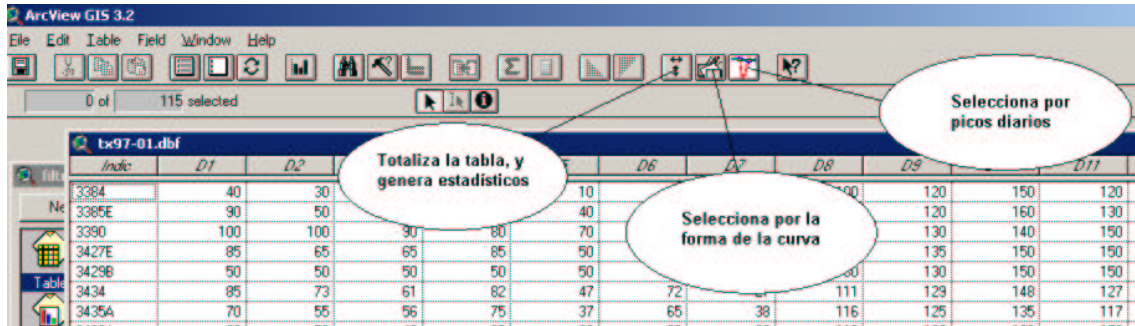


Temperatura media de las máximas del día 2-1-1997, obtenida a partir de todo el conjunto de estaciones disponibles y utilizando como método de interpolación el kriging. En blanco aparecen las estaciones que los filtros utilizados han desechado.

Como puede observarse el grid obtenido tras el filtrado es mucho más creíble, ya que está es mucho más acorde con la geografía subyacente que el obtenido utilizando todo el conjunto de estaciones, aportando estas un ruido innecesario y molesto a la hora de realizar cualquier tipo de estudio.

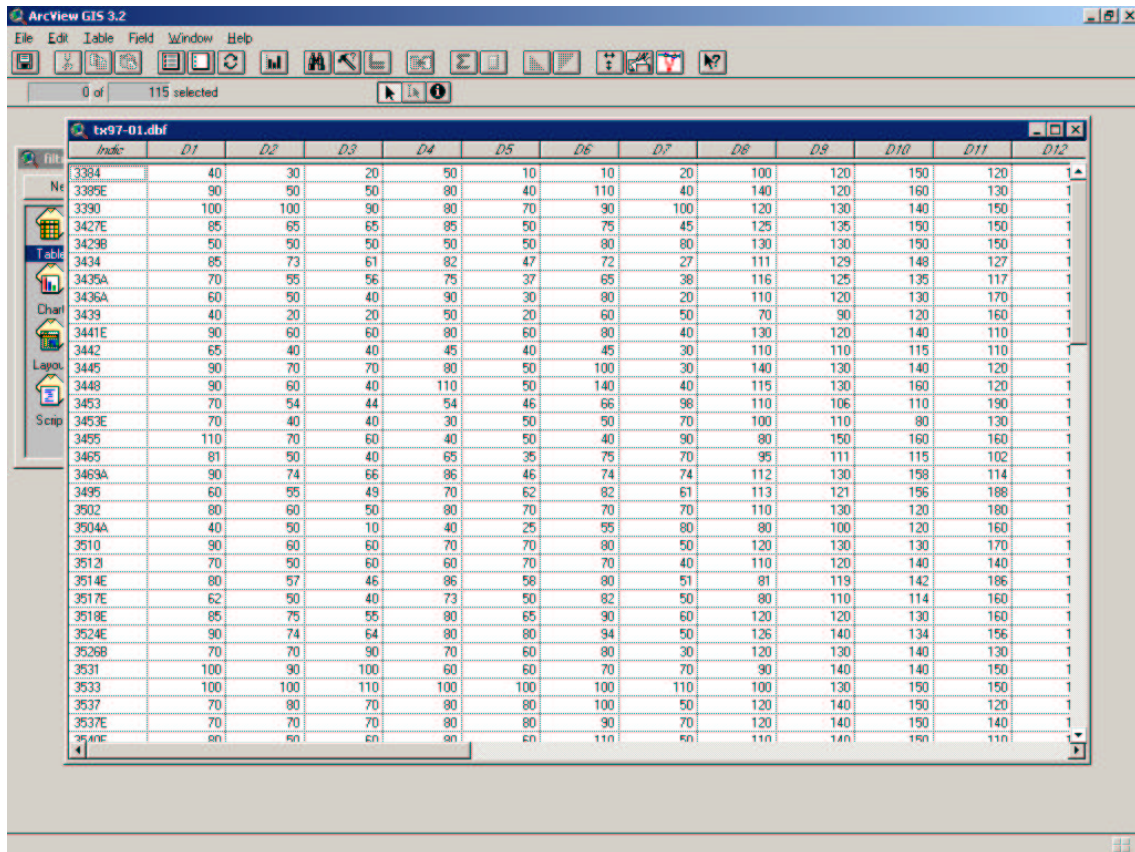
Integración en ArcView.

La implementación práctica de los algoritmos descritos se ha realizado mediante el lenguaje de programación de ArcView: Avenue y se han integrado en el panel de botones para la utilización con tablas de datos.



Estas tablas de datos deberán tener la siguiente estructura:

Cada fichero de tabla de datos debe corresponder a un mes, cada registro contiene los datos diarios de cada estación. En el primer campo de los registros debe figurar el indicativo de la estación, siguen tantos campos como días tenga el mes en cuestión.



Con las versiones actuales de los scripts el proceso de ejecución sería el siguiente:

1º.- Se debe ejecutar el script 'estadísticos', que calcula los estadísticos necesarios para los filtros y añade dos registros al final y un nuevo campo 'Mensual' a la tabla. Si la tabla está ya totalizada el scrip generará un mensaje y dará opción a totalizar nuevamente o dejarla tal como está.

2°.- Se ejecuta el script 'selecciona', que hace una selección de estaciones por la forma de la curva mensual. Estos registros seleccionados deben exportarse a un fichero de tabla de datos .dbf (File+Export) para después pasarles el siguiente filtro.

3°.- Se volvería a ejecutar el script 'estadísticos', para generar los nuevos estadísticos necesarios para el siguiente filtro.

4°.- Se ejecuta el script 'picos', que realiza la segunda de las selecciones propuestas en este documento, es decir eliminando las estaciones que presenten picos diarios en sus curvas demasiado pronunciados.

5°.- Se exportan los registros seleccionados, mediante el procedimiento anteriormente descrito a un nuevo fichero de tabla de datos (.dbf). en este fichero estarán las estaciones que han pasado ambos filtros.

Limitaciones de los métodos y futuros desarrollos.

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. Esos errores, en principio, la única manera de identificarlos sería con un modelo espacial lo suficientemente bueno como para que dejaran 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, quedaría por lo tanto desarrollar un método de establecimiento de umbrales óptimos de forma objetiva.

Conclusiones.

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.

- 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.
- 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 desarrollados permiten eliminar ruido indeseado y tener cierta certeza de que las estaciones empleadas ofrecen una cierta garantía.