

ANÁLISIS DE TENDENCIAS EN TEMPERATURAS EXTREMAS POR ESTADÍSTICOS DE ORDEN EXTREMALES

José A. LÓPEZ DÍAZ
Instituto Nacional de Meteorología
(jalopez@inm.es)

RESUMEN

En este trabajo se presenta un estudio de las tendencias en temperaturas extremas máximas y mínimas utilizando un ajuste de máxima verosimilitud a la función general de valores extremos. Se ha tomado como ejemplo para ilustrar esta técnica la serie de León (Virgen del Camino). Se ha efectuado en primer lugar un análisis de las tendencias por medio de un ajuste a los extremos anuales. Esto se completa con un ajuste a los conjuntos de valores extremos del año de hasta cinco elementos. Por último se ha pretendido mejorar este enfoque agrupando las temperaturas extremas diarias en clusters antes de hallar los extremos anuales.

Palabras clave: Tendencia de extremos, cluster de extremos

ABSTRACT

In this paper a study of trends in extreme temperatures is presented that uses a maximum likelihood fit to the generalised extreme value function. The technique has been illustrated with the series of León (Virgen del Camino). In the first place the trend has been estimated through the annual extreme values. This has been complemented through a fit to the sets of annual extremes of size up to five elements per year. Finally an attempt has been made to improve on the above approach by clustering the daily extremes before computing the annual extremes.

Key words: Trends in extremes, extreme cluster

1. INTRODUCCIÓN

El análisis de tendencias en series de extremos es un campo de creciente interés al ser los extremos climáticos un elemento de importancia creciente en la caracterización del régimen climático y sus posibles variaciones. Hasta ahora ha sido común el análisis de las tendencias en series de extremos mediante test de tendencia globales no paramétricos como el de Mann para la evaluación de la significación estadística, y estimadores de la tendencia como el basado en la regresión lineal u otros más robustos como el basado en el test de Mann que se presenta más adelante. Sin embargo de esta forma no se aprovecha la peculiar naturaleza estadística de los extremos que como se sabe está bien fundamentada teóricamente en el sentido de la convergencia asintótica de la ley de distribución hacia la función general de valores extremos. Se puede plantear un test de tendencia comparando el logaritmo de la función de verosimilitud para ajuste a un modelo sin tendencia frente a otro que incluya una tendencia en el parámetro de localización. Además es posible usar la teoría de extremos para

ajustar una función que extiende la FGVE para el caso de varios extremos por bloque (en este caso año). Así se puede ajustar a la serie formada por los dos valores más altos (o bajos según el caso) por año, o en general k extremos por año. Y también mediante la comparación de las log verosimilitud evaluar la tendencia.

Cuando se trabaja con varios extremos por año un potencial problema estriba en la posible dependencia entre extremos del mismo año. Esta no independencia viola una de los supuestos para la validez del modelo asintótico de la teoría de extremos. Por tanto parece conveniente investigar la posibilidad de incrementar la independencia de los extremos a base de agrupar los extremos diarios en clusters.

En este trabajo se han aplicado los procedimientos descritos a la serie de temperaturas máximas y mínimas de León. Esta serie comienza en 1938 y presenta la particularidad de tener tendencia decreciente en las máximas anuales.

2. TEMPERATURA MÁXIMA DE LEÓN

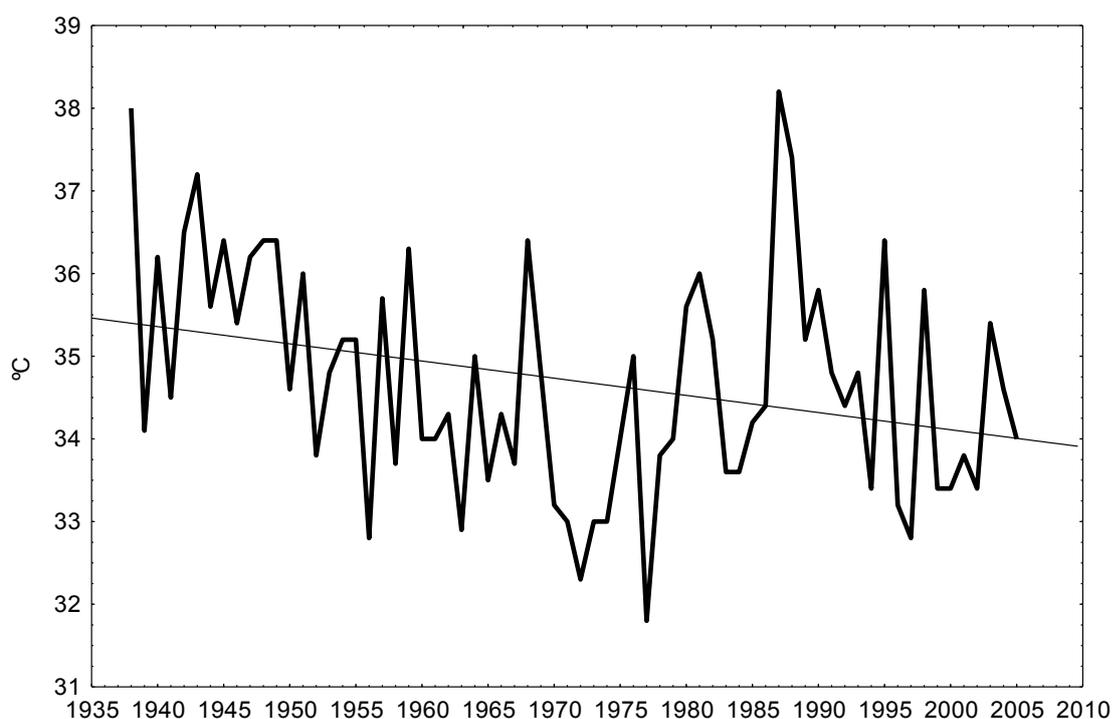


Fig. 1: Serie de temperaturas máximas anuales de León

La serie de temperaturas máximas anuales de la estación de León (Virgen del Camino) comprende el periodo 1938-2005 y presenta la peculiaridad de mostrar una tendencia decreciente global con el test de Mann-Kendall, con una pendiente de la recta de regresión de -2.1 °C / 100 años. En la tabla 1 se presenta también la pendiente del estimador de tendencia de Mann que es un estimador robusto de la tendencia inspirado en el test de Mann (Kunkel *et al* 1999). Se define como la mediana de las pendientes que determinan todos los pares de elementos de la serie dada. En el caso de León es de -2.3 °C /100 años. Esta tendencia

ANÁLISIS DE TENDENCIAS EN TEMPERATURAS EXTREMAS POR ESTADÍSTICOS DE ORDEN EXTREMALES

decreciente parece estar bastante condicionada por los años al comienzo de la serie. Por ejemplo en Salamanca Matacán, con serie comenzando en 1945, hay tendencia también decreciente en las temperaturas máximas (pendiente de regresión $-2.1\text{ }^{\circ}\text{C} / 100\text{ años}$) pero no es globalmente significativa con el test de Mann. En León para ese mismo periodo tampoco hay tendencia global significativa. En cambio en Burgos Villafraía las temperaturas máximas muestran una tendencia global creciente significativa al 5% con pendiente $2.6^{\circ}\text{C} / 100\text{ años}$ ($2.9\text{ }^{\circ}\text{C} / 100\text{ años}$ con el estimador de Mann).

Pend. Reg. Lin. ($^{\circ}\text{C} / 100\text{ a}$)	Pend. Mann ($^{\circ}\text{C} / 100\text{ a}$)	p-valor Mann
-2.08	-2.27	0.009

Tabla 1: ESTADÍSTICOS DE TENDENCIA GLOBAL DE LAS TEMPERATURAS MÁXIMAS EN LEÓN

El análisis de la tendencia basado en un test no paramétrico como el de Mann no aprovecha la particular estructura estadística de los datos al ser estos extremos anuales. Para sacar partido de este hecho hemos de movernos al terreno paramétrico y utilizar el hecho de que los extremos anuales se distribuyen asintóticamente (es decir suponiendo que el año contuviera infinitas realizaciones) siguiendo la distribución general de valores extremos (GVE). La distribución GVE contiene tres parámetros, uno de forma k , uno de escala α y otro de localización ξ . La función de distribución de la GVE toma la forma (Maidment 1993):

$$G(x) = \exp \left\{ - \left[1 - k \frac{x - \xi}{\alpha} \right]^{1/k} \right\}$$

definida para $1 + k(x - \xi)/\alpha > 0$. Aquí se ha usado la convención habitual en hidrología y climatología para el signo de k .

Nº extre./bloque	k	b ($^{\circ}\text{C} / 100\text{ a}$)	$2(1 - I_{b=0})$	p-valor $b \neq 0$
1	0.16	0	-----	-----
1	0.18	-2.12	5.49	0.02
2	0.16	0	-----	-----
2	0.16	-1.38	5.10	0.02
3	0.25	0	-----	-----
3	0.25	-1.03	3.68	0.06
4	0.27	0	-----	-----
4	0.27	-0.97	2.88	0.09
5	0.28	0	-----	-----
5	0.28	-0.96	4.30	0.04

Tabla 2: ANÁLISIS DE TENDENCIA EN TEMPERATURA MÁXIMA POR MEDIO DE ESTADÍSTICOS DE ORDEN

Se puede contrastar la presencia de una tendencia comparando el ajuste del modelo anterior a los datos con el modelo que incluye una tendencia lineal en el parámetro de localización ξ . De esta forma el valor de ξ en el año i -ésimo será $a + b i$ con a y b constantes a determinar por el ajuste. Se puede plantear el test de hipótesis con hipótesis nula $b = 0$ y alternativa $b \neq 0$. Un estadístico adecuado a este test se obtiene comparando el valor de la log-verosimilitud para el modelo con $b = 0$ y con b variable. En efecto se tiene que $2 (\log L(b) - \log L(b = 0))$ se distribuye bajo la hipótesis nula siguiendo una ley χ^2 con 1 grado de libertad (donde $L(b)$ se refiere al modelo que incluye la tendencia b). De esta forma se han obtenido las primeras filas de la tabla 2, en la que se observa que el estimador de máxima verosimilitud de b es -2.12 °C /100 años, muy similar al obtenido por regresión lineal y que la tendencia en ξ es significativa la 5% (el p-valor del test anteriormente descrito es 0.02).

Se puede teóricamente mejorar el anterior análisis a base de aumentar el tamaño de la muestra, lo que indudablemente debe conducir a un test más determinante. Para ello se puede efectuar el ajuste de una función análoga a la GVE para un conjunto de valores máximos en cada año, es decir, para los 2 valores más altos, o los 3 más altos, etc. Para este caso la función de densidad tiene la forma (Coles 2001):

$$f(x^{(r)}, x^{(r-1)}, \dots, x^{(1)}) = \exp \left\{ - \left[1 - k \frac{x^{(r)} - \xi}{\alpha} \right]^{1/k} \right\} \times \prod_{j=1}^r \alpha^{-1} \left[1 - k \frac{x^{(j)} - \xi}{\alpha} \right]^{1/k-1}$$

donde $x^{(r)} \leq x^{(r-1)} \leq \dots \leq x^{(1)}$ son los r valores más altos de cada año. La convergencia de los r valores más altos hacia esta forma de distribución es más lenta que para el máximo, pero a cambio se aumenta el número de datos sobre los que se hace el ajuste. De forma análoga al caso anterior se puede plantear un test de hipótesis para contrastar una tendencia lineal en el parámetro de localización ξ utilizando la diferencia de log-verosimilitudes. Los resultados de este análisis se encuentran en la tabla 2, para un número de valores por año desde 2 a 5.

Se aprecia un gradual aumento en el valor del parámetro de forma k , lo cual es un elemento en contra de la validez del modelo ya que este parámetro debería ser en teoría el mismo independientemente del número de valores por año si el ajuste es bueno. También observamos que la estimación de la pendiente b disminuye gradualmente, hasta que para 5 datos por año se reduce a -0.96 °C/año, la mitad del valor para el máximo anual. Sin embargo los p- valores del test de tendencia son similares al máximo anual aunque resultan ser no significativos al 5% para 3 y 4 datos por año.

3. ANÁLISIS POR CLUSTERS DE LA TENDENCIA EN TEMPERATURAS MÁXIMAS

Una posible causa de que el ajuste con varios elementos por año sea peor es que no se ha tenido en cuenta el efecto de cluster en los datos. Este efecto surge de la tendencia de los valores más altos a aparecer juntos como consecuencia de situaciones sinópticas que perduran varios días. Si sucede esto estos valores estarán correlacionados y se viola uno de los supuestos del modelo teórico que supone observaciones independientes en cada bloque.

Para definir los clusters de temperaturas máximas diarias se ha adoptado aquí un criterio basado en las diferencias entre días sucesivos. Esto tiene la ventaja de permitir un criterio sencillo de aplicar ya que no incluye umbrales para cada día. La idea es considerar que varios días sucesivos forman un cluster si las diferencias sucesivas son pequeñas. El cluster se romperá si el total de diferencias (absolutas) excede determinado valor o bien una diferencia es suficientemente grande. En concreto se han adoptado los dos criterios siguientes para la formación de un cluster:

- a) el valor medio de las diferencias sucesivas absolutas no puede exceder el valor medio de las diferencias sucesivas absolutas a lo largo de toda la serie diaria
- b) una diferencia absoluta no puede ser superior al cuantil 0.7 de la serie de diferencias absolutas.

Se han formado entonces los clusters de cada año y se han tomado los valores más altos para efectuar el análisis por bloques anterior. Los resultados se recogen en la tabla 3. Apreciamos que también ahora aumenta el valor de k de 0.16 para $n=1$ hasta 0.32 para $n=5$ elementos por año, lo cual parece indicar que el ajuste del modelo no ha mejorado sustancialmente con relación al caso anterior. Por otra parte los resultados del test de tendencia son diferentes del caso anterior, apreciándose que ahora ningún test con más de un elemento por bloque es significativo al 5%, y eso que los estimadores de la tendencia en el parámetro de localización son algo más grandes que en el análisis sin clusters.

Nº extre./bloque	k	b (°C /100 a)	$2(1 - I_{b=0})$	p-valor $b \neq 0$
1	0.16	0	-----	-----
1	0.18	-2.12	5.49	0.02
2	0.25	0	-----	-----
2	0.26	-1.08	1.43	0.23
3	0.29	0	-----	-----
3	0.28	-1.38	3.49	0.06
4	0.30	0	-----	-----
4	0.30	-1.22	2.90	0.09
5	0.32	0	-----	-----
5	0.32	-1.13	2.07	0.15

Tabla 3: ANÁLISIS DE TENDENCIA TEMPERATURAS MÁXIMAS POR MEDIO DE ESTADÍSTICOS DE ORDEN CON CLUSTERS

4. TEMPERATURA MÍNIMA DE LEÓN

La serie de temperaturas mínimas anuales de León presenta la esperada tendencia creciente como se aprecia en la figura 2 y en la tabla 4. La pendiente de la recta de regresión es $3.27\text{ }^{\circ}\text{C} / 100$ años y la del estimador de Mann de $3.33\text{ }^{\circ}\text{C} / 100$ años, con un p-valor para el test de Mann de 0.03, por tanto significativa al 5%.

Se ha procedido a un análisis de tendencias paralelo al descrito antes para las temperaturas máximas. En primer lugar se ha comparado el ajuste de máxima verosimilitud a la función GVE sin tendencia en el parámetro de localización y con tendencia en el mismo. Los resultados se recogen en la tabla 5. Se observa que este análisis no muestra con claridad una

tendencia significativa ya que para 2 y 5 elementos por bloque sí da tendencia significativa, pero para el resto (1, 3 y 4) no. Además el estimador del parámetro de forma k (de 0.02 a 0.11) también aumenta al incrementar el número de elementos por bloque, en forma análoga a las máximas, aunque de forma menos acentuada. Esto permite suponer un mejor ajuste de los modelos. El estimador de la pendiente en el parámetro de localización disminuye con el número por bloque, pero menos que para las máximas.

Pend. Reg. Lin. (°C/ 100 a)	Pend. Mann (°C /100 a)	p-valor Mann
3.27	3.33	0.03

Tabla 4: ESTADÍSTICOS DE TENDENCIA GLOBAL DE LAS TEMPERATURAS MÍNIMAS EN LEÓN

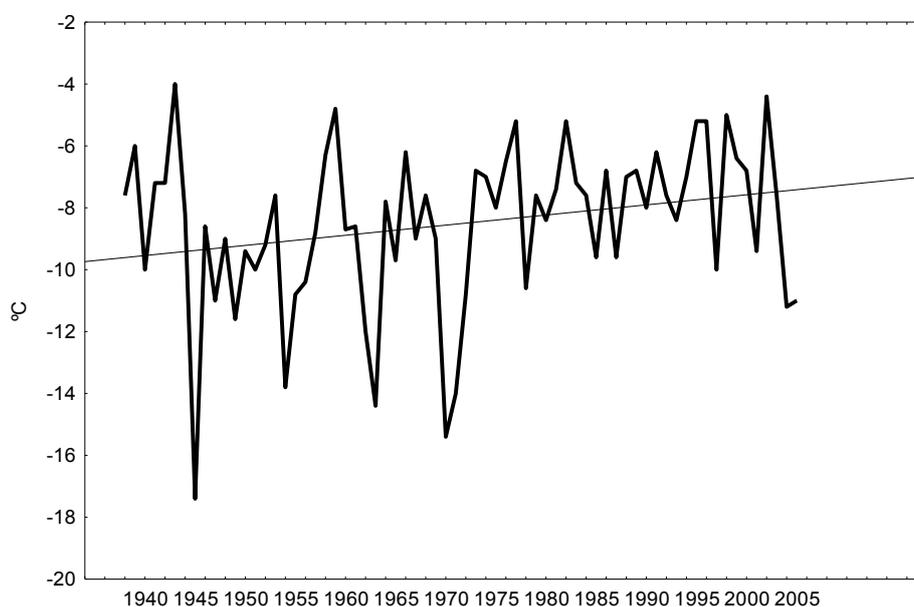


Fig. 2: Serie de temperaturas mínimas anuales de León

Nº extre./bloque	k	b (°C /100 a)	$2(l - l_{b=0})$	p-valor $b \neq 0$
1	0.02	0	-----	-----
1	0.04	1.91	2.31	0.13
2	0.07	0	-----	-----
2	0.10	1.94	4.10	0.04
3	0.09	0	-----	-----
3	0.10	1.24	3.06	0.08
4	0.10	0	-----	-----
4	0.11	1.27	3.26	0.07
5	0.10	0	-----	-----
5	0.11	1.55	4.96	0.03

Tabla 5: ANÁLISIS DE TENDENCIA DE LAS TEMPERATURAS MÍNIMAS POR MEDIO DE ESTADÍSTICOS DE ORDEN

5. ANÁLISIS POR CLUSTERS DE LA TENDENCIA EN TEMPERATURAS MÍNIMAS

Los resultados del análisis por clusters de la tendencia en temperaturas mínimas se recogen en la tabla 6, que muestra un contraste evidente con el análisis anterior sin clusters. A medida que crece el número de mínimos por bloque considerados la tendencia creciente en las mínimas es más significativa estadísticamente. Así el p-valor para un dato por año es 0.13, pero para 5 datos por bloque es 0.0009. Además el estimador de la tendencia b es bastante más estable que en la tabla 5. Estos resultados son más acordes con lo que cabría esperar teóricamente, ya que al aumentar el número de datos por bloque se aumenta el tamaño de la muestra y por consiguiente si hay verdaderamente tendencia esta tiene que resultar más significativa cuantos más elementos por bloque se incluyan en el análisis

Nº extre./bloque	k	b (°C /100 a)	$2(l - l_{b=0})$	p-valor $b \neq 0$
1	0.02	0	-----	-----
1	0.04	1.91	2.31	0.13
2	0.12	0	-----	-----
2	0.13	2.08	5.37	0.02
3	0.12	0	-----	-----
3	0.13	1.69	7.16	0.007
4	0.13	0	-----	-----
4	0.15	2.73	11.7	$6 * 10^{-4}$
5	0.10	0	-----	-----
5	0.11	2.10	10.97	$9 * 10^{-4}$

Tabla 6: ANÁLISIS DE TENDENCIA POR MEDIO DE ESTADÍSTICOS DE ORDEN CON CLUSTERS

6. DISCUSIÓN DE LOS RESULTADOS

En el análisis de las tendencias en las temperaturas máximas no está claro que la introducción de los clusters haya conducido a un análisis más preciso. En realidad la tendencia decreciente lejos de perfilarse al aumentar el número de datos por año se va diluyendo en el estudio con clusters, en el otro se mantiene. Podemos sospechar que aquí lo que sucede es que el número de observaciones independientes reales por año no es suficiente para incluir varios valores máximos por año. Este efecto podría incluso acentuarse al introducir los clusters, ya que en ese caso los distintos máximos están más separados. Algún indicio a favor de esta tesis nos lo proporciona el hecho de que el valor del parámetro k aumenta considerablemente con el número de máximos por año con clusters o sin ellos. Este argumento se puede apoyar haciendo un ajuste de la función GVE para máximos de años agrupados de dos en dos. En teoría el parámetro k no debería variar. Los valores de k obtenidos para un máximo por grupo (dos años), dos por grupo y 5 por grupo son respectivamente 0.16, 0.23, 0.24 con clusters, y valores similares sin ellos. El crecimiento con la agrupación es apreciablemente menor, así que efectivamente parece que un año no proporciona un número suficiente de observaciones de temperatura máxima realmente independientes como para tomar hasta 5 máximos por año.

La situación es diferente para las temperaturas mínimas. Aquí el aumento de k es más moderado y la tendencia se acentúa considerablemente al aumentar el número de datos por

año considerando clusters. En cambio se mantiene vacilante si no se consideran clusters. Así que en el caso de las mínimas obtenemos un mejor acuerdo con lo que teóricamente cabría esperar y los clusters consiguen aumentar la independencia efectiva entre los mínimos del año sin llegar a manifestarse el fenómeno de escasez efectiva de observaciones en el año.

REFERENCIAS

COLES, S. (2001). An Introduction to Statistical Modeling of Extremes, Springer Series in Statistics, ISBN 1852334592

KUNKEL, E.K. y ANDSAGER, K.. (1999). “Long-term trends in extreme precipitation events over coterminous United States and Canada”, *Journal of Climate*, 12, pp. 2515-2527

MAIDMENT, D.R, (ed) (1999). Handbook of Hydrology, McGraw-Hill.