

# **VALIDACIÓN DE PROPIEDADES DE CLUSTERING DE EVENTOS EXTREMOS DE PRECIPITACIÓN MEDIANTE EL ÍNDICE EXTREMAL EN TRES MODELOS CLIMÁTICOS CON DOWNSCALING.**

José Antonio LÓPEZ DÍAZ  
Programa de Técnicas Climáticas, AEMET  
[jalopez@inm.es](mailto:jalopez@inm.es)

## **RESUMEN**

Mediante la aplicación del índice extremal se ha efectuado una clasificación de las ocurrencias de precipitación diaria en percentiles extremos en clusters, tanto para las series observadas en un conjunto de observatorios en el periodo 1961-1990 como para las series simuladas por la técnica de downscaling estadístico por análogos de la FIC a partir de tres modelos climáticos generales (ECHAM4, HadAM3, CGCM2). Aplicando una técnica de bootstrapping se han generado muestras simuladas de las realizaciones de clusters para las observaciones y los modelos que permiten valorar estadísticamente las diferencias en cuanto a propiedades de clustering de eventos extremos. Esta técnica se ha aplicado por separado al periodo oct-nov y mar-abr. Se observan patrones espaciales bastante coherentes en cada modelo, con diferencias significativas entre los modelos en su grado de aproximación a las propiedades de clustering observadas.

**Palabras clave:** Extremos, Precipitación, Índice Extremal, Cluster

## **ABSTRACT**

A classification of the occurrences of extreme daily precipitation has been carried out using the extremal index, both for observed series in a number of observatories through 1961-1990 and for the simulated series generated by statistical downscaling with the analogue technique of the FIC on the basis of three GCM's (ECHAM4, HadAM3, CGCM2). In order to appraise the statistical differences regarding the clustering of extreme events a bootstrapping technique has been used, whereby synthetic cluster series for observations and models were generated. This procedure has been applied both to Oct-Nov and Mar-Apr. One can see pretty coherent spatial patterns for each model, with significant differences between models in their degree of approximation to the observed clustering properties.

**Key words:** Extremes, Precipitation, Extremal Index, Cluster

## **1. EL ÍNDICE EXTREMAL**

### **1.1 Definición**

En este trabajo las propiedades de clustering de los sucesos extremos diarios consistentes en la superación de un umbral extremo de precipitación o temperatura se han analizado basándose fundamentalmente en el índice extremal. Este índice da idea de la desviación de la

hipótesis de variables aleatorias independientes e idénticamente distribuidas que es la más sencilla posible para un proceso estacionario, esto es, con propiedades estadísticas invariantes frente a traslaciones temporales.

Más concretamente, si  $X(n)$  es una sucesión estacionaria de variables aleatorias con función de distribución  $F$  (y  $F^* = 1 - F$ ), definimos  $M(n) = \max ( X(i), 1 \leq i \leq n)$ . Decimos que el proceso  $X(n)$  tiene un índice extremal  $\theta \in [0, 1]$  si para cada  $\tau > 0$  existe una sucesión  $u(n)$  tal que para  $n \rightarrow \infty$ ,

$$a) \quad n F^*(u(n)) \rightarrow \tau$$

$$b) \quad P ( M(n) \leq u(n) ) \rightarrow \exp (- \theta \tau)$$

La condición a) implica que en promedio de  $n$  realizaciones de la variable aleatoria  $X(i)$  se supera el umbral  $u(n)$  en un número  $\tau$  (fijo) de casos, cuando  $n$  crece indefinidamente. Entonces si el proceso fuera aleatorio simple, es decir, con variables aleatorias independientes e idénticamente distribuidas, en la condición b) tendríamos  $\theta = 1$ . Esto se deduce de que en este caso la variable aleatoria número de excedencias del umbral  $u(n)$  tendría una distribución límite de Poisson con parámetro  $\tau$ . A medida que  $\theta$  se acerca a 0 más tendencia al clustering de los eventos de superación del umbral existe.

## 1. 2 Tiempo entre excedencias de un umbral extremo

Otra forma de captar el significado del índice extremal es analizando el proceso de tiempos entre excedencias de un umbral extremo. Definamos la variable aleatoria  $T(u)$  como el tiempo que hay que esperar entre una superación del umbral  $u$  y la siguiente superación, es decir,  $T(u) = \min \{ n \geq 1: X(n+1) > u \}$  dado que  $X(1) > u$ .

En el caso de v.a. independientes deducimos de  $P ( T(u) > n ) = F(u)^n$  que

$$\lim ( P ( F^*(u) T(u) > x ) ) = \exp (-x)$$

cuando  $u$  tiende a  $\omega = \sup \{ x: F(x) < 1 \}$ , finito o infinito. Por tanto  $F^*(u) T(u)$  es asintóticamente estándar exponencial, como corresponde a un proceso de Poisson.

Para el caso de un proceso con índice extremal  $\theta$  se cumple:

$$\lim ( P ( F^*(u) T(u) > x ) ) = \theta \exp (- \theta x)$$

cuando  $x$  tiende a  $\omega$ . Esto significa que la distribución de tiempos entre excedencias en el límite tiene una masa en  $T = 0$  y la distribución límite de  $F^*(u) T(u)$  es de la forma:

$$(1-\theta) \varepsilon(0) + \theta \text{Exp} (\theta^{-1})$$

donde  $\varepsilon(0)$  es la distribución degenerada en el origen y  $\text{Exp} (\theta^{-1})$  la exponencial de media  $\theta^{-1}$ .

## 1. 3 Estimación del índice extremal

Si los tiempos de excedencia del umbral son  $S(i)$ ,  $i = 1, 2, \dots, N$ , y por tanto los tiempos entre excedencias son  $T(i) = S(i+1) - S(i)$ ,  $i = 1, 2, \dots, N-1$ , el estimador utilizado de  $\theta$  es

$$\theta^* (u) = \frac{2 \{ \sum_{i=1}^{N-1} (T_i - 1) \}^2}{(N-1) \sum_{i=1}^{N-1} (T_i - 1) (T_i - 2)}$$

Cuando el valor obtenido con la fórmula anterior es superior a 1 se toma 1 como estimador. Este estimador tiene sesgo de primer orden nulo (FERRO (2003)).

#### 1. 4 Procedimiento de “declustering”

En un proceso estacionario de excedencias de un umbral con índice extremal distinto de 1 el proceso límite de tiempos de excedencia es Poisson compuesto. Por tanto podemos clasificar los tiempos inter-excedencia en dos tipos:

- 1) tiempos inter-cluster independientes
- 2) conjuntos independientes de tiempos intra-cluster

El segundo tipo lo forman los tiempos dentro de uno de los clusters identificados. La teoría del índice extremal permite identificar los clusters de una forma objetiva. En concreto, suponemos que de los  $N-1$  tiempos inter-excedencias, ordenados de mayor a menor, los primeros  $C-1 = \text{Floor}(\theta N)$  son los de tipo 1, el resto forman el tipo 2 (por tanto la proporción de tiempos inter-excedencias que están dentro de un cluster, y los que están entre clusters distintos está regulada por el índice extremal  $\theta$ ).

Esto quiere decir que si  $T_a$  y  $T_b$  son dos tiempos entre excedencia del tipo 1 sucesivos cronológicamente, entonces todos los tiempos  $T_{a+1}, T_{a+2}, \dots, T_{b-1}$  (que pueden ser vacíos) son los tiempos intra-cluster correspondientes a unos de los  $C$  clusters (el primer cluster empieza en  $T_1$  y el último termina en  $T_{N-1}$ ). Por tanto tenemos  $C$  conjuntos (que pueden ser vacíos) de tiempos intra-cluster, separados por  $C-1$  tiempos inter-cluster.

#### 1. 5 Técnica de “bootstrapping”

Este tipo de técnicas permiten la generación de muestras aleatorias ficticias para incrementar la precisión de los estimadores. En nuestro caso supongamos que queremos hacer inferencias acerca de un funcional  $H$  cualquiera del cluster, que en este trabajo va a ser el índice extremal  $\theta$ . La técnica de “bootstrapping” no paramétrico para obtener límites de confianza del estimador de  $\theta$  contiene los siguientes pasos:

- a) Muestreamos con reemplazo  $C-1$  tiempos inter-cluster del conjunto total de los mismos
- b) Muestreamos con reemplazo  $C$  conjuntos de tiempos intra-cluster del total de los mismos
- c) Formamos un proceso ficticio, intercalando en sucesión cronológica los tiempos inter-cluster entre los clusters
- d) Calculamos para este proceso el funcional en cuestión, en particular  $\theta$

Repitiendo los pasos anteriores un número arbitrario  $m$  de veces tendremos  $m$  estimadores aproximadamente independientes de  $\theta$ , con lo que podremos dar unos intervalos de confianza para  $\theta$ .

## 2. VERIFICACIÓN DEL ÍNDICE EXTREMAL PARA PRECIPITACIÓN DIARIA EN MODELOS CLIMÁTICOS

Se han utilizado las simulaciones de tres modelos (ECHAM4, HadAM3, CGCM2) de la precipitación diaria en el periodo 1961-1990, con la técnica de downscaling estadístico por análogos de la FIC. Al objeto de preservar en lo posible la estacionariedad, requisito necesario para poder aplicar la técnica del índice extremal descrita antes, se han considerado dos periodos bi-mensuales correspondientes a los máximos de otoño y primavera aproximadamente: oct-nov y mar-abr.

Además en cada uno de los anteriores periodos, para reducir la falta de estacionariedad (aunque dos meses es poco), se han computado las ocurrencias de excesos en la precipitación diaria sobre un umbral nominal, el percentil 90, obtenido por ajuste con Gamma en periodos móviles de 15 días. De esta forma para cada observatorio y día del año de uno de los periodos oct-nov o mar-abr se ha formado la serie de 15 días centrada en ese día, y a lo largo de los 30 años, se ha ajustado la gamma biparamétrica a los 15 x 30 valores, y se ha computado el percentil 90 de la gamma ajustada.

Se han comparado las estimaciones del índice extremal obtenidas a partir de las observaciones y de cada uno de los tres modelos en un conjunto de 39 observatorios repartidos por la España peninsular y Baleares (ver tabla 1).

BARCEL.-PRAT	BARCEL.-MONTSENY	SSEB-IGUELDO	SANTAN.-CENTRO	ASTURIAS-GRADO
CORUÑA	VIGO-PEINADOR	PONFERRADA	ORENSE-ALLARIZ	BURGOS-VILLAFRIA
NAVACERRADA	VILLANUBLA	LEON B.A.	SALAM.-MATACÁN	MOLINA DE AR.
BARAJAS	RETIRO	TOLEDO-LAS VENTAS	C.REAL-SOCUÉLLAMOS	BADAJ.-TALAVERA
CORDOBA-AER.	GRANADA B.A.	SEVILLA-MORON	CADIZ-SAN FER.	MALAGA-AER.
MURCIA-ALCANTAR.	ALICANTE-C.JARDÍN	CUENCA	ALBACETE-L.LLANOS	VALENCIA
LOGR.-AGONCILLO	ZARAG.-DAROCA	ZARAG.-AER.	GERON.-LA MOLINA	HUESC.-MONFLORITE
TORTOSA	PALMA-CALVIA	PALMA-FELANICH	IBIZA-AER.	

Tabla 1: LISTA DE OBSERVATORIOS

Para cada modelo y las observaciones, en cada periodo del año, y para cada observatorio, se han realizado 1000 simulaciones con la técnica de “bootstrapping” descrita antes, obteniéndose 1000 estimadores de  $\theta$ .

Para valorar la desviación de cada modelo se ha dividido la diferencia entre la mediana de sus estimaciones de  $\theta$  con la mediana de las estimaciones basadas en las observaciones por el rango intercuartílico de las estimaciones de las observaciones, y multiplicado por 1.35 para corresponder aproximadamente a una desviación estándar. Esto proporciona una estimación del grado de discrepancia entre cada modelo y las observaciones bastante robusta:

$$1.35 \frac{(\text{Mediana } \theta \text{ mod} - \text{Mediana } \theta \text{ obs})}{\text{rango interQ } \theta \text{ obs}}$$

## 2.1 Valores observados del índice extremal

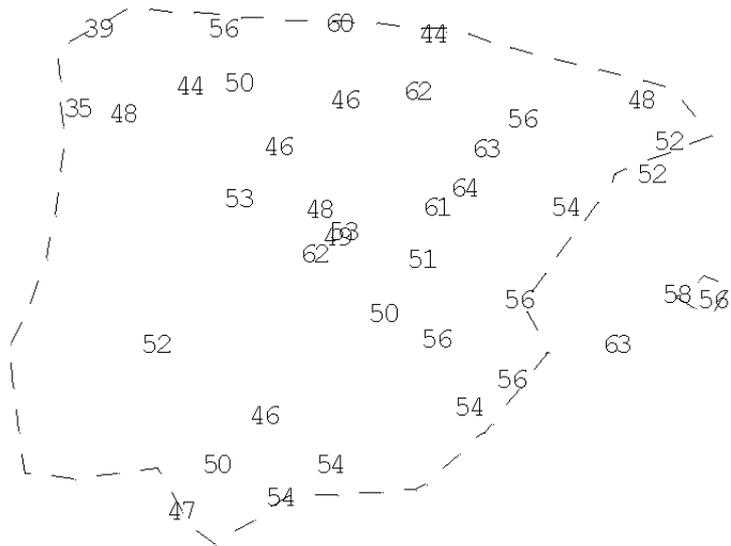


Fig. 1: Valores del índice extremal (en centésimas) obtenidos a partir de las observaciones en oct-nov

En oct-nov  $\theta$  toma los menores valores (más clustering) en el NW, SW y puntos del N. Los mayores valores se dan en el valle del Ebro, puntos del centro y Baleares (fig. 1).

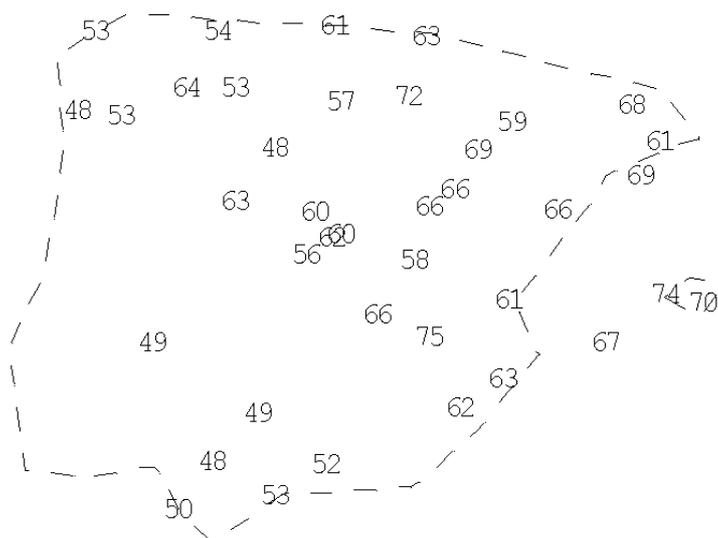


Fig. 2: Valores del índice extremal (en centésimas) obtenidos a partir de las observaciones en mar-abr

En mar-abr también las zonas de influencia atlántica tienen algo menos de índice  $\theta$ , pero los valores altos aparecen más extendidos por la cuenca mediterránea (fig. 2).

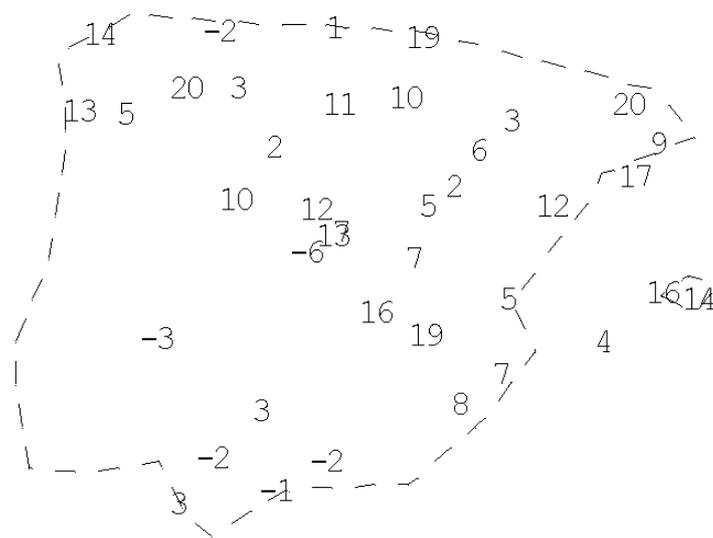


Fig. 3: Diferencias (en centésimas) entre el índice extremal de mar-abr y oct-nov en las observaciones

La figura 3 permite valorar las diferencias de  $\theta$  entre primavera y otoño. Las mayores diferencias de  $\theta$  entre primavera y otoño aparecen en la cuenca mediterránea norte y puntos del interior, las menores en el SW y puntos del N.

## 2. 2 Verificación del índice extremal en octubre-noviembre

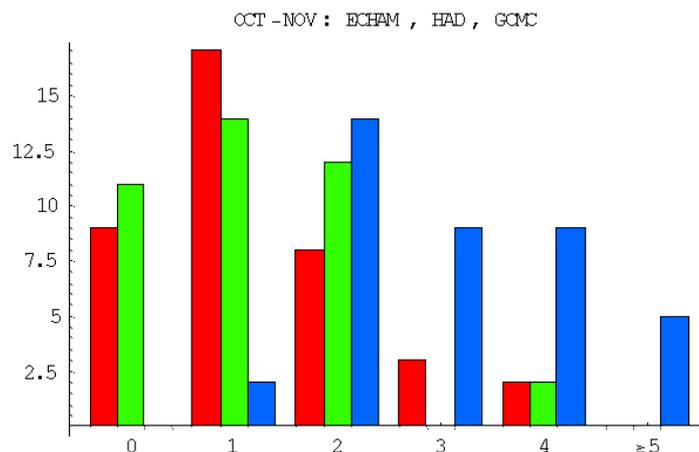


Fig.4: Histograma de los Valores absolutos redondeados de las desviaciones estándar del estimador de  $\theta$  frente a observaciones para los tres modelos en oct-nov, con colores en el orden indicado.

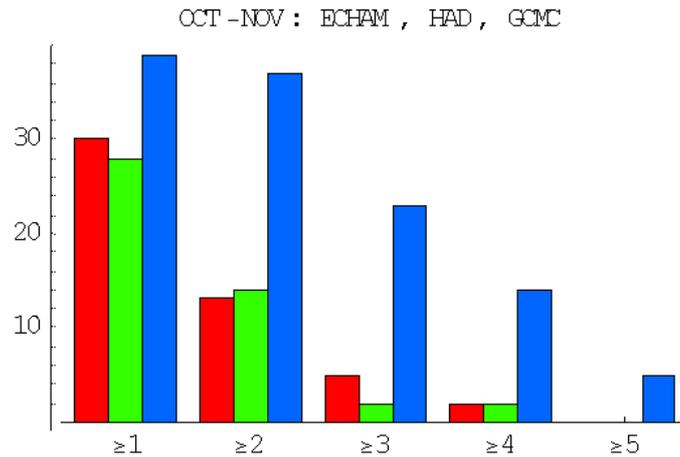


Fig. 5: Como la fig. 4 pero con clases definidas por exceso del umbral indicado.

La desviación absoluta media (para todos los observatorios) de  $\theta$  en cada uno de los tres modelos con relación a las observaciones fue de 1.4 para EC, 1.18 para HA y 3.3 para CG. Las desviaciones medias fueron respectivamente de -0.1, 0.2 y 3.2.

A la vista de lo anterior y de las figuras 4 y 5 deducimos que el modelo CGCM2 da claramente los peores resultados. Los otros dos están bastante igualados, siendo el Hadley mejor para las grandes desviaciones ( $\geq 3$ ), como pone de manifiesto la fig. 5.

Analizamos ahora cómo se distribuye espacialmente el ajuste de los modelos a las observaciones en cuanto al índice extremal. La figura 6 refleja los observatorios en los que un modelo destaca sobre el resto. En este aspecto destaca el Hadley, que domina en la cuenca mediterránea y NW.

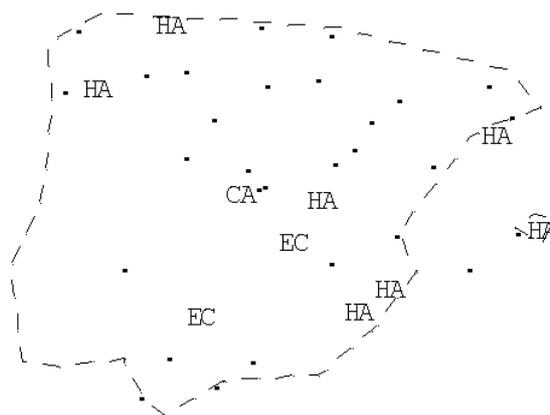


Fig. 6: Observatorios en que un modelo (el indicado) tiene una desviación absoluta de  $\theta$  inferior en más de una unidad al siguiente mejor modelo en oct-nov.

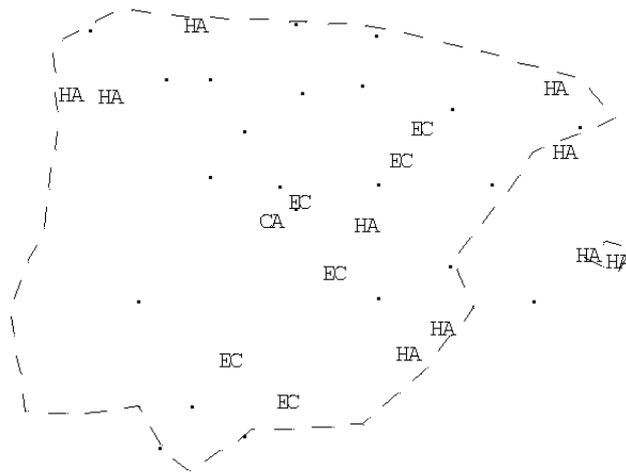


Fig. 7: Observatorios en que un modelo tiene una desv. abs. de  $\theta$  inferior en 0.5 o más a la del siguiente mejor modelo en oct-nov.

En 17 observatorios del total es mejor el EC, en 21 el Had y en 1 el CG. Las figuras 6 y 7 ponen de manifiesto que el Echam es mejor en varios observatorios de Andalucía y del interior peninsular.

Por último la fig. 8 permite centrarse en la comparación entre los dos modelos que mejores resultados dan en oct-nov: el Hadley y el Echam.

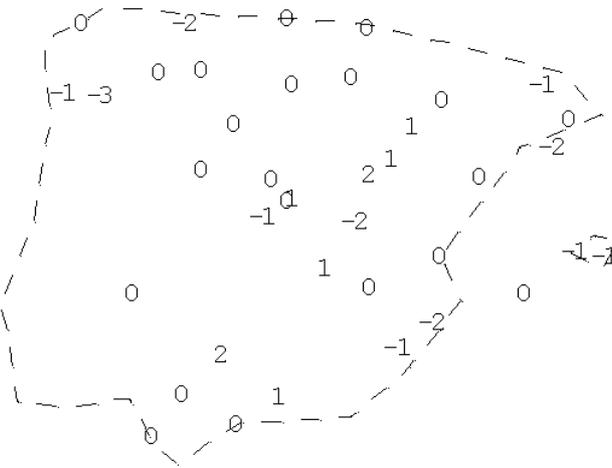


Fig. 8: Diferencias en desviación absoluta de  $\theta$  (respecto a observaciones) entre el modelo Hadley y el Echam, valores negativos favorecen a Hadley.

### 2. 3 Verificación del índice extremal en marzo-abril

La desviación absoluta media (para todos los observatorios) de  $\theta$  en cada uno de los tres modelos fue de 1.8 para EC, 1.4 para HA y 1.5 para CG. Las desviaciones medias fueron respectivamente de 1.15, -0.9 y -0.5.

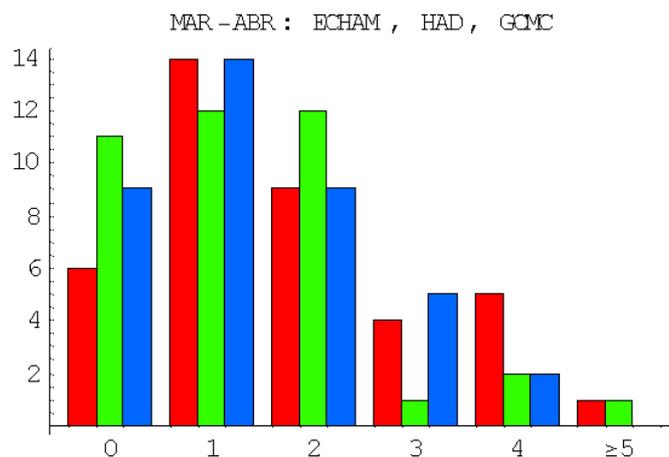


Fig 8: Histograma de los Valores absolutos redondeados de las desviaciones estándar del estimador de  $\theta$  frente a observaciones para los tres modelos en mar-abr, con colores en el orden indicado.

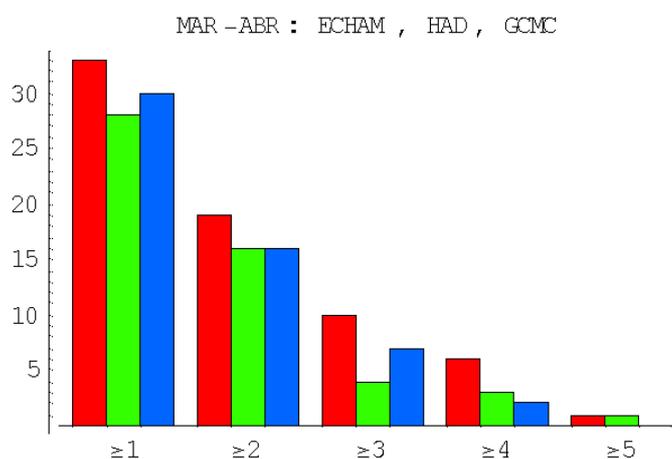


Fig. 9: Como la fig. 8 pero con clases definidas por exceso del umbral indicado.

En primavera los tres modelos están más igualados, el Echam queda un poco peor que los otros dos. En las desviaciones  $\geq 3$  de nuevo el Hadley es el mejor (aunque en  $\geq 4$  vence por poco el canadiense, ver fig. 9).

Analizamos ahora cómo se distribuye espacialmente el ajuste de los modelos a las observaciones en cuanto al índice extremal.

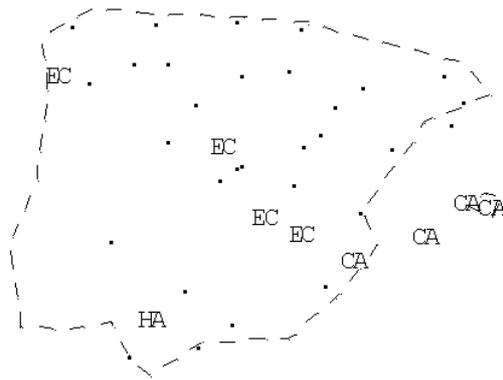


Fig. 10: Observatorios en que un modelo (el indicado) tiene una desviación absoluta de  $\theta$  inferior en más de una unidad al siguiente mejor modelo en mar-abr.

La fig. 10 muestra que el Echam es mejor en la zona centro y el modelo canadiense en la cuenca mediterránea (aquí casi no aparece el Hadley).

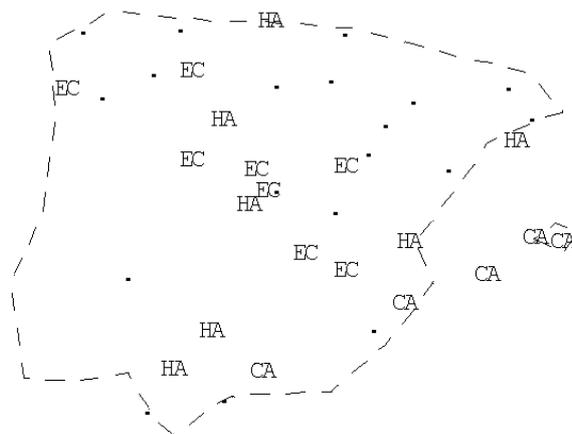


Fig. 11: Observatorios en que un modelo tiene una desv. abs. de  $\theta$  inferior en 0.5 o más a la del siguiente mejor modelo en mar-abr.

La fig. 11 muestra que el Hadley es superior en la vertiente N de la cuenca del Ebro y puntos de Andalucía, el Echam sobre todo en el interior peninsular y el canadiense está más repartido, con más presencia en la cuenca mediterránea S. Del total de observatorios, en 15 es mejor el EC, en 14 el Had y en 10 el CG.

### 3. BIBLIOGRAFÍA Y REFERENCIAS

(1) FERRO, C.A.T. (2003) "Inference for clusters of extreme values", J.R. Statist. Sco. B, 65, pp. 545-556

(2) EFRON, B., R. TIBSHIRANI (1993). *An Introduction to the Bootstrap*. Chapman and Hall.