

# Estudio de la borrasca del 21 al 27 de mayo de 2018 mediante la nueva técnica de "clustering" de AEMET

<https://doi.org/10.31978/639-19-010-0.595>

Alejandro Roa Alonso<sup>1</sup> ([aroad@amet.es](mailto:aroad@amet.es))  
Álvaro Subías Díaz-Blanco<sup>1</sup> ([asubiasd@amet.es](mailto:asubiasd@amet.es))

<sup>1</sup>AEMET / Dirección de Producción e Infraestructuras

## RESUMEN

Durante los años 2017 y 2018 se ha implementado por parte del Área de Técnicas y Aplicaciones de la Dirección de Producción e Infraestructuras de AEMET un nuevo procedimiento de agrupamiento o *clustering* de los miembros del modelo ENS-IFS del ECMWF, basado en el desarrollado en el ECMWF (CORTI y FERRANTI, 2011). A partir del mes de abril de 2018 dicha técnica está operativa dentro del Sistema Nacional de Predicción de AEMET, si bien todavía se mantiene el antiguo método, desarrollado a finales de los 90, como respaldo.

El nuevo método de *clustering* implica no solo una nueva forma de cálculo de los distintos grupos o escenarios, con su correspondiente representación gráfica. También se calculan para cada pasada una serie de índices o *scores* que permiten hacer un seguimiento y validación del propio método, al mismo tiempo que se hace un seguimiento diario del comportamiento de la técnica, comparando de modo subjetivo los análisis y observaciones con las predicciones de los días previos basadas en el método.

**PALABRAS CLAVE:** *clustering*; ens-ifs; ecmwf; predecibilidad; *scores*.

## 1. INTRODUCCIÓN

En este trabajo se estudiará desde el punto de vista de la nueva técnica de *clustering* una situación de baja predecibilidad asociada a la presencia de una baja sobre la península ibérica. A lo largo de la semana del 21 al 27 de mayo de 2018 esta baja experimentó una compleja evolución, de modo que en algunos momentos fue una baja con frentes en superficie y en otros una dana, con importantes efectos en el tiempo en forma de fuertes chubascos y tormentas que afectaron a casi toda España.

Lo que se pretende aquí no es tanto valorar la bondad del método, que aún está por verificar objetiva y subjetivamente, como mostrar un ejemplo de su aplicación práctica por medio de los distintos productos calculados y representados, así como del uso de los índices o *scores* relacionados. Las cuestiones teóricas en relación con la nueva técnica de *clustering* pueden encontrarse en otro trabajo presentado en este mismo Sexto Simposio por los mismos autores, con el título «Adaptación de la nueva técnica de clusters de ECMWF en AEMET», además de en la ponencia del mismo título presentada en las XXXV Jornadas Científicas de la Asociación Meteorológica Española (SUBÍAS y ROA, 2018) y en la Nota Técnica núm. 1 del Área de Innovación de AEMET (RODRÍGUEZ, 2014).

## 2. ANÁLISIS DE LA SITUACIÓN DEL 27 DE MAYO DE 2018

A lo largo de la semana del 21 al 27 de mayo de 2018 se mantuvo una baja sobre la península ibérica que experimentó una compleja evolución, de modo que en algunos momentos fue una baja con frentes en superficie y en otros una dana, con importantes efectos en el tiempo atmosférico en forma de fuertes chubascos y tormentas que afectaron a casi toda España. En la figura 1.a podemos observar la situación a escala sinóptica del 27 de mayo a las 12 UTC en niveles medios, según el modelo HRES-IFS del ECMWF, y en la figura 1.b la Guía Técnica de Diagnóstico de Niveles Altos elaborada por el jefe de turno del Sistema Nacional de Predicción para la misma hora. En las figuras 2.a y 2.b tenemos la situación también a nivel sinóptico en niveles bajos para el mismo día y hora.

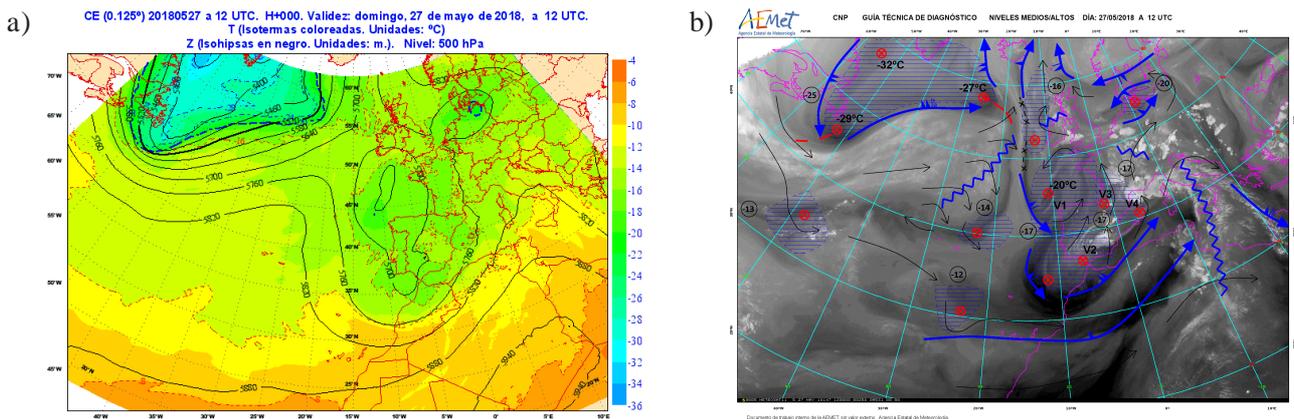


Figura 1. Situación sinóptica en niveles medios y altos del día 27 de mayo a las 12 UTC: a) T y Z en 500 hPa según modelo HRES-IFS y b) Guía Técnica de Diagnóstico de Niveles Altos.

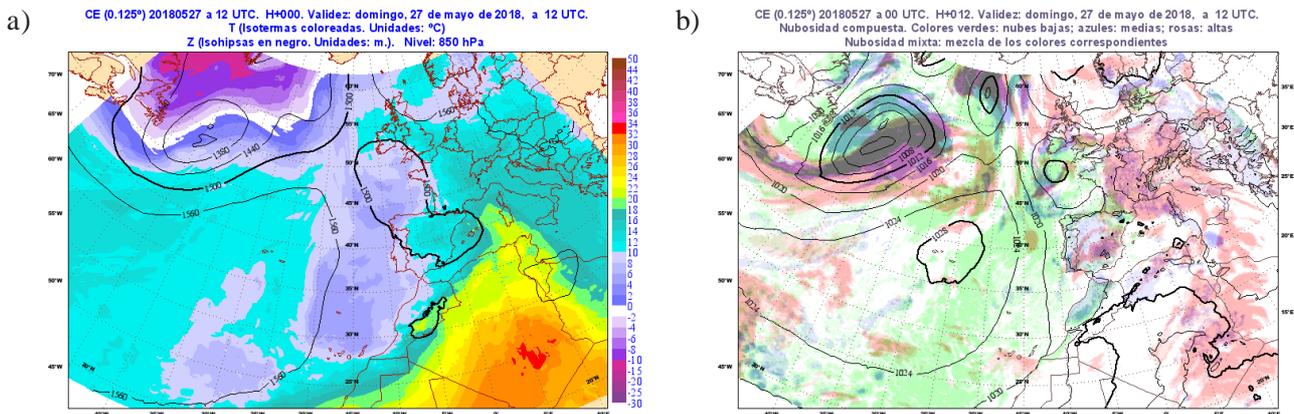


Figura 2. Situación sinóptica en niveles bajos del día 27 de mayo a las 12 UTC según modelo HRES-IFS: a) T y Z en 850 hPa y b) presión al nivel del mar (PSL) y nubosidad por capas.

En la figura 3 tenemos la precipitación recogida en 24 horas según el Banco de Datos Climatológico de AEMET. Se observa una zona de gran cantidad de precipitación en el interior del este peninsular, con valores en el rango de 40-60 mm en 24 horas en gran parte de la provincia de Cuenca y en el oeste de la de Valencia. En otras zonas de la península ibérica se recogieron importantes precipitaciones, pero en cambio hay algunas otras en que fueron prácticamente inexistentes, caso de Galicia, Cádiz, la mayor parte de Córdoba, Sevilla y Málaga, Alicante, Tarragona, Barcelona y Girona, además de las Baleares.

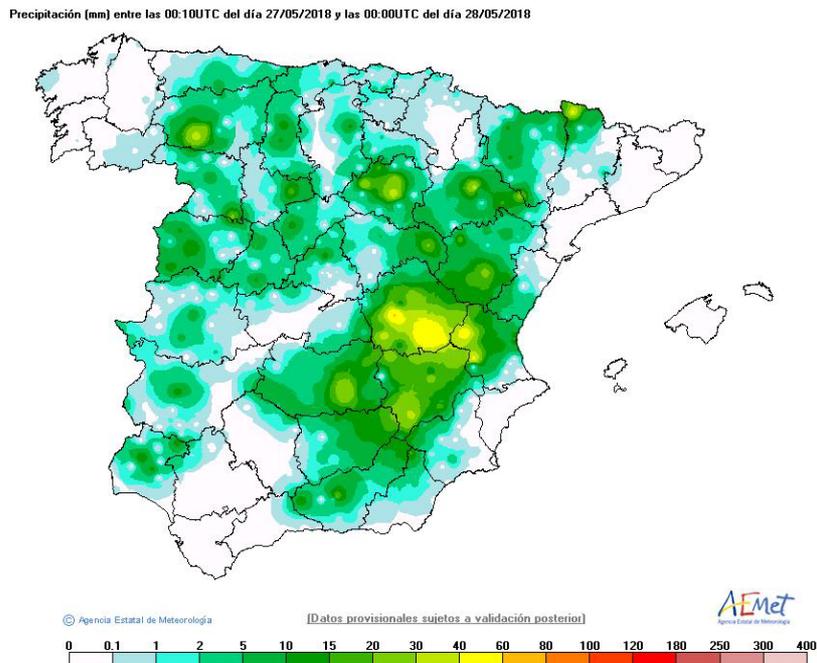


Figura 3. Precipitación recogida durante las 24 horas del día 27 de mayo de 2018 según el Banco de Datos Climatológico de AEMET.

### 3. PREDICCIÓN PARA EL 27 DE MAYO ELABORADA EL DÍA 22 UTILIZANDO LA NUEVA TÉCNICA DE *CLUSTERING*

En las figuras 4a, 4b y 4c se muestran los grupos creados a partir del *ensemble* del modelo ENS-IFS (50 miembros más uno de control), según la nueva técnica de *clustering* a partir de la pasada del 22 de mayo a las 00 UTC para el D+5, es decir, para el 27 de mayo. En vez de representar el campo medio del geopotencial en 500 hPa de los distintos grupos, que es el campo a partir del que se hacen los cálculos, ahora se dibuja el llamado «miembro representante», que es el más cercano al centroide de cada grupo. Se representa además la temperatura en 500 hPa correspondiente al mismo representante. El grupo 1 (figura 4a) cuenta con 23 miembros, siendo el número 27 el representante; el grupo 2 (figura 4b) cuenta con 18 miembros, incluyendo el de control, siendo el representante el 48; finalmente, el grupo 3 (figura 4c), cuenta con 10 miembros, de los que el número 3 es el representante.

En la figura 5 se muestra el campo de precipitación en 24 horas, de 00 a 24 del día 27 de mayo, asociado al representante de cada uno de los *clusters* que ha establecido el modelo. En la figura 6 se muestra la distribución de precipitación para cada uno de esos *clusters*, concretamente la probabilidad de que se supere el umbral

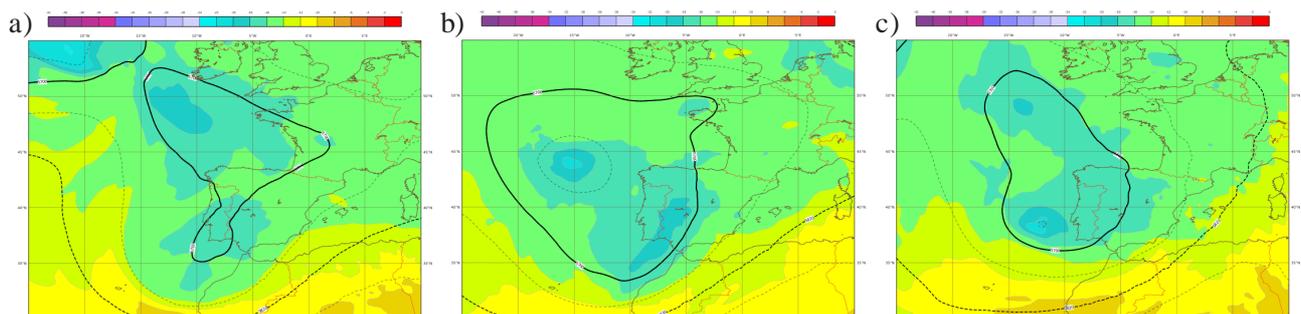


Figura 4. *Clusters* originados con la pasada de las 00 del día 22 del ENS-IFS para el día 27.

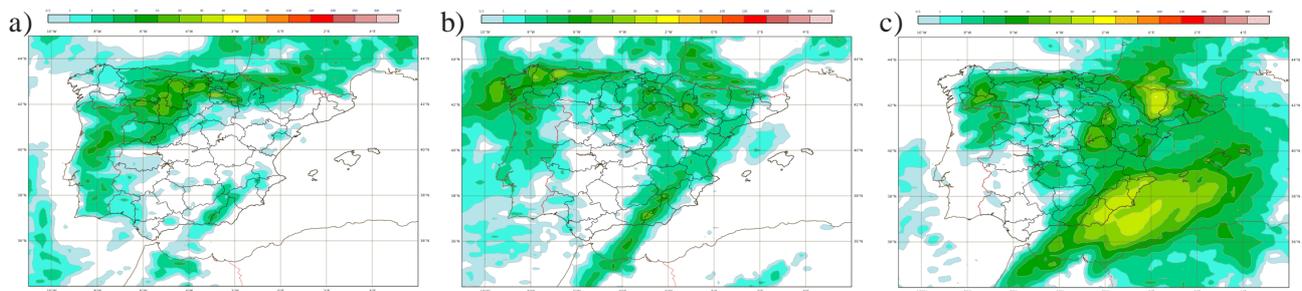


Figura 5. Precipitación en 24 horas asociada a cada uno de los representantes de los *clusters* de la figura 4.

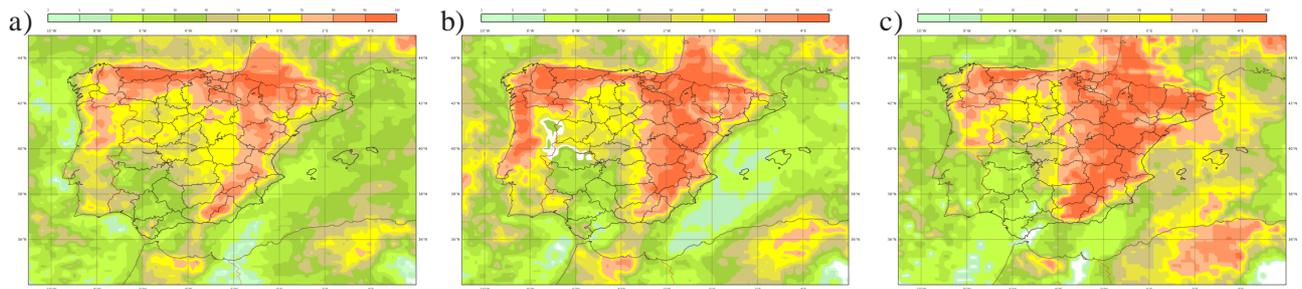


Figura 6. Probabilidad de precipitación en 24 horas mayor de 1 mm asociada a cada uno de los *clusters* de la figura 4.

1 mm en 24 horas, para el mismo periodo. Ambos productos son útiles por sí mismos, al tiempo que complementarios. Comparando con la figura 3, precipitación recogida durante el día 27, podemos observar que el representante del tercero de los *clusters* es el que más se aproxima a la observación; y además que, si bien las probabilidades de precipitación asociadas a los grupos 2 y 3 son similares, no lo son las precipitaciones asociadas a sus respectivos representantes.

#### 4. SIGNIFICADO Y USO DE LOS SCORES

Tal y como se explica en el artículo teórico del que este es complementario (SUBÍAS y ROA, 2018), en la nueva técnica de agrupamiento a partir del ENS-IFS del ECMWF se trabaja en el espacio de componentes principales en vez de en el espacio de puntos de rejilla, reduciendo notablemente el tiempo de cálculo al tratar el agrupamiento en un espacio de dimensión reducida, típicamente un factor 2, en contraposición al espacio de puntos de rejilla, de dimensión igual al número de puntos que se consideren en el dominio de estudio. Para evaluar el funcionamiento de la técnica se han definido los siguientes índices o *scores*:

- **varper:** varianza explicada de cada una de las componentes principales. Se realiza la reordenación habitual, de mayor a menor, de forma que se hace visible el hecho de que muy pocas componentes explican la mayor parte de la varianza explicada. Esto ocurrirá escogiendo variables «suaves» como pueden ser Z500 o T850, por ejemplo.
- **varcen:** media de la varianza entre los centroides.
- **varint:** media entre todos los grupos de la varianza interna de cada grupo.
- **varopt (r):** cociente entre la varianza entre centroides y la varianza interna ( $\text{varcen}/\text{varint}$ ). Fijado un número de *clusters* cuanto mayor sea este valor más nos interesará la agrupación, porque quiere decir que tendrá centroides más separados y dentro de cada grupo los miembros más juntos. Esta función objetivo pretende buscar nubes de puntos bien definidas. Lamentablemente este valor únicamente permite comparar agrupamientos que tengan el mismo número de grupos.

- **sign:** significación estadística de 2 a 6 grupos. Fijada la mejor agrupación  $Q$  para un número de *clusters* prefijado se considera su valor  $r_Q$ . La significación estadística está dada por el porcentaje de perturbaciones a la agrupación  $Q$  que tengan el *score*  $r < r_Q$ . Esto se hace equivalente a que  $Q$  sea máximo local de  $r_Q$  (al menos en un sentido porcentual).

Para establecer el número óptimo de grupos buscamos un máximo de significación estadística que supere un determinado umbral  $t_{min}$ . Sin embargo, se considera una hiperheurística dependiente del alcance con cuatro casos que discriminan entre el corto plazo y los distintos rangos del medio plazo.

- Desde D+0 hasta D+2,  $t_{min} = 55,0$ .
- Desde D+3 hasta D+6,  $t_{min} = 1,0$ .
- Desde D+7 hasta D+9,  $t_{min} = 1,0$  y se elige el máximo absoluto.
- Desde D+10 hasta D+14 se toman 6 grupos siempre.

En los casos i), ii) se elige el primer valor  $ncl$  (número de *clusters*) tal que se cumple que  $sign(n) > sign(n+1) - tol$ , siendo  $tol$  un nivel de tolerancia. Se trata, en esencia, del primer máximo local.

En los casos i), ii), iii), una vez fijado el número de *clusters* óptimo  $n$ , se toma recursivamente  $n-1$  si  $|sign(n-1) - sign(n)| \leq tol$ . Esto quiere decir que si la agrupación óptima con  $n$  grupos tiene similar significación estadística que  $n-1$  grupos, nos quedamos con este último caso. Esto se hace de forma recursiva.

En la figura 7 vemos la representación de los distintos miembros del ENS-IFS de la pasada del día 22 a las 00 UTC para el día 27 representados en un diagrama cuyos ejes se corresponden con las dos componentes principales que explican el 80 % de la varianza en este caso (pueden ser más, pero normalmente no pasan de tres componentes principales). Cada uno de los miembros se representa con su número, correspondiendo el «0» al miembro de control. Los miembros correspondientes a cada grupo se representan del mismo color, el centroide se marca con un aspa, y el miembro representante se recuadra. Se observa que este miembro representante es el más cercano al centroide de cada grupo: para el grupo 1 (el más poblado con 23 miembros y en azul) es el miembro 27 el más cercano al centroide; para el grupo 2 (18 miembros, en verde, que incluye al de control o «0») lo es el 48; y para el grupo 3 (10 miembros, en rojo), el miembro 3 es el más próximo. Al ser dos las componentes principales y ser la gráfica bidimensional, la distancia entre los distintos miembros entre sí y respecto al centroide se conserva, de modo que se observa que realmente el representante es el punto más cercano al centroide. Si fueran tres o más las componentes principales, la proyección en la gráfica bidimensional haría que las distancias aparentes al centroide no se conservaran, y pudiera parecer que el representante no es el miembro más cercano.

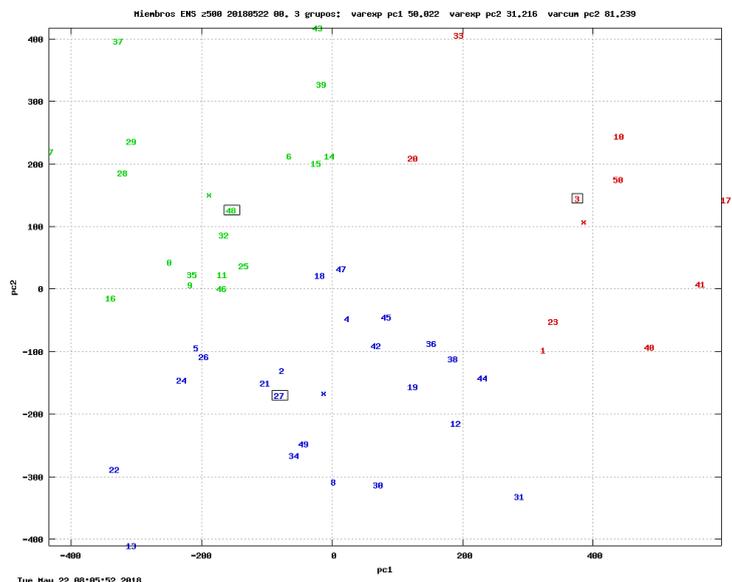


Figura 7. Representación de los distintos miembros del ENS-IFS del día 22 a las 00 UTC para el día 27 (D+5) en un diagrama cuyos ejes son las componentes principales.

El hecho de que el número óptimo de grupos para el alcance D+5 haya sido establecido en tres se deduce de la aplicación de las reglas que hemos visto anteriormente. En la figura 8 se muestra la gráfica correspondiente al índice *sign*, que es la significación estadística. Utilizando las reglas hiperheurísticas el agrupamiento óptimo, para la pasada del día 22 a las 12 UTC y alcance D+5, es de tres *clusters* o grupos, puesto que se alcanza un máximo local, que también es absoluto. [Nota: aunque en el eje de abscisas la numeración empieza en 0, ese punto se corresponde con un grupo o, lo que es lo mismo, todos los miembros del *ensemble*, por lo que el 2 indica tres grupos.]

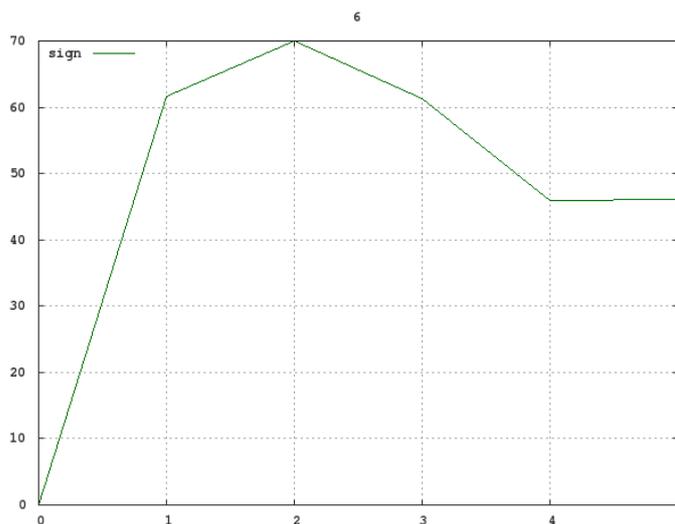


Figura 8. Significación estadística para el agrupamiento de los miembros del ENS-IFS del día 22 a las 00 UTC y alcance D+5 desde un grupo (valor 0) hasta seis (valor 5).

## 5. CONCLUSIONES

La adaptación de la nueva técnica de *clustering* desarrollada en ECMWF era una necesidad para la predicción operativa de medio plazo en AEMET, puesto que el anterior desarrollo tenía cerca de veinte años, y ya estaba en desuso en el propio ECMWF.

Las ventajas de esta nueva técnica de *clustering* pueden resumirse en los siguientes aspectos: 1) se trabaja en el espacio de componentes principales en vez de en el espacio de puntos de rejilla, lo que reduce la dimensionalidad del problema y acelera los cálculos; 2) se ha establecido, mediante unas reglas hiperheurísticas, un método de cálculo del número óptimo de grupos, lo que evita intervenciones subjetivas por parte de los predictores, aunque tampoco estas son por completo descartables; 3) el cálculo y representación de *scores* asociados a este método de cálculo permite hacer un seguimiento y validación.

Queda pendiente para un futuro próximo hacer una verificación objetiva del método, que complemente la subjetiva que ya se está haciendo, así como seguir trabajando en nuevos desarrollos que permitan extraer todo su potencial.

## REFERENCIAS

- CORTI, S., y FERRANTI, L., 2011. New Clustering Products. ECMWF Newsletter No. 127 - Spring 11.
- RODRÍGUEZ, A., 2014. Nuevas técnicas de agrupamiento de predicciones por conjuntos del ECMWF aplicadas a campos de superficie. Nota Técnica núm. 1. Área de Innovación. AEMET.
- SUBÍAS, A. y ROA, A., 2018. Adaptación de la nueva técnica de clusters de ECMWF en AEMET. XXXV Jornadas Científicas de la Asociación Meteorológica Española.