

# Adaptación de la técnica de ECMWF de agrupamiento de EPS en AEMET

<https://doi.org/10.31978/639-19-010-0.145>

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

<sup>1</sup>AEMET / Dirección de Producción e Infraestructuras / Área de Técnicas y Aplicaciones de Predicción

## RESUMEN

Desde marzo de 2018 se ha establecido en AEMET un nuevo procedimiento de agrupamiento de los miembros del *ensemble* de ECMWF. Este procedimiento ha sido realizado inicialmente en ECMWF (CORTI y FERRANTI, 2011) y adaptado en AEMET por los autores, a partir de un estudio previo para campos de superficie (RODRÍGUEZ, 2014).

En la actualidad este es el método usado en AEMET para el agrupamiento de los distintos miembros del *ensemble* de ECMWF efectuado en dos dominios, península ibérica y Baleares y otro centrado en las islas Canarias. Se realizará además un agrupamiento en la zona del estrecho de Drake para la campaña antártica de AEMET en 2019.

**PALABRAS CLAVE:** predicción por conjuntos; *clustering*.

## 1. INTRODUCCIÓN

Para la realización del agrupamiento se trabaja en el espacio de componentes principales de geopotencial en 500 hPa. Se propone una varianza explicada del 80 % para establecer un truncamiento de la dimensión de dicho espacio. Es en ese espacio, típicamente de dimensión de 2 a 3, en el que se realiza un agrupamiento basado en la técnica de *K*-medias. Fijado un número de grupos la mejor elección se realiza para aquel grupo que maximice el cociente  $r_n$  entre la varianza entre centroides y el promedio de la varianza interna de cada grupo. Se obtiene la mejor agrupación para cada número de grupos, desde 1 hasta 6. Este procedimiento permite obtener la elección óptima del número de grupos basada en un criterio de significación estadística. Fijado el número de grupos dicha significación está dada por el porcentaje de grupos calculados a partir de una perturbación gaussiana de la agrupación óptima asociada a dicho número, con coeficiente  $r$  menor que  $r_n$ . La elección del número de grupos es aquella que maximiza dicha significación junto con algunos criterios dependientes del corto, medio o largo plazo (hiperheurística).

La técnica de agrupamiento permite la separación del conjunto de 51 miembros en subgrupos de los cuales se puede extraer el mismo tipo de información que la que se extrae para el total del *ensemble*. En particular una vez obtenidos los grupos se pueden obtener, por ejemplo, probabilidades de superación de umbrales para la precipitación.

Se elige el representante de cada grupo como el miembro más cercano a la media, la cual se desecha por carecer de coherencia meteorológica dado el carácter no lineal de las ecuaciones que gobiernan el estado de la atmósfera.

## 2. MÉTODO

Un agrupamiento de miembros de un *ensemble* que se realice de forma automática requiere la elección de al menos un campo de salida del modelo sobre el que realizar una comparación. Típicamente se elige el geopotencial en 500 hPa (z500) al tratarse de un campo suave (figura 1) a la vez que representativo sinópticamente del estado de la atmósfera.

El anterior método de agrupamiento utilizado en AEMET, desarrollado por ELVIRA y MARTÍN (1999) realiza un agrupamiento en el campo de z500. El algoritmo trata de minimizar la pérdida de varianza al fusionar dos grupos utilizando la distancia de Ward entre ellos que consiste en la distancia L2 entre sus centroides pesada por un factor que tiene en cuenta las poblaciones de ambos, de modo que se favorece la fusión con pocos miembros.

Desde 2014 en ECMWF se utiliza una nueva técnica de agrupamiento desarrollada por L. Ferranti y S. Corti. En ella se comparan campos de z500 en el espacio de componentes principales en vez del espacio de puntos de rejilla. Fijado un alcance en el que se desee realizar un agrupamiento se calculan las funciones empíricas ortogonales (EOF) a partir de la variabilidad de los 50 miembros del EPS junto con el miembro de control. Estas funciones constituyen los modos que diagonalizan la matriz de covarianza del campo analizado. La varianza explicada de los distintos modos decrece muy rápidamente cuando se ordenan de forma decreciente (figura 1), de modo que para explicar una varianza de al menos el 80 % basta escoger simplemente muy pocos modos (típicamente 2 o 3). Los campos en los que se realiza la presente técnica han de ser suaves, es decir, que presenten un decaimiento rápido, o dicho de otra forma, que pocos modos expliquen mucha variabilidad.

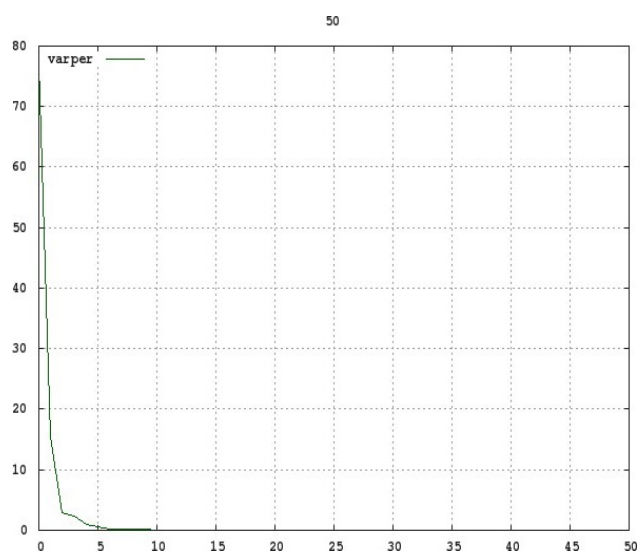


Figura 1. Varianza explicada de cada una de las componentes principales presentadas en orden decreciente. Los dos primeros modos explican más del 80 % de variabilidad.

Una vez que se tienen los modos se proyectan los 51 campos en las funciones empíricas ortogonales y se obtienen las componentes principales de cada uno de ellos. En una situación en la que nos quedamos con 2 modos, basta considerar dos parámetros para identificar cada uno de los campos. Por lo que es posible hacer una representación en un plano en la que tenemos 50+1 puntos. El objetivo consiste en la identificación de grupos o particiones en dicho plano (figura 3).

El objetivo es conseguir particiones en las que los centroides están lo más separados entre sí y dentro de cada grupo los miembros muy próximos al centroide. Para ello se define el valor  $r_Q$  como el cociente entre la varianza entre centroides y la media de las varianzas internas de cada grupo. Lo deseable es que este cociente sea lo mayor posible.

A partir de una partición  $Q$  se realiza un procedimiento de reasignación en el que dos grupos se intercambian un miembro. El procedimiento de reasignación de elementos a grupos más próximos devuelve una nueva partición en la que el procedimiento se estabiliza, es decir, no hace crecer el coeficiente  $r$ . De la partición obtenida consideramos el miembro representante que es el más cercano a la media (*ensemble mean*). El representante escogido de esta manera es un miembro del EPS coherente en todos sus campos, eso no ocurre necesariamente para el *ensemble mean* en el que la media de determinados campos puede no representar un campo físico real (figura 2).

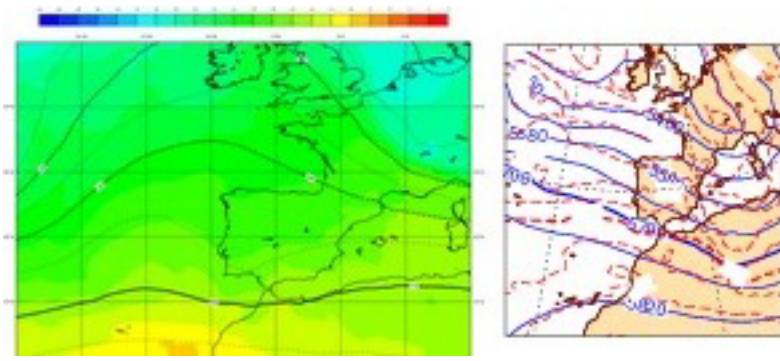


Figura 2. Grupo formado por tres miembros (15, 27, 38) en el que se muestra el representante (izquierda) y la media del *ensemble* (derecha). Pese a representar el mismo grupo pueden apreciarse las diferencias en las isolíneas del campo de geopotencial.

Una vez obtenida una partición  $Q$  se compara la fuerza del agrupamiento con una función de distribución de referencia. Para ello se generan particiones  $Q_m$  con el mismo número de miembros que  $Q$ . Estas muestras se generan a partir de cadenas de Markov cuya media, varianza y autocorrelación a primer orden se obtienen a partir de  $Q$  (CORTI y FERRANTI, 2011). Para cada conjunto inicial de datos  $Q_m$  se realiza un agrupamiento, obteniéndose el coeficiente  $r_{Q_m}$ . La significación estadística estará dada por el porcentaje de muestras sintéticas  $Q_m$  tales que  $r_{Q_m} < r_Q$ .

Inicialmente se parte de un número  $n$  de grupos prefijado, se toman al azar un conjunto de particiones iniciales y se elige la partición  $Q_n$  que maximiza el cociente  $r_{Q_n}$ . Una vez hecho esto se compara la significación de las particiones  $Q_2, \dots, Q_6$  y se elige como número de grupos óptimo aquel que maximiza dicha significación (figura 4). Se tiene en cuenta una hiperheurística si el alcance se sitúa en el corto plazo, de modo que se inhibe el número de grupos, o si se sitúa en el largo plazo, de modo que este se favorece.

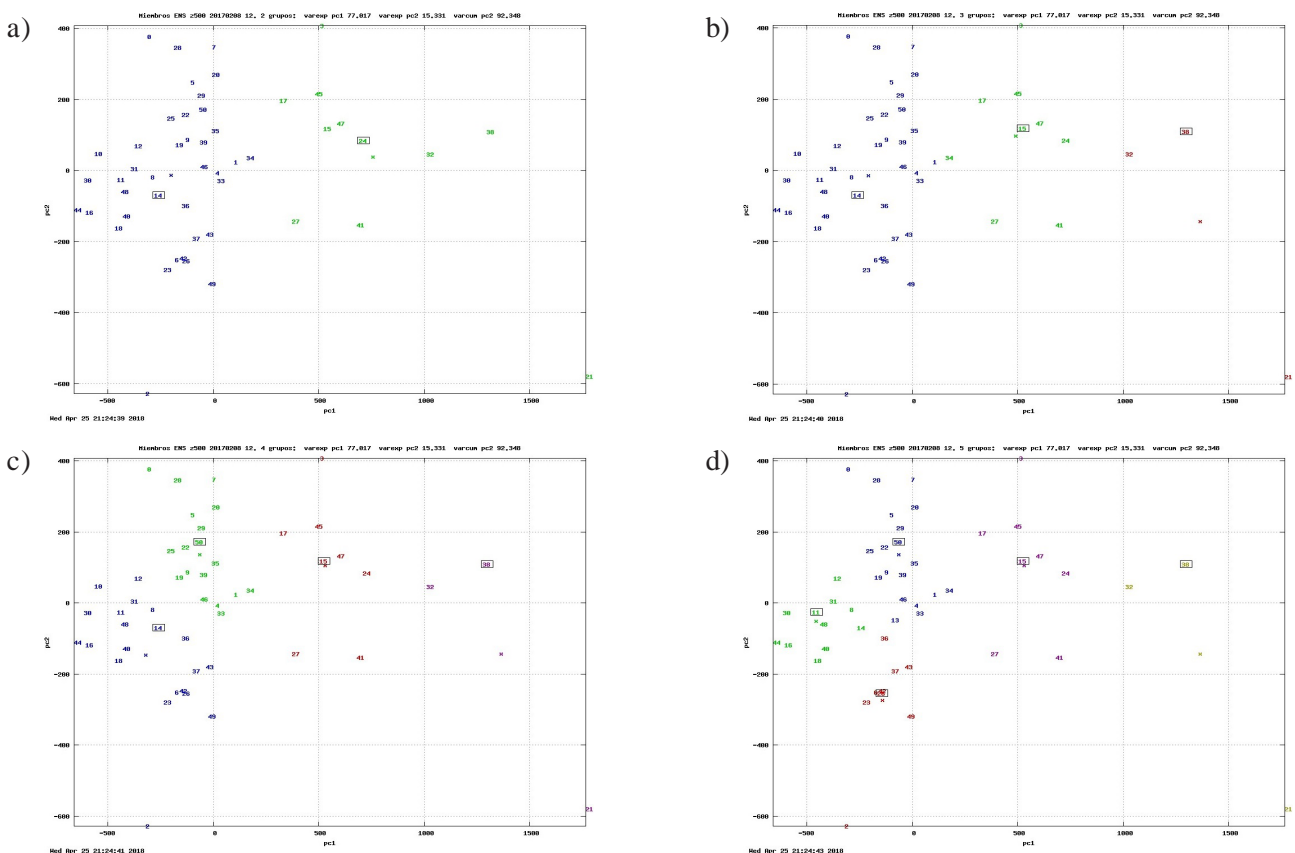
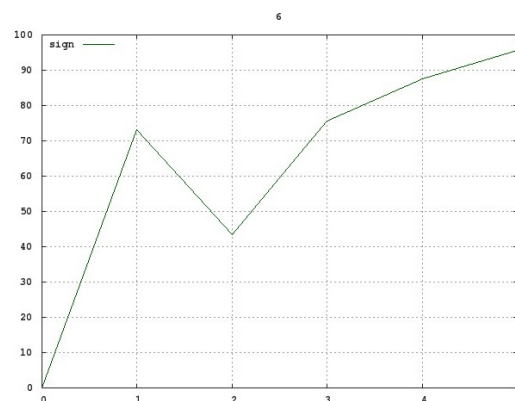


Figura 3. Agrupaciones  $Q_2, Q_3, Q_4, Q_5$  de los 51 miembros en el espacio de componentes principales (figuras a, b, c, d, respectivamente). En las imágenes se muestra la pertenencia de un determinado miembro a un grupo según su color recuadrando el representante. La letra  $x$  representa la media de cada grupo (*ensemble mean*).

Figura 4.  
 Significación estadística según el número de grupos,  
 permite establecer el número de grupos óptimo.  
 En este caso el número óptimo es de dos grupos en el que  
 se sitúa el primer máximo local de significación  
 (hiperheurística de corto plazo).  
 El eje horizontal representa el número de grupos  
 (menos uno).



### 3. RESULTADOS

Una vez identificados los miembros que pertenecen a un determinado grupo se procede a su representación gráfica. Debido a que un grupo constituye un subconjunto de un EPS se puede obtener el mismo tipo de información. Además de mostrar el miembro representante podemos obtener probabilidades como por ejemplo la superación de umbrales de precipitación o la probabilidad de que esté el cielo cubierto.

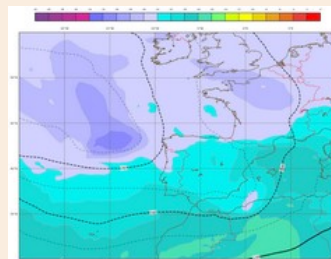
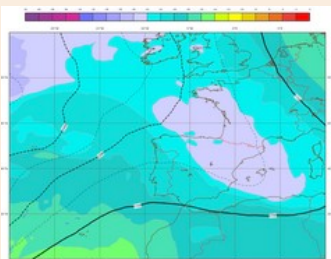
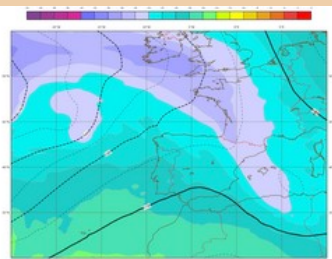
Agrupamiento basado en z500. Número de *clusters*: 2. Varianza explicada: 80,0  
**Validez: mar 14 feb 2017 12 UTC (D+6)**  
 20170208 a 12 UTC HH+144 (rango: 132-156)

**EPS**  
 control  
 miembros: 51

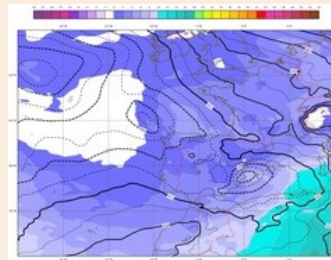
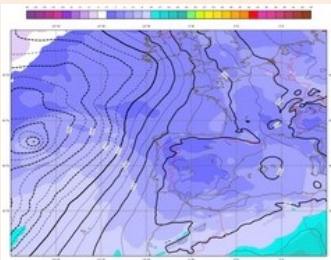
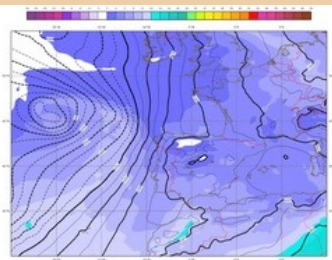
**grupo: 1**  
 representante: 14  
 miembros: 40 **control**

**grupo: 2**  
 representante: 24  
 miembros: 11

**Temperatura (°C) y altura geopotencial (m) en 500 hPa**



**Temperatura en 850 hPa (°C) y presión en superficie (hPa)**



**Probabilidad de precipitación de 0 a 24 UTC superior a 1 mm**

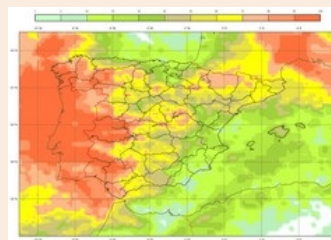
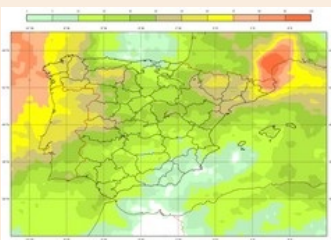
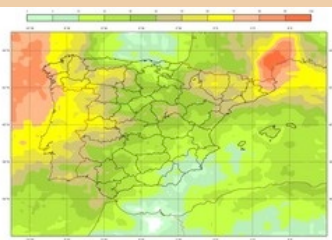


Figura 5. Visualización principal del agrupamiento. En este caso se muestran 2 grupos, pueden observarse las diferencias de probabilidad de precipitación entre ambos.

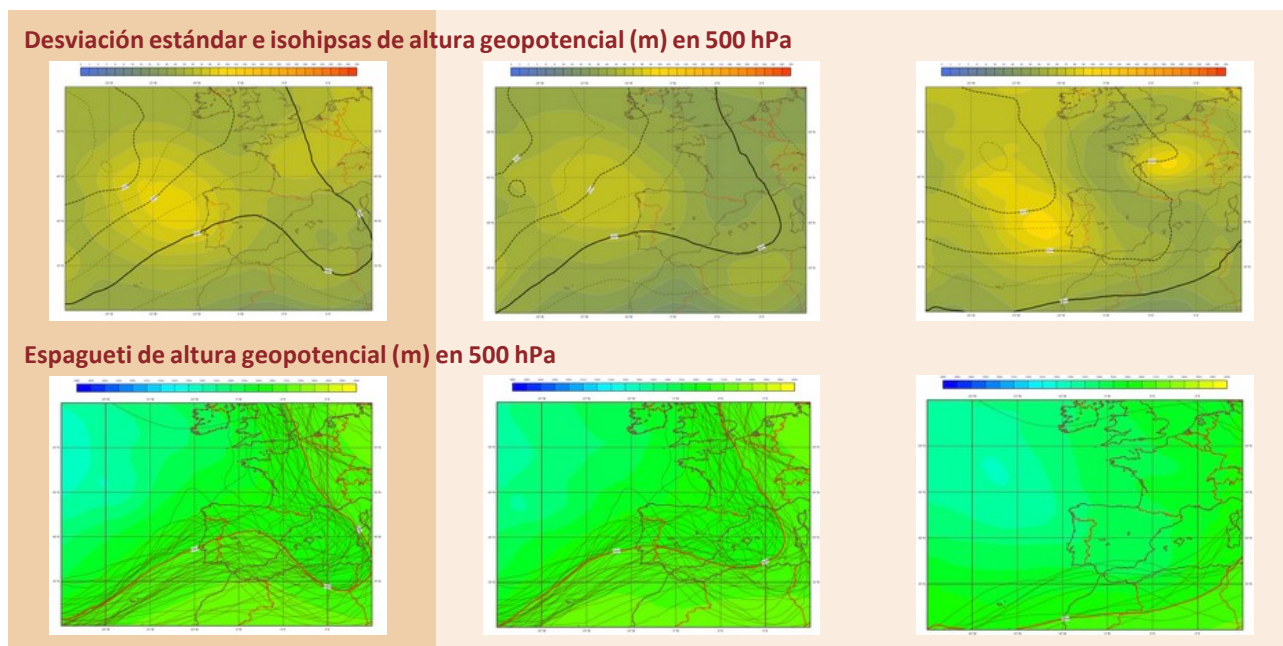


Figura 6. Desviación estándar de  $z_{500}$  y mapa con espaguetis en la isohipsa cuyo miembro de control pasa por Madrid. Los grupos representan la misma situación que en la figura 5.

Esta técnica de agrupamiento se realiza en pasadas de 00 y 12 UTC del modelo ENS de ECMWF para los alcances cuya validez sea las 12 UTC, desde el D hasta el D+14. El cálculo de probabilidades tiene en cuenta las 24 horas del día en el que está centrado el alcance antes mencionado.

Como visualización principal se muestran varias columnas, la primera de ellas siendo el EPS con los 51 miembros y como campo de referencia el miembro de control. Las siguientes columnas están dadas por los distintos grupos, además se indican tanto el número de miembros que lo forman como el representante que se utiliza para la visualización (figura 5).

La técnica de componentes principales separa modos de máxima variabilidad ortogonales entre sí, es decir, con covarianza nula entre ellos. Lo que intenta el presente agrupamiento es desglosar, en el mapa de desviación estándar del EPS conjunto, distintos polos de variabilidad en distintos grupos. Ello puede contemplarse en el mapa de desviación estándar de  $z_{500}$  para el EPS y para cada uno de los grupos (figura 6).

#### 4. CONCLUSIONES

La presente técnica de agrupamiento establece una comparación en el espacio de componentes principales en vez de hacerlo en el espacio de puntos de rejilla, lo que permite que cada campo sea etiquetable con un reducido número de valores. Este conjunto es de dos o tres valores según la situación meteorológica, mientras que para el cálculo en el espacio de puntos de rejilla se necesitan todos los puntos para su completa descripción (en nuestro ejemplo  $176 \times 126$ ). Este hecho permite el desarrollo de algoritmos que solo son posibles debido a la baja dimensionalidad del problema.

A su vez la nueva representación gráfica de campos permite la intercomparación entre los distintos grupos y el EPS conjunto, dado que los grupos constituyen un subconjunto del *ensemble* y por tanto se puede extraer análoga información.

Una posible mejora de esta técnica consistiría en establecer un peso en la zona de mayor interés a la hora de calcular covarianzas. De modo que, por ejemplo, si nos interesa la variabilidad sobre la península ibérica entonces una gran variabilidad en una región como el sur de las islas británicas no debería de ser causante de consideración de nuevos grupos.

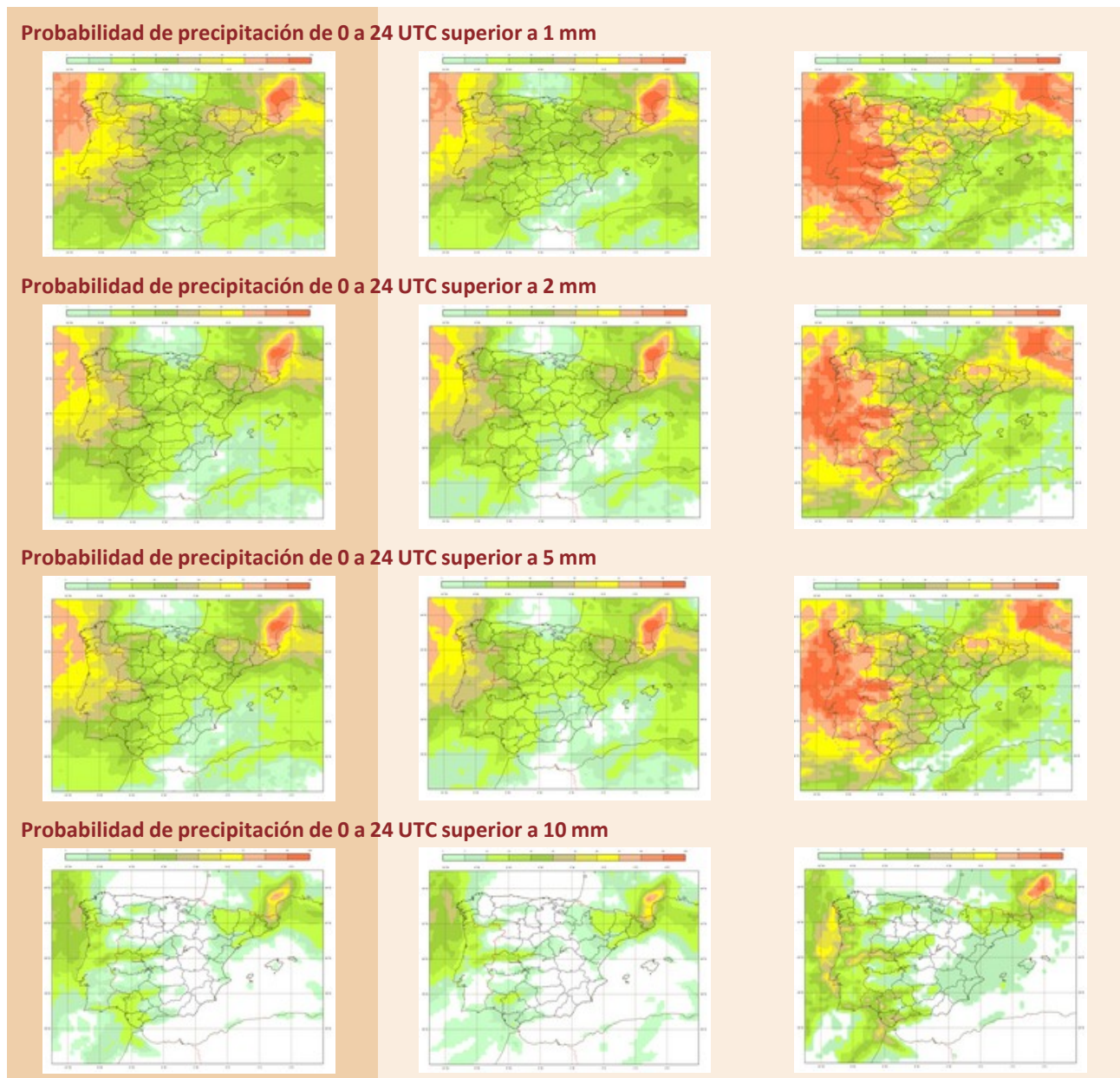


Figura 7. Probabilidad de superación de umbrales de precipitación de 1, 2, 5, 10 mm para los distintos grupos junto con el EPS.

## REFERENCIAS

- ELVIRA, B. y MARTÍN, A., 1999. Sistema de predicción por conjuntos. Nota técnica STAP 33. AEMET.
- ELVIRA, B. y MARTÍN, A., 1999. Verificación del EPS a través de los grupos y supergrupos. Nota técnica STAP 36. AEMET.
- 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 n.º 1. Área de Innovación. AEMET.