

VALIDACIÓN DE LA PRECIPITACIÓN MENSUAL DEL MODELO REGIONAL 'REMO' APLICANDO LA BASE DE DATOS 'MOPREDAS' EN ESPAÑA DURANTE 1989-1999

Miriam ÁLVAREZ -GALLEGO¹, Giovanni LIGUORI¹, William CABOS NARVÁEZ¹ y José Carlos GONZALEZ HIDALGO²

¹*Universidad de Alcalá, Departamento de física, Madrid, España.*

²*Universidad de Zaragoza, Zaragoza, España.*

william.cabos@uah.es

RESUMEN

En el presente estudio se ha llevado a cabo la validación del modelo regional REMO (JACOB et al., 1997), creado en el Instituto Max Planck de Meteorología de Hamburgo, utilizando MOPREDAS (MOnthly PRECipitation DATaset of Spain; HIDALGO et al., 2010) como base de datos de alta resolución de las precipitaciones mensuales de España que cubre el intervalo de 1945- 2005, extrayéndose para este trabajo el periodo 1989-1999.

La validación se ha llevado a cabo teniendo en cuenta tres diferentes resoluciones del modelo REMO: R0088 (0,088°, aproximadamente 10 km), R022 (0,22° aprox. 25km) y R044 (0,44° aprox. 50 km.) aplicándoles datos del reanálisis: ERA Interim, con un dominio centrado en la península Ibérica, para las simulaciones de R0088 y R022, y ERA 40, con un dominio que cubre toda Europa, aplicado únicamente para la simulación de R044.

Se han utilizado diversas técnicas estadísticas tanto univariadas como multivariadas, con énfasis en el procedimiento de clasificación del Análisis Cluster basado en componentes principales, dirigido a identificar y a comparar las pautas espacio-temporales características de las climatologías observada y simulada. Con este método se realizó la regionalización de España en nueve clusters de precipitación similar aplicando el algoritmo k-means.

La utilización del análisis de *clusters* ofrece a la validación una perspectiva novedosa, dado que permite evaluar la bondad del modelo en términos de su reproducción de los regímenes de precipitación observados en diferentes zonas del dominio.

Palabras clave: Precipitación, España, Análisis Cluster, Validación, Modelo Regional.

ABSTRACT

We validate the monthly precipitation over Spain simulated by the regional climate model REMO (JACOB et al, 1997) which has been developed at the Max Planck Meteorology Institute in Hamburg, comparing it to the highest quality dataset available for Spain: MOPREDAS (MOnthly PRECipitation DATABASE of Spain; HIDALGO et al., 2010) extracting for this study the period 1989-1999 from 1945-2005.

This validation has been carried out using three different model resolutions: R0088 (0.088°, approx. 10 km), R022 (0.22°, approx. 25 km) and R044 (0.44°, approx. 50 km). Both R0088 and R022 simulations have been driven by ERA Interim reanalysis with a relatively small domain centered in Spain while R044 has been driven by ERA40 with a large domain covering all Europe.

Diverse statistical tools have been applied, with a special emphasis on Cluster Analysis method based on principal components. Cluster Analysis identifies and compares the observed and the

simulated climatology results. Spain has been regionalized using the k-means algorithm in nine clusters, which divides areas of similar precipitation. Cluster analysis reveals very important spatial and temporal information highlighting the differences and common traits between the observed data and the model results coming up to a more specific and sophisticated validation.

Key words: Precipitation, Spain, Cluster Analysis, Validation, Regional Model.

1. INTRODUCCIÓN

Los recursos hídricos constituyen en España un factor socio-económico y ecológico de carácter crítico, susceptible de agravarse según sugieren diversas proyecciones de clima futuro. A escala regional, estas proyecciones adolecen, sin embargo, de notables incertidumbres cuya elucidación requieren simulaciones de alta resolución espacial tanto de clima presente como bajo distintos escenarios de futuro. Como paso previo a la obtención de simulaciones de esta clase, presentamos aquí la validación de la precipitación mensual del modelo regional REMO respecto a los datos observados MOPREDAS.

El modelo climático regional REMO está basado en el Europamodel, el primer modelo numérico de predicción del tiempo del Servicio Meteorológico Alemán. Posteriores desarrollos del modelo han sido llevados a cabo por el Instituto Max Planck de Meteorología, donde los parámetros físicos del modelo global ECHAM4/T106 fueron implementados en el Europamodel (JACOB y PODZUN, 1997).

En la nueva base de datos MOPREDAS (MONTHly PREcipitation DAtabase of Spain) ha sido construida utilizando 6821 series de datos temporales originales obviando las series menores de 10 años. Este trabajo creó 2670 series completas y homogéneas para el periodo de tiempo 1946-2005, siendo a día de hoy la base de datos más completa y extensa sobre la precipitación mensual de toda España. MOPREDAS ha sido creado con la intención de estudiar el comportamiento de la precipitación en las provincias españolas, así como ayudar en la validación del *downscaling* de los modelos climáticos aplicando un mayor nivel de detalle espacial. En cada estación, el dato de precipitación ha sido interpolado en una malla regular con una resolución de $1/10^\circ$ a lo largo de toda España (HIDALGO et al., 2010).

La validación se ha llevado a cabo teniendo en cuenta tres diferentes resoluciones del modelo REMO: R0088 ($0,088^\circ$, aproximadamente 10 km), R022 ($0,22^\circ$ aprox. 25km) y R044 ($0,44^\circ$ aprox. 50 km.) aplicándoles datos del reanálisis: ERA Interim, con un dominio centrado en la península Ibérica, para las simulaciones de R0088 y R022, y ERA 40, con un dominio que cubre toda Europa, únicamente para la simulación de R044.

Normalmente, los modelos son validados comparando los datos del modelo respecto a los datos observados de regiones arbitrarias y de áreas prefijadas o bien mediante el uso de patrones de dominio global aplicándoles el análisis de EOFs. En este trabajo, el análisis cluster se aplica como la herramienta principal a la hora de validar el modelo, creando una nueva metodología en el campo de las validaciones. El algoritmo K-means es una variante del algoritmo creado por Mac Queen's (ANDERBERG, 1973). En investigaciones geofísicas, los investigadores necesitan clasificar datos en unidades homogéneas y determinar sus características para poder entender la naturaleza del fenómeno y predecir posibles cambios futuros. Por ejemplo, el análisis sinóptico del tiempo o en la regionalización climática, observaciones y variables se dividen en diferentes grupos para revelar importante información que se encuentra sumergida entre los datos, que será imprescindible para

futuras investigaciones (GONG y RICHMAN, 1995).

En su aplicación, podemos dividir las regiones que tengan una precipitación similar en clusters pudiendo estudiar las diferencias entre los datos observados y los del modelo. El análisis cluster revela información espacial y temporal importante, obteniéndose unos resultados en la validación más específicos y sofisticados. Es muy importante la aplicación de este tipo de métodos no supervisados pero entrañan una grandísima dificultad. El objetivo de los clusters es la subdivisión de datos que no contienen una etiqueta específica y que deben ser unificados bajo el mismo cluster homogéneo (FRED y JAIN, 2002).

En primer lugar, para entender mejor los patrones espaciales creados con el análisis cluster que más tarde se explicarán, se ha creído conveniente mostrar la variación de la orografía dependiendo de las tres resoluciones del modelo REMO, ya que la orografía tiene una fuerte influencia sobre la precipitación (Fig. 1(a)). Así mismo, se muestra la media anual de las precipitaciones de los 10 años (Fig. 1(b)) que también tiene una importancia vital en el entendimiento posterior.

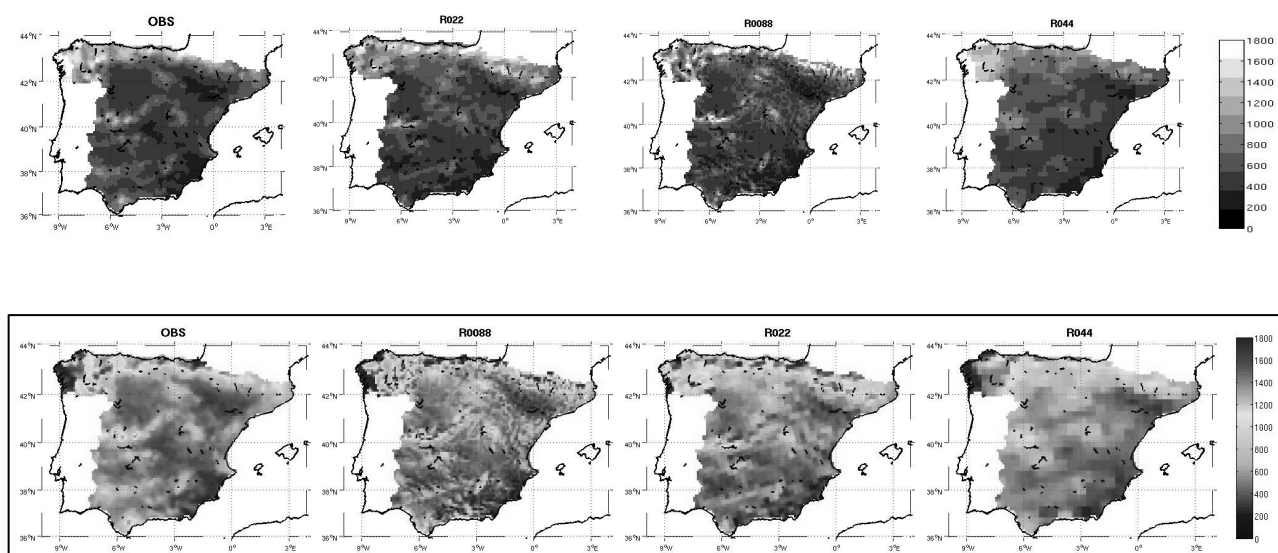


Fig. 1(a) En el panel superior: Comparación entre la orografía real y la estimada por las tres resoluciones. Fig. 1(b) En el panel inferior: Media anual de los 10 años.

2. METODOLOGÍA

Se aplicó el algoritmo k-means como método no jerárquico sobre las diez primeras EOFs (componentes principales que tienen en cuenta la temporalidad de los datos) explicando un 95% de varianza. El análisis de las componentes principales ha sido utilizado como una herramienta de reducción de datos extrayendo la varianza dominante como un prefiltro previo a la aplicación del análisis cluster. Esta técnica asume una relación lineal de los datos (RICHMAN y ADRIANTO, 2010).

Se debe tener en cuenta, que existen cientos de algoritmos a la hora de aplicar el análisis cluster, por ello es difícil asignar el algoritmo adecuado a nuestros datos (FRED y JAIN, 2002). La distancia euclídea es la distancia más difundida en las ciencias atmosféricas (UNAL, 2003) pero la distancia elegida en este estudio ha sido la distancia coseno, la cual también es una medida de similitud entre los datos, pero tiene en cuenta el ángulo que forman entre ellos.

A la hora de decidir determinísticamente cuál es el número de clusters adecuado nos basamos en el

coeficiente silhouette (ROUSSEEUW, 1986).

Este coeficiente asigna el número óptimo de clusters. En la (Fig. 2) las curvas tanto de las tres resoluciones del modelo como de los datos observados asciende con un fuerte pendiente hasta los ocho clusters y a partir de él, el ascenso se suaviza hasta los doce clusters donde la relación del aumento del número de clusters con la pendiente es prácticamente constante. En este trabajo se ha elegido nueve como el número de clusters apropiado ya que su pendiente es más suavizada y supone un número manejable de clusters.

Para asegurar la estabilidad de los nueve centroides de cada uno de los clusters se decidió que la mejor manera sería realizar 100 réplicas, es decir, la distancia del punto al centroide sería recalculada 100 veces, de este modo los centroides se fijarían y no habría variación en el origen de los centroides. De este modo, la robustez en la aplicación del análisis cluster se refuerza, ya que sin realizar esta mejora el algoritmo k-means nos proporcionaría centroides al azar impidiendo realizar cualquier tipo de trabajo con clusters.

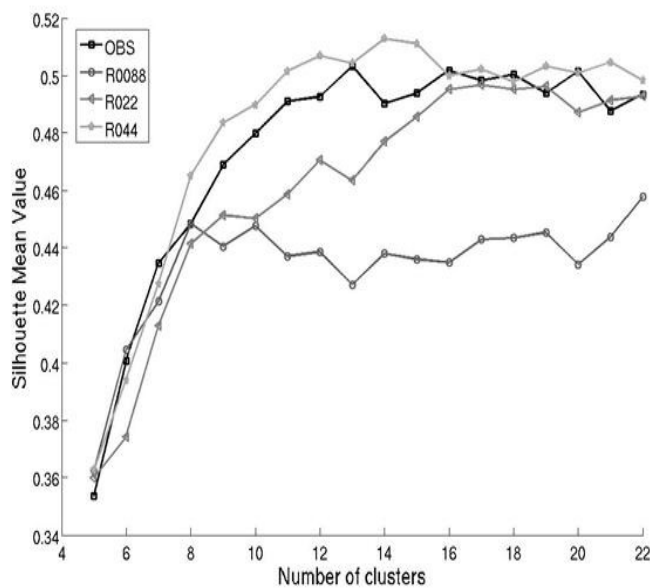


Fig. 2: Coeficiente Silhouette. Método para la obtención del número apropiado de clusters.

3. RESULTADOS

El gráfico OBS contiene la distribución espacial de los clusters de la base de datos MOPREDAS.

Los clusters sólo tienen en cuenta los valores de precipitación anual media de los 10 años.

El gráfico R0088 (ERA Interim) representa la distribución espacial del modelo a la resolución más alta, aproximadamente 10 kilómetros, al aplicar los nueve clusters. Su mayor grado de detalle se muestra en la formación de pequeños subclusters dentro de otros clusters, dando información sobre la orografía de España, ya que en zonas de alta montaña la precipitación, en términos generales, es mayor. En R022 (ERA Interim) que constituye la resolución del modelo a 25 kilómetros, proporciona una distribución de los patrones espaciales más homogeneizado que en el caso anterior. Se diferencia principalmente de R0088, en los clusters 7 y 3 provocando una conformación espacial distinta en el cluster 4. Además, se incluye un subcluster en el cluster 1 perteneciente del cluster 8.

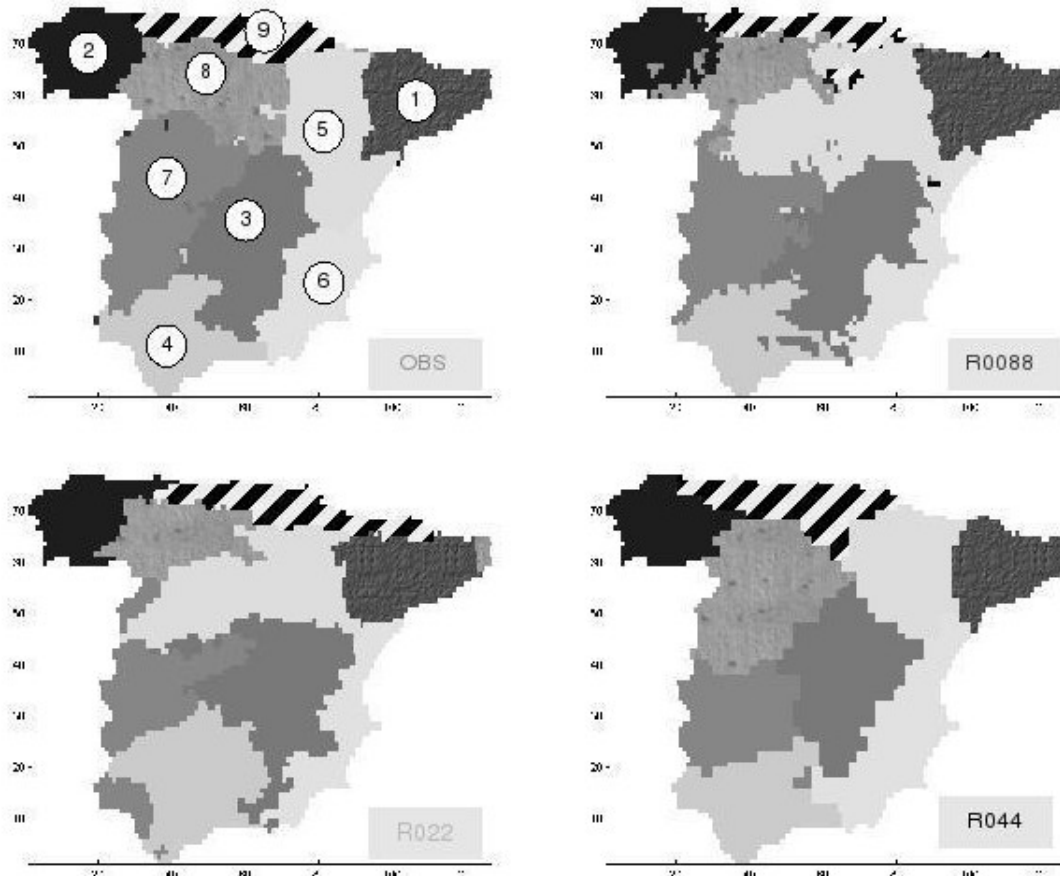


Fig. 3: Distribución espacial de los clusters tanto en los observados (arriba a la izquierda) como en las tres resoluciones del modelo.

En R044 las condiciones de contorno son las del reanálisis ERA40. Mediante esta simulación de baja resolución obtenemos los clusters muy unificados. Las mayores diferencias con los observados se encuentran entre los clusters 7 y 8.

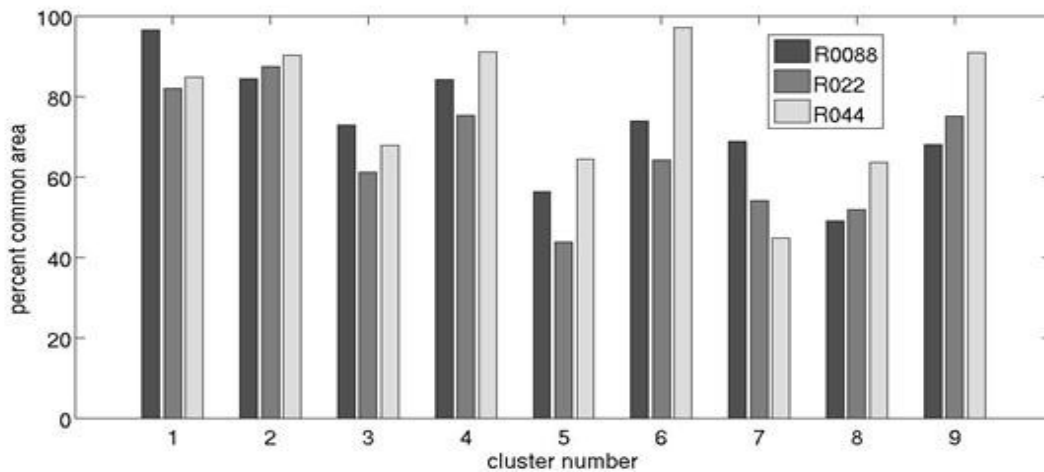


Fig. 4: Porcentaje de área común entre los datos observados y las tres resoluciones.

En el gráfico (Fig.4) se hace una comparativa del porcentaje de área común entre los datos observados y las áreas de las tres resoluciones del modelo. Como se observa en la (Fig.4) y en la (Fig.5) los clusters que peor simulan: cluster 5, cluster 7 y cluster 8 son los que tienen porcentaje de área menor. En el resto de los clusters el porcentaje de área es superior al 70%. En futuras líneas de investigación, nuestros estudios se centrarán en métodos que homogeneización de clusters aplicando análisis de correlación, de tendencias, de regresiones o métodos como H-estadística, el cual se basa en la técnica L-momento, es utilizado como un método de homogeneización de la media anual de precipitación dentro de cada cluster (DINPASHO, 2004) y (SRINIVAS, 2008).

A continuación se muestran las climatologías de cada cluster.

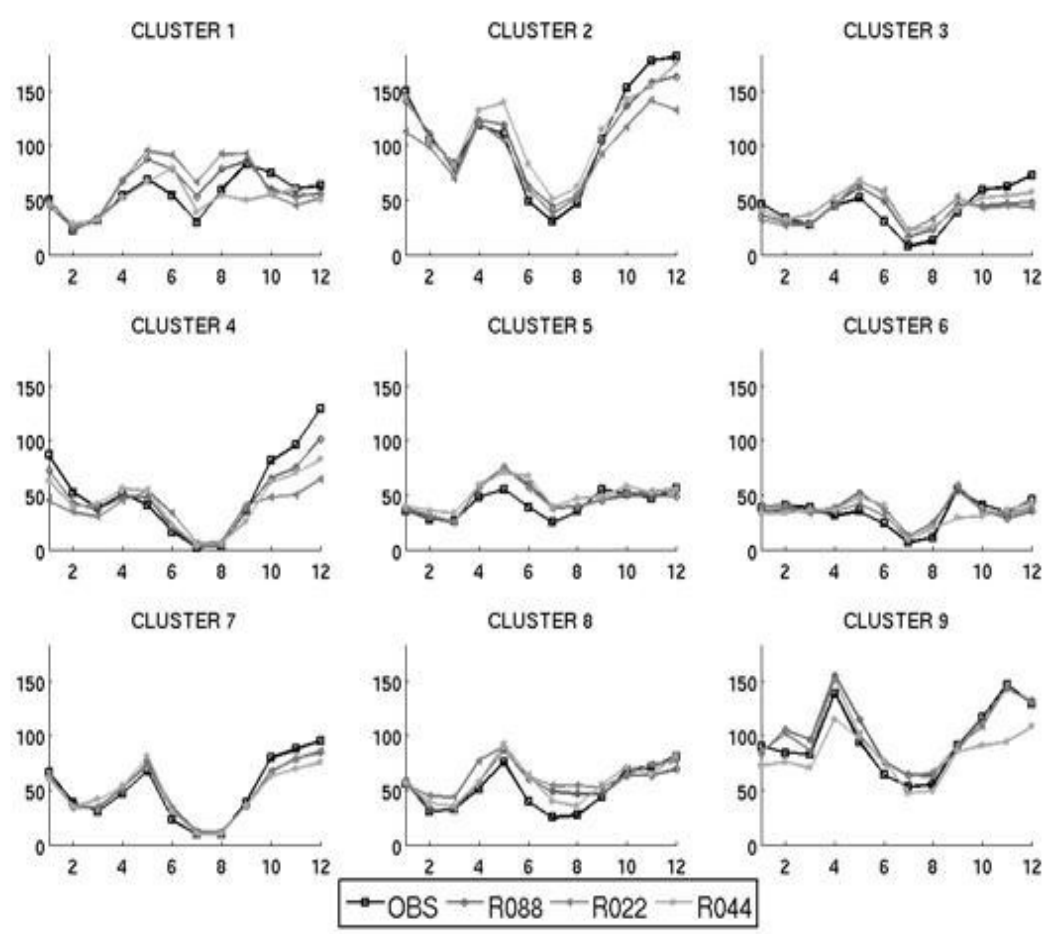


Fig. 5: Climatología de los 9 clusters comparando las 3 resoluciones del modelo y los datos observados.

De manera global, todas las resoluciones reproducen satisfactoriamente el mínimo de verano y las subidas en la época de otoño-invierno, así como las tendencias generales del ciclo. Detallando el análisis, la confluencia de las tendencias en el cluster 1 es menor que en el resto de los clusters, pero aún así los resultados en las tendencias son buenas. En los clusters 2, 3, 4 y 7 se obtienen las mayores discrepancias en las tendencias a lo largo de los meses de Septiembre, Octubre, Noviembre y Diciembre. En el resto de los clusters las climatologías se ajustan muy bien a

las tendencias de los observados.

Se debe tener en cuenta que en la mayoría de las climatologías las tendencias de las simulaciones tienen valores de precipitación media mayor que la de los observados.

La “mejor” aplicación del cluster está muy relacionada con la minimización de las 3 formas de bias: metodológicas, letent e informativas. Tanto las fuentes como las soluciones a este problema están muy discutidas (FOVELL y FOVELL, 1993)

4. DISCUSIÓN

Mediante la aplicación del Análisis Cluster se ha podido validar el modelo regional REMO extrayendo información más concisa y más detallada pudiendo entender mejor los pros y contras del modelo. El detalle obtenido en la regionalización de España permite un nivel de profundidad mayor que en otros estudios pudiendo hacer una comparativa más exhausta. Mediante esta técnica la regionalización de España se hace teniendo en cuenta los patrones naturales de precipitación y pudiendo entender mejor cómo funciona el modelo y en dónde los ajustes son mejores y dónde son mejorables. La climatología de los clusters no sólo depende de la evolución temporal sino también de la ubicación espacial dentro de España.

La simulación R044 ofrece los mejores resultados. Esto puede ser debido a que las simulaciones que tienen condiciones de borde a gran escala (ERA40) son mejor representadas que las simulaciones que sólo engloban a España (ERA Interim) ya que son simulaciones más completas que valoran mayor cantidad de variables.

La orografía se aprecia mejor mediante R0088, la alta resolución nos da una información más precisa de las altas cumbres apreciándose como subclusters dentro de otros clusters. En las resoluciones más bajas estos subclusters no se aprecian.

5. AGRADECIMIENTOS

Este trabajo está subvencionado por el Ministerio de Educacion y Ciencia del Gobierno de España. Giovanni Ligouri agradece el apoyo dado por la Universidad de Alcalá.

6. REFERENCIAS

- ÁLVAREZ -GALLEGO, M ; LIGUORI, G.; CABOS NARVÁEZ, W.; GONZÁLEZ HIDALGO, J.C (2009). “Validación del modelo regional REMO en la vertiente Mediterránea española”. Proyecto de fin de carrera. Universidad de Alcalá de Henares.
- ANDERBERG, M.R (1973). “Cluster analysis for applications”. *Academic Press*, pp 359.
- DINPASHO, Y.; FAKHERI-FARD, A.; MOGHADDAM, M.; JAHANBAKHS, S.; MIRNIA, M; (2004). “Selection of variables for the purpose of regionalization of Iran's precipitation climate using multivariate methods”. 109-123.
- FOVELL, R.; FOVELL, M. (1993). “Climate Zones of the Conterminous United States Defined Using Cluster Analysis”. *Journal of Climate*, pp 2103-2135.
- FRED, A.; K.JAIN, A.; (2002). "Data Clustering Using Evidence Accumulation," *Pattern Recognition, International Conference on*, vol. 4, pp. 40276, 16th International Conference on Pattern Recognition (ICPR'02).
- GONG, X.; RICHMAN, M.B (1995). “On the application of cluster analysis to growing season precipitation data in North America east of the Rockies” *Journal of Climate* 8 pp. 897–931.

- HIDALGO et al. (2010). "A new tool for monthly precipitation analysis in Spain: MOPREDAS database (monthly precipitation trends December 1945-November 2005)". *International Journal of Climatology*. Published Online: 10 Marzo 2010.
- JACOB, D.; PODZUN,R. (1997). "Sensitivity studies with the regional climate model REMO". *Meteorology Atmospheric Physics* 63: 119-129.
- RICHMAN, M.; ADRIANTO, I. (2010). "Classification and regionalization through kernel principal component analysis Physics and Chemistry of the Earth" *Physics and Chemistry of the Earth*.
- ROUSSEEUW, P(1986). "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis" *Journal of Computational and Applied Mathematics*. Volumen 20, pp 53-56.
- SRINIVAS, V.; TRIPHATI, S.; RAMCHANDRA, A.; GOVINDARAJU, R. (2008)."Regional flood frequency analysis by combining self-organizing feature map and fuzzy clustering". *Journal of Hydrology*, Volume 348, Issues 1-2, 1 January 2008, Pages 148-166.
- UNAL, Y.; KINDAP, T.; KARACA, M; (2003). "Redefining the climate zones of turkey using cluster anlysis". *International Journal of Climatology*. Volumen 23, Issue 9, pp 1045-1055.