

MODELO ESTADÍSTICO DE RECONSTRUCCIÓN DE LA PRECIPITACIÓN REGIONAL EN EL MEDITERRÁNEO OCCIDENTAL

M^a Luisa MARTÍN*; M^a Yolanda LUNA**; Ana MORATA*** y Francisco VALERO***

* *Departamento de Matemática Aplicada, Universidad de Valladolid*

** *Servicios de Desarrollos Climatológicos, Instituto Nacional de Meteorología*

*** *Departamento de Astrofísica y CC. de la Atmósfera, Universidad Complutense de Madrid*

RESUMEN

El objetivo fundamental de este trabajo es demostrar la influencia de la circulación atmosférica en el nivel de 300 hPa sobre la precipitación en la cuenca mediterránea occidental. Para ello se formula un modelo de reconstrucción de la precipitación regional a partir de los resultados obtenidos del análisis de la variabilidad acoplada entre ésta y la circulación atmosférica, en el que posteriormente se incluye la influencia en el comportamiento de la precipitación de factores orográficos locales a cada observatorio.

Palabras clave: Precipitación, cuenca mediterránea occidental, modelo de reconstrucción, factores orográficos.

ABSTRACT

The main goal of this work is to prove the influence of 300 hPa level atmospheric circulation over regional precipitation in the western Mediterranean Basin. A reconstruction model of regional precipitation is formulated in terms of the results obtained from a coupled-variability analysis. The model is improved including local orographic factors.

Key words: *Precipitation, western Mediterranean basin, reconstruction model, orographic factors.*

1. INTRODUCCIÓN

Las principales características del clima en la cuenca del Mediterráneo occidental y, por tanto su variabilidad, se encuentran fuertemente vinculadas con la circulación atmosférica a gran escala y, por tanto, a las corrientes en chorro de niveles altos de la atmósfera, concretamente en la superficie bórica de 300 hPa (MARTÍN *et al.*, 2004; VALERO *et al.*, 2004). Asimismo, el clima de esta región también está muy influido por sus características orográficas, los fuertes contrastes térmicos entre el mar y la tierra y, finalmente, por efectos más locales (MILLAN *et al.*, 1995). El análisis de las interacciones entre la circulación atmosférica a gran escala y la precipitación regional resulta determinante para profundizar en el conocimiento de la variabilidad conjunta. Para ello se utilizan métodos estadísticos multivariantes, tales como el análisis de descomposición en valores singulares (DVS), capaces de extraer información común entre diferentes campos para identificar patrones que ocurran de forma simultánea. El objetivo es formular un modelo de reconstrucción de la precipitación regional a partir de los resultados de varios análisis DVS. En este trabajo, se pretende demostrar la influencia de la circulación atmosférica en altura sobre la precipitación en superficie en el área de la cuenca mediterránea occidental, más que desarrollar un modelo predictivo *ad hoc*. No obstante, la constatación de la relación entre las variables de altura y la predicción pondrían la base para el desarrollo de modelos predictivos de la precipitación.

El estudio consta, de dos fases, la primera de ellas consiste en la calibración de un modelo de reconstrucción de la variable regional a partir de los modos de variabilidad obtenidos entre ésta y los campos de las variables de circulación en altura. La segunda fase ha consistido en el análisis de la posible influencia de factores locales, fundamentalmente geográficos, en la variabilidad residual que no ha sido reconstruida con las variables de circulación a gran escala.

2. DATOS

La variable regional consta de 26 series de precipitación mensual acumulada (PCP) correspondientes a observatorios del sur de Francia, litoral mediterráneo peninsular y norte de África (Fig. 1), extraídas de la base de datos observacional Southwestern Europe Dataset (SED) (GONZÁLEZ-ROUCO *et al.*, 2001). Esta base de datos contiene 92 series de precipitación mensual acumulada de observatorios de toda la Península Ibérica, sur de Francia y norte de África. Las series fueron sometidas a un exhaustivo control de calidad consistente en corrección de datos anómalos, homogeneización e interpolación de datos inexistentes que se describe en detalle en dicho trabajo. Para este estudio se parte, así, de 92 series con un período de 42 años comprendidos entre 1948 y 1989. Las 26 series finalmente empleadas presentan un comportamiento similar en el área de la cuenca mediterránea occidental y han sido extraídas de la base SED mediante un procedimiento objetivo no lineal que puede verse en detalle en MORATA (2004). Los valores de las variables son valores mensuales para la estación de primavera entendida como: marzo-abril-mayo.

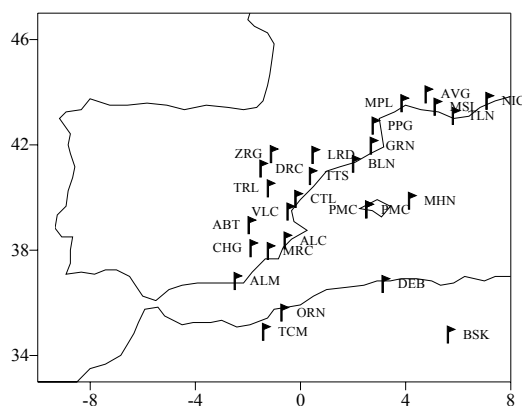


Fig. 1. Distribución espacial de las estaciones SED en el Mediterráneo occidental

Las variables de circulación a gran escala empleadas en el desarrollo de este trabajo han sido extraídas de los reanálisis del *National Centre for Atmospheric Research, NCEP / NCAR* (KALNAY *et al.*, 1996). Las variables escogidas son valores medios mensuales de altura geopotencial en el nivel de 300 hPa (Z300) y valores medios mensuales de las componentes zonal (U300) y meridional (V300) del viento en el mismo nivel. Los datos están organizados en una malla de 2,5° x 2,5° (longitud x latitud) que comprende un dominio espacial que abarca el Atlántico Norte, el Mar Mediterráneo y Europa, desde 20°N a 85°N de latitud y desde 100°W a 40°E de longitud, y el periodo temporal entre 1948 y 1989.

3. FORMULACIÓN DEL MODELO DE RECONSTRUCCIÓN

El método de Descomposición en Valores Singulares es una herramienta para el estudio de relaciones lineales entre dos campos. Se fundamenta en encontrar combinaciones lineales de dos conjuntos de variables, tales que la combinación lineal dé cuenta de la mayor covarianza temporal posible bajo condiciones de ortogonalidad en los coeficientes. El método puede entenderse como la extensión de la diagonalización de matrices simétricas cuadradas a matrices rectangulares (BRETHERTON *et al.*, 1992). Se trata así de la diagonalización de la matriz de covarianza entre las observaciones de dos campos físicos definidos sobre dos dominios espaciales que permite identificar modos acoplados de variabilidad entre los campos estudiados. El análisis DVS constituye una herramienta potente en el análisis de regionalización y predicción estadística a partir de la información extraída de los modos de variabilidad acoplada obtenidos (WIDMAN y SCHÄR, 1997).

La matriz diagonal de elementos de la descomposición hace posible la reconstrucción de la serie de coeficientes de expansión derechos a partir de los izquierdos y viceversa (WALLACE *et al.*, 1992). Los campos originales izquierdo y derecho, $S(s,t)$ y $Z(z,t)$, mediante la descomposición singular se pueden expresar como:

$$\begin{aligned} s(t) &= \sum_{k=1}^{N_s} a_k(t) \mathbf{q}_k \\ z(t) &= \sum_{k=1}^{N_z} b_k(t) \mathbf{p}_k \end{aligned} \quad (3.1)$$

donde \mathbf{q}_k y \mathbf{p}_k son los vectores singulares izquierdo y derecho, respectivamente. Las series temporales $a_k(t)$ y $b_k(t)$ son las series de coeficientes de expansión correspondientes, respectivamente, calculados mediante la proyección de sus vectores singulares sobre los datos originales.

El coeficiente de expansión derecho (izquierdo), contiene información de una parte de la covarianza conjunta, linealmente acoplada con el coeficiente de expansión izquierdo (derecho), de la forma:

$$b_k(t) = \frac{\sigma_k}{\langle a_k^2(t) \rangle} a_k(t) \quad (3.2)$$

donde σ_k es el autovalor.

Se define la correlación entre las series temporales de coeficientes de expansión izquierdos y derechos, $r[a(t), b(t)]$, mediante la aproximación

$$r_k [a_k(t), b_k(t)] = \frac{\sigma_k}{\left(\langle a_k^2(t) \rangle \cdot \langle b_k^2(t) \rangle \right)^{\frac{1}{2}}} \quad (3.3)$$

El coeficiente de expansión izquierdo (derecho) puede entonces expresarse como:

$$\hat{b}_k(t) = r_k a_k(t) \frac{\langle b_k^2(t) \rangle^{\frac{1}{2}}}{\langle a_k^2(t) \rangle^{\frac{1}{2}}} \quad (3.4)$$

El valor de correlación entre las series temporales de coeficientes de expansión izquierdos y derechos constituye una medida de la capacidad de reconstrucción de un campo a partir del otro. Con esto, la fracción que puede ser reconstruida del campo derecho usando N modos DVS es:

$$\hat{z}(t) = \sum_{k=1}^N \hat{b}_k(t) \mathbf{p}_k \quad (3.5)$$

Utilizando los patrones acoplados obtenidos del análisis DVS es posible reconstruir la parte linealmente dependiente de uno de los campos a partir del otro.

4. VALIDACIÓN CRUZADA DEL MODELO

La técnica de validación cruzada fue desarrollada por STONE (1974; 1977); posteriormente, EFRON (1983) y BUNKE y DROGE (1984) desarrollaron varios métodos para determinar el comportamiento estadístico de ésta. La validación cruzada presenta una importante ventaja frente a otras técnicas de validación al disponer del número total de datos en las sucesivas calibraciones del modelo para las validaciones. La técnica de validación cruzada es una técnica de muestreo en la que el modelo se calibra extrayendo secuencialmente un pequeño conjunto de observaciones sobre las que posteriormente se valida. Si n es el número total de datos disponibles, el modelo se calibra de manera sucesiva sobre j conjuntos de longitud $n-m$ y se aplica, asimismo, sobre j subconjuntos de longitud m . El número de particiones, j , y, por tanto, de validaciones cruzadas, puede llegar a ser de $n!/(m!(n-m)!)$. El modelo de validación cruzada se calibra sobre sucesivos subconjuntos de datos con el mismo número de grados de libertad, realizando las predicciones para cada observación de manera independiente. El método de validación cruzada es un método robusto que permite evaluar la capacidad predictiva del modelo y determinar si el modelo formulado presenta error de sesgo (MICHAELSEN, 1987).

En este trabajo, el análisis de validación cruzada se ha realizado extrayendo segmentos de tres meses correspondientes a un mismo año del conjunto total de datos. Para cada nueva validación se realiza el mismo proceso de manera independiente, de modo que se tienen tantas validaciones como años tiene el registro. Los experimentos de validación cruzada pretenden reproducir el campo de anomalías de precipitación estandarizadas observado a lo largo de todo el registro analizado. De cada una de las series de los observatorios que constituyen la base de datos de precipitación se han extraído simultáneamente las tres observaciones mensuales correspondientes a cada año del registro, para predecir cada uno de los datos extraídos. El nuevo conjunto de datos, constituido por los valores reconstruidos obtenidos para cada observatorio, se ha utilizado para elaborar un nuevo conjunto que abarque todo el registro y que constituye el campo de precipitación reconstruido. Se muestran a continuación los resultados obtenidos para cada estación.

Se han comparado los resultados obtenidos para el campo de PCP de los experimentos de validación cruzada independientes realizados con los tres campos de altura, Z300, U300 y V300. La figura 2 muestra los coeficientes de correlación lineal entre las series de precipitación reconstruidas y las observadas para cada observatorio frente a la latitud del mismo. Se ha resaltado mediante símbolos, el valor de correlación más alto en cada observatorio obtenido de los experimentos realizados con Z300 (círculo negro), U300 (triángulo rojo) y V300 (cuadrado verde), destacando de este modo el campo que mejor reproduce el comportamiento de la precipitación en cada uno de ellos.

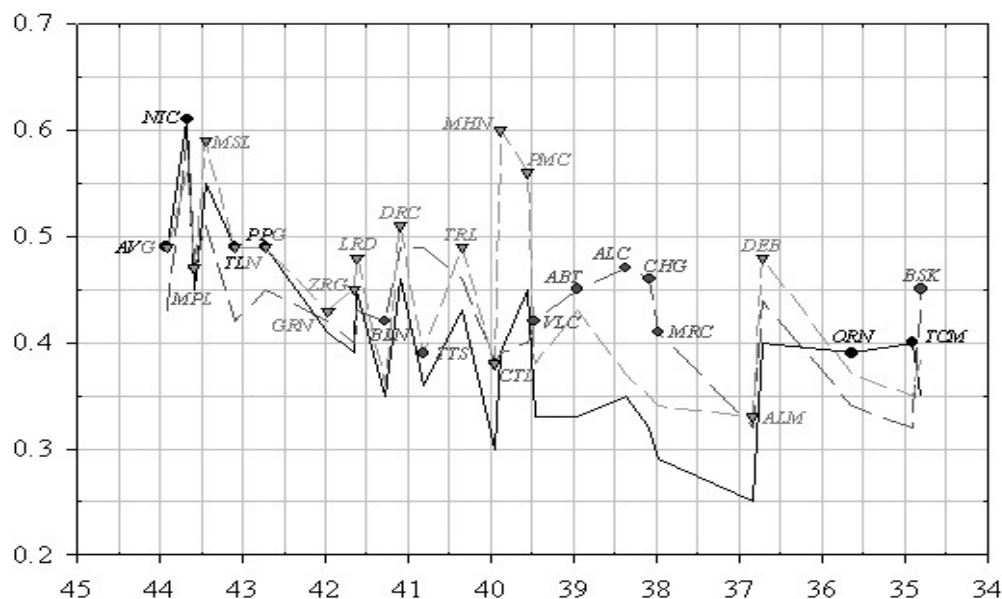


Fig. 2. Representación del coeficiente de correlación temporal, ρ , (eje de ordenadas), frente a la latitud ($^{\circ}$ N) del observatorio (eje de abscisas) entre las series de anomalías estandarizadas de PCP observada y las series reconstruidas con: Z300 (línea continua); U300 (línea discontinua larga) y V300 (línea discontinua corta).

Mediante símbolos se indica el campo que mayor correlación ofrece en cada uno de los observatorios

El papel del campo de viento zonal, U300, es comparativamente insignificante teniendo en cuenta el papel que juegan Z300 y V300. En la zona sur de la región, la variable V300 (línea y símbolos en color verde) juega el papel principal en relación con las variables de altura con la PCP. En esta zona, no se observa dependencia del valor de correlación con la latitud. El valor medio de correlación en esta zona es de 0,45. Los observatorios de Tortosa, Teruel, Albacete y los de las Islas Baleares, así como los observatorios norteafricanos a excepción de Tlemcem, muestran valores de correlación más altos, por encima del valor de correlación medio de la zona; las estimaciones para observatorios tales como Castellón, Alicante, Cehegim, Murcia y Almería las correlaciones están por debajo de 0,45.

Por otra parte, el porcentaje de varianza de las series reconstruidas (no mostrado) no supera el 30%, obteniéndose los valores más altos nuevamente en los observatorios del sur de Francia. El valor del sesgo de la serie correspondiente a Palma de Mallorca es próximo al 30% y tiene signo positivo, lo que indica que el valor medio de las anomalías reconstruidas es superior al de las anomalías observadas. Por otra parte, Murcia muestra un sesgo cercano al 25% en sentido inverso

al anterior. El resto de observatorios muestran valores del sesgo próximos a cero, lo que resulta indicativo de una reconstrucción sin errores sistemáticos.

5. CORRECCIÓN DEL MODELO DE RECONSTRUCCIÓN

En la sección anterior se ha visto que el comportamiento de la precipitación regional está gobernado por los campos de circulación a gran escala. Sin embargo, los resultados obtenidos con el modelo de reconstrucción de la PCP regional a partir de éstos no reproducen parte de la variabilidad de la PCP. Diversos autores han analizado la influencia de factores orográficos sobre el comportamiento del campo de precipitación (SOTILLO, 2003; QUADRELLI *et al.*, 2001). En esta sección, se introduce la influencia de este tipo de factores sobre los residuos del campo de anomalías de precipitación que no han sido reproducidos por el modelo de reconstrucción formulado a partir de los campos de circulación atmosférica en altura.

El modelo inicial de reconstrucción formulado con los campos de circulación analizados ha permitido reproducir, en esencia, la evolución del campo de anomalías de PCP. Sin embargo, el modelo únicamente reproduce parte de la amplitud de la oscilación del campo original de anomalías de PCP.

En esta sección se pretende determinar el efecto que los factores locales ejercen sobre la fluctuación de las anomalías de PCP. Para ello, se ha establecido que la razón entre el campo de anomalías observadas y reconstruidas con el modelo, es función dependiente de los siguientes factores locales en cada observatorio:

- latitud, φ .
- altitud, h .
- distancia al mar, d .

Cada una de las series reconstruidas debe ser corregida de la forma:

$$z_m^*(t) = C_m \hat{z}_m(t) \quad (5.1)$$

donde $\hat{z}_m(t)$ es la serie de anomalías de PCP reconstruidas en el observatorio m -ésimo con el modelo desarrollado; $z_m^*(t)$ es la serie de anomalías corregida en el mismo observatorio y C_m es un coeficiente de corrección local adimensional dependiente de φ , h y d , para cada observatorio. Este coeficiente ha sido calculado para cada observatorio, C_m , mediante un modelo de regresión lineal múltiple considerando como variable dependiente el vector, \mathbf{C} , constituido por los valores de la mediana de la razón entre el campo de anomalías de PCP original, $z_{i,m}(t)$, y el campo reconstruido con el modelo inicial, $\hat{z}_{i,m}(t)$, para cada observatorio, donde el subíndice i es el índice temporal y m indica el observatorio; las variables independientes han sido los vectores de latitud, $\boldsymbol{\varphi}$, altitud, \mathbf{h} , y distancia al mar, \mathbf{d} , de cada observatorio. El modelo de regresión lineal múltiple se plantea entonces para cada observatorio como:

$$C_m = \sum_{n=1}^3 a_n L_i(m) + a_0 \quad (5.2)$$

donde C_m es el coeficiente de corrección local obtenido para el observatorio m -ésimo; a_i son los coeficientes del modelo de regresión y $L_i(m)$ son las variables locales en este observatorio.

Una vez determinado el modelo de regresión, se ha efectuado un test t-Student para determinar, a un nivel de confianza del 95%, si los coeficientes del modelo para cada una de las variables independientes son significativamente distintos de cero. El coeficiente de corrección local a cada observatorio, viene descrito por la siguiente expresión:

$$C_m = 3.83 \cdot 10^{-4} h(m) + 1.25 \quad (5.3)$$

donde $h(m)$ es el valor de la altitud en unidades de m, de cada observatorio, y el coeficiente de regresión tiene unidades de m^{-1} ; los coeficientes de las variables latitud y distancia al mar han sido eliminados de la expresión 5.3 porque no han resultado significativamente distintos de cero en el test de significación realizado. El coeficiente de determinación obtenido por el modelo es $R^2 = 0,53$.

La figura 3a muestra la distribución espacial de los valores de C_m obtenidos a partir de la expresión 5.3. Asimismo, la figura 3 (b, c y d) muestra el patrón de varianza de las series corregidas.

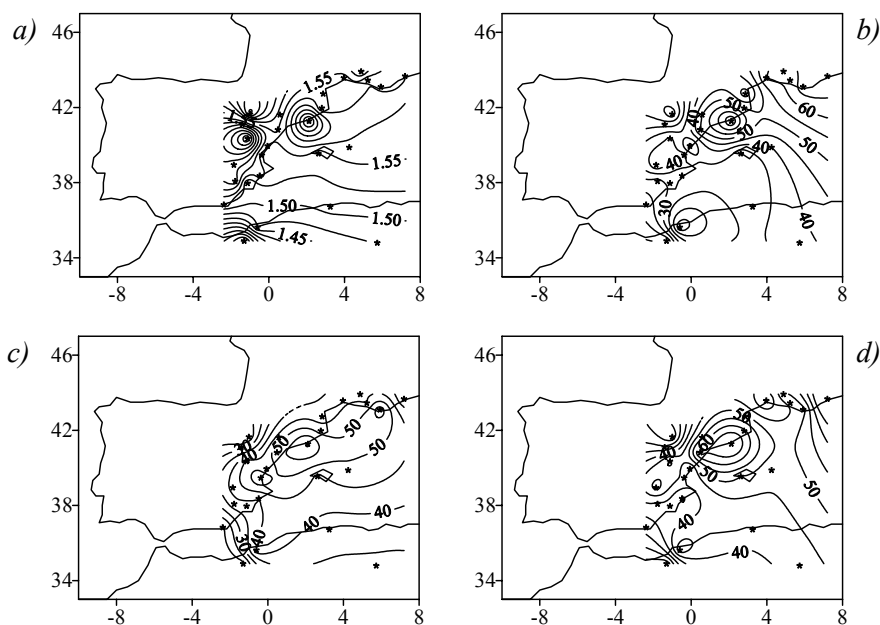


Fig. 3. Distribución espacial de: a) valores del coeficiente de corrección local, C_m ; b), c) y d) varianza de las series reconstruidas corregidas de anomalías estandarizadas de PCP empleando los campos Z300, U300 y V300, respectivamente

Se observa, que los valores de C_m tienen un marcado carácter local dependiente de la altitud en cada punto. Por otra parte, en toda la región se observa un aumento considerable de la varianza explicada por el nuevo conjunto de series corregidas. Asimismo, los resultados del modelo de reconstrucción inicial permite diferenciar dos regiones norte y sur identificadas por los valores de correlación máximos en función del predictor Z300 y V300, respectivamente. Los valores más altos de varianza explicada por las series obtenidas a partir de Z300 se obtienen de nuevo para las estaciones del norte de la región, mientras que los más bajos se obtienen en las situadas al suroeste. Asimismo, la figura 2c, correspondiente a los resultados de U300, muestra notable mejora para la zona del noreste peninsular. El patrón de varianza correspondiente a la variable V300 no muestra un patrón de distribución espacial tan definido como en los otros dos casos.

Una comparación más exhaustiva entre estimaciones y observaciones puede realizarse comparando la función de distribución acumulada de cada variable, ya que ésta caracteriza la distribución completa del conjunto de datos y no depende del número de observaciones ni de la función de distribución de éstas. Esta comparación se lleva a cabo mediante la representación gráfica de los cuantiles de ambos conjuntos. Si éstos se distribuyen de la misma forma, el gráfico resultante será una línea recta con pendiente unidad que pasa por el origen. Si una muestra es función lineal de la otra, su gráfico será también lineal pero con diferente pendiente y diferente origen. Esta metodología constituye un método de validación no paramétrico que no requiere ninguna suposición acerca de la distribución estadística de los datos. La representación de los cuantiles (percentiles) de las estimaciones realizadas frente a los de las observaciones, viene dada por los gráficos cuantil-cuantil (C-C).

En la figura 4 se muestran los gráficos C-C de las series observadas frente a las series reconstruidas por el modelo inicial (izquierda) por un lado, y, por otro, frente a las series corregidas (derecha). Se analiza, de este modo, la mejora introducida por los factores locales en la predecibilidad de anomalías evaluando asimismo, si se produce infra o sobreestimación de las precipitaciones. Analizando los resultados obtenidos con el modelo inicial (Fig. 4, izquierda) se observa un comportamiento similar en todos los casos; el modelo reproduce una pequeña parte de los datos situados en el centro de la distribución con relativa exactitud, infraestimando los valores máximos y sobrestimando los mínimos.

En la figura 4 (derecha) puede observarse que las series corregidas se ajustan con mayor fidelidad a las observaciones ya que la mayor parte de los percentiles se encuentra próximos a la recta. Las estimaciones de los valores mínimos son más exactas para todas las series que las de los máximos. Se observa que aproximadamente todos los valores situados por debajo del percentil del 85% son reproducibles con exactitud, mientras que los situados por encima de este se encuentran infraestimados.

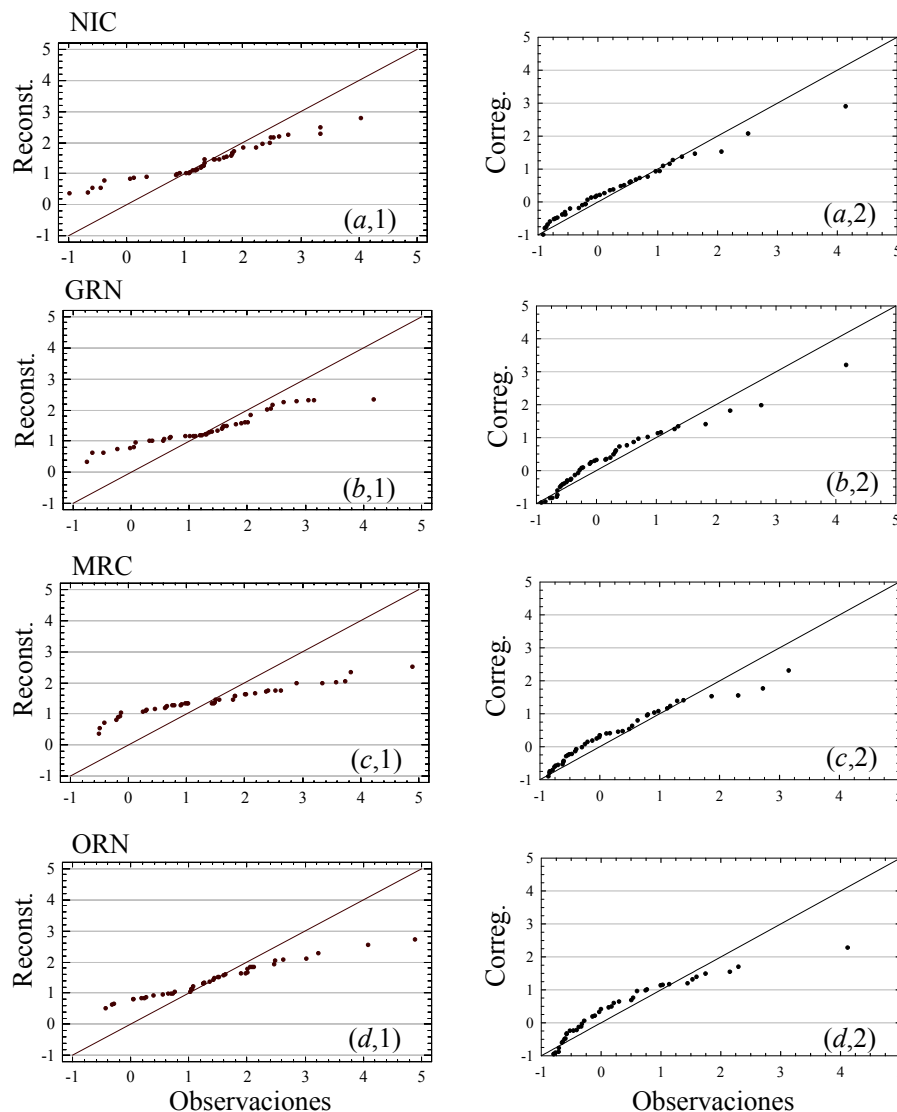


Fig. 4. Resultados del proceso de validación cruzada. Izquierda: representación C-C de las observaciones (eje de abcisas) frente a las reconstrucciones corregidas (eje de ordenadas). Derecha: representación C-C de las observaciones (eje de abcisas) frente a las reconstrucciones sin corregir (eje de ordenadas) realizadas con: (a y b) Z300 y (c y d) V300

6. CONCLUSIONES

A partir de los modos acoplados de variabilidad se ha demostrado la influencia de la circulación en altura sobre la precipitación regional en esta zona de tan difícil predicción. En la alta troposfera la circulación atmosférica se encuentra vinculada a la estructura del campo de precipitación regional estudiado. Asimismo, se ha comprobado que, en el nivel de 300 hPa, el campo de altura geopotencial constituye la variable de gran escala que estima con mayor exactitud las anomalías de precipitación en primavera.

Por otra parte, la consideración de factores locales en cada observatorio, tales como altitud y latitud, permite mejorar la conexión de la señal en altura con la precipitación regional.

7. AGRADECIMIENTOS

Este trabajo se ha llevado a cabo gracias a la financiación del Ministerio de Ciencia y Tecnología, mediante el proyecto de investigación REN2003-03647.

8. REFERENCIAS

- BRETHERTON, C.S., SMITH, C. y WALLACE, J.M. (1992). "An intercomparison of methods for finding coupled patterns in climate data". *J. Climate*, 5, pp. 541-560.
- BUNKE, O. y DROGE, B. (1984). "Bootstrap and cross-validation estimates of the precipitation error for linear regression models". *Ann. Stat.*, 12, pp. 1400-1424.
- EFRON, B. (1983). "Estimating the error rate of a prediction rule: Improvement on cross-validation". *J. Amer. Stat. Assoc.*, 78, pp. 316-331.
- GONZÁLEZ-ROUCO, J.F. *et al.* (2001). "Quality Control and Homogenization of Monthly Precipitation Data in the Southwest of Europe". *J. Climate*, 14, pp. 964-978.
- HOUZE, R.A. Jr. (1993). *Cloud Dynamics*. Academic Press, San Diego. 573 pp.
- KALNAY, E. *et al.* (1996). "The NCEP/NCAR 40-years Reanalysis, Project". *Bull. Amer. Meteorol. Soc.*, 77(3), pp. 437-471.
- MARTIN, M.L. *et al.* (2004). "North atlantic teleconnection patterns of low-frequency variability and their links with spring-time precipitation in the western Mediterranean". *Int. J. Climatol.*, 24, pp. 213-230.
- MICHAELSEN, J. (1987). "Cross-validation in statistical climate forecast models". *J. Clim. Appl. Meteorol.*, 26, pp. 1589-1600.
- MILLAN, M.; ESTRELA, M.J. y CASELLES, V. (1995). "Torrential precipitation on the Spanish east coast: The role of the Mediterranean sea surface temperature". *Atmos. Res.*, 36, pp. 1-16.
- MORATA, A. (2004). *Influencia de la circulación atmosférica sobre la precipitación en la cuenca mediterránea occidental*. Tesis Doctoral, Universidad Complutense de Madrid. 193 pp.
- QUADRELLI, R.; PAVAN, V. y MOLTENI, F. (2001). "Wintertime variability of Mediterranean precipitation and its links with large-scale circulation anomalies". *Clim. Dyn.*, 17, pp. 457-466.
- SOTILLO, M.G. *et al.* (2003). "Role of orography in the spatial distribution of the precipitation over the Spanish Mediterranean zone". *Clim. Res.*, 23, pp. 247-261.
- STONE, M. (1974). "Cross-validated choice and assessment of statistical predictions". *J. R. Statist. Soc.*, B36, pp. 111-147.
- STONE, M. (1977). "Asymptotics for and against cross-validation". *Biometrika*, 64, pp. 29-38.
- VALERO, F. *et al.* (2004). "Coupled modes of large-scale climatic variables and regional precipitation in the western Mediterranean in autumn". *Climate Dynamics*, 22, pp. 307-323.
- WALLACE, J.M.; C. SMITH y C. S. BRETHERTON (1992). "Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies". *J. Climate*, 5, pp. 561-576.
- WIDMANN, M. y SCHÄR, C. (1997). "A principal component and long-term trend analysis of daily precipitation in Switzerland". *Int. J. Climatol.*, 17, pp. 1333-1356.