

# VALIDACIÓN DE LA BASE DE DATOS RETICULAR DE LA AEMet: TEMPERATURA DIARIA MÁXIMA Y MÍNIMA

M. Yolanda LUNA<sup>1</sup>, Ana MORATA<sup>1</sup>, M. Luisa MARTIN<sup>2</sup>, Daniel SANTOS-MUÑOZ<sup>1</sup>,  
Javier DE LA CRUZ<sup>3</sup>

<sup>1</sup> *Agencia Estatal de Meteorología, Madrid.*

<sup>2</sup> *Escuela Universitaria de Informática de Segovia. Universidad de Valladolid*

<sup>3</sup> *Consejería de Educación. Comunidad de Madrid.*

yluna@inm.es

## RESUMEN

Para cubrir la necesidad de conjuntos de datos reticulares, de alta resolución espacial y a escala temporal diaria, de elementos climatológicos tales como temperatura y precipitación, el Área de Climatología de la Agencia Estatal de Meteorología ha desarrollado una base de datos reticular de temperatura máxima y mínima con un espaciado de 25x25 km y un registro temporal desde el 1 de enero de 1931 hasta el 31 de diciembre de 2006. Estos datos deben ser validados y deben ofrecerse junto con una medida de su precisión. Un método clásico consiste en eliminar del análisis espacial con el que se efectúa la interpolación unos pocos e independientes observatorios, y posteriormente utilizarlos para validar. Así, se mantienen la mayoría de las observaciones en la interpolación. En este trabajo el objetivo principal es la validación de los datos de temperatura diaria máxima y mínima, y por ende los valores mensuales y anuales de la base de datos reticular para dar una estimación de su calidad, es decir, del grado de correspondencia entre las estimaciones reticulares y las observaciones. El estudio de validación se realizará mediante una comparación estadística exhaustiva entre los observatorios independientes y el punto de retícula más cercano, estimando además las medidas de error más habituales: el sesgo, el error absoluto medio, el error cuadrático medio y el coeficiente de correlación lineal. Una vez validados, estos datos serán de gran utilidad en la obtención de escenarios de cambio climático, en el estudio de la variabilidad climática y, en general, en los modelos de evaluación de riesgos climáticos.

**Palabras clave:** Temperatura máxima, temperatura mínima, interpolación espacial, datos reticulares, validación, Península Ibérica

## ABSTRACT

In order to fulfil the necessity for climatological data onto a regular grid, the Climatic Area of Spanish Meteorological Agency (Agencia Estatal de Meteorología) has developed a regular database by spatial interpolation of maximum and minimum daily temperature. All available observed daily data for the period from 1 January 1931 to 31 December 2006 is used in the interpolation over a 25x25 km regular grid covering the Iberian Peninsula and the Balearic Islands. In order to assess estimation uncertainty, an independent validation procedure has to be carried out and a description of the precision of the grid database should be provided. A classical approach for this procedure is to split the data sample in two parts, and use those to parts for estimation and validation respectively. The objective of this study is to provide a

validation of the maximum and minimum temperature of the grid database giving a measure of the correspondence between the absolute values of observed and estimated data. Classical similarity measures between observed data and the estimated one in the closest grid point are used to validate the estimated data: bias, mean absolute error, mean quadratic error and correlation coefficient. After validation, the grid daily temperature data will be very useful in general for climatic risk analysis models and in particular for climatic variability and climate change scenarios studies.

**Key words:** Maximum temperature, minimum temperature, spatial interpolation, validation, Península Ibérica

## 1. INTRODUCCION

En la actualidad existe una necesidad creciente de conjuntos de datos reticulares, de alta resolución espacial y a escala temporal diaria, de elementos climatológicos tales como temperatura y precipitación. El desarrollo de los datos reticulares es un paso previo y absolutamente necesario en la realización de cualquier estudio climático, especialmente en el análisis de la variabilidad climática observada y en la obtención de escenarios de cambio climático regionalizados, siendo estos objetivos de vital importancia en la actualidad debido al incremento de origen antropogénico de gases de efecto invernadero (IPCC, 2007). Además, estos datos climáticos reticulares son imprescindibles para alimentar modelos hidrológicos, ecológicos, y en general, de evaluación de riesgos naturales.

Existe una discrepancia entre los datos climáticos disponibles, normalmente series tomadas en estaciones irregularmente distribuidas en el espacio y en el tiempo, y los datos necesarios para los modelos y los estudios de variabilidad, que deberían estar regularmente distribuidos y sin huecos en las series. De hecho, existe una necesidad creciente de bases de datos climáticos de precipitación y temperatura distribuidos en una retícula regular. Además, el análisis de la variabilidad climática pasada y futura necesita datos observados en varias escalas espacio-temporales. Para paliar este problema, el Área de Climatología de la Agencia Estatal de Meteorología (AEMet) ha generado una base de datos a escala diaria interpolando los datos observados en una retícula regular de 25 x 25 km (LUNA et al., 2004). El tamaño de la retícula no es tan pequeño como para necesitar un cálculo computacional excesivo, pero sí adecuado para estudios de variabilidad climática incluyendo la evaluación de los impactos de cambio climático en España. Se hace necesario, por tanto, construir bases de datos de buena calidad no solo para detectar el fenómeno, sino también para validar los modelos climáticos con los que se intenta predecir.

La homogeneidad de la base de datos reticular ya fue analizada en ALMARZA y LUNA (2006). Es este trabajo se concluyó que, al igual que las series de datos observados, los datos de la base reticular presentan inhomogeneidades y tendencias. Por otra parte, esta base de datos ha sido comparada con los las bases de datos de reanálisis a escala global del ECMWF y del NCEP, así como con los datos hindcast regionales HIPOCAS (SOTILLO et al., 2006, VALERO et al., 2008). Estos estudios han demostrado que análisis climáticos a escala regional realizados con los datos reticulares son más realistas que los obtenidos con los datos globales, demostrando la utilidad de este tipo de datos. Sin embargo, esta base de datos debe ser validada y debe ofrecerse junto con una medida de su precisión. Un método clásico consiste en eliminar

del análisis espacial con el que se efectúa la interpolación unos pocos e independientes observatorios, y posteriormente utilizarlos para validar. Así, se mantienen la mayoría de las observaciones en la interpolación. El objetivo de este trabajo es analizar la calidad de los datos de temperatura máxima y mínima diarios contenidos en la base de datos reticular de LUNA et al. (2004), realizando un estudio de validación estadística de los mismos frente a observaciones. Para ello, se construyó una nueva base reticular en la que no se incorporaron un conjunto de observatorios, que como datos independientes se han utilizado para la validación. Estos observatorios son El puerto de Navacerrada (Madrid), Palma de Mallorca Aeropuerto, A Coruña, Alicante y San Sebastián. El trabajo está organizado de la siguiente manera: en la Sección 2 se describen los datos utilizados, los resultados más relevantes se discuten en la Sección 3, finalizando con la exposición de las conclusiones más destacables de este estudio en la Sección 4.

## 2. DATOS

Se utiliza la base de datos climática diaria obtenida a partir de observaciones in situ provenientes de la red de la AEMet. El Área de Climatología ha elaborado una base de datos reticular mediante la interpolación estadística de estas observaciones in situ. Con esta base de datos reticular se trata de contar con observaciones diarias completas necesarias como datos de entrada de modelos medioambientales y para los estudios de variabilidad climática a escala regional diaria. Todos los datos termoplumiométricos contenidos en la Base de Datos Climática Histórica de la AEMet fueron extraídos para el periodo comprendido entre el 1 de enero de 1931 y el 31 de diciembre de 2006. El número de observaciones disponible no es constante con el tiempo, depende de la fecha como puede observarse en la Figura 1. Aún así, las observaciones disponibles, irregularmente distribuidas en el espacio, proporcionan una buena cobertura sobre el dominio como se puede observar en el ejemplo de la Figura 2a.

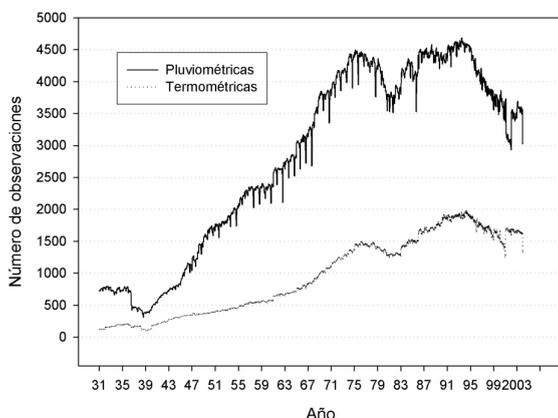


Fig. 1: Evolución temporal del número de observatorios.

Se eligió un tamaño de retícula de 25 km (Figura 2b) debido a que dicha resolución era apropiada para los modelos de evaluación de riesgos medioambientales y para los estudios de variabilidad climática, incluyendo los de impactos del cambio climático en España. El método geoestadístico utilizado para llevar a cabo la interpolación fue el Kriging debido a que preserva mayor cantidad de varianza original de los datos observados que otros métodos estadísticos (SHEN et al., 2001). Esta técnica se encuentra implementada en todos los Sistemas de Información Geográfica utilizados para el análisis espacial. Además, el Kriging es ampliamente utilizado en otros países de nuestro entorno, permitiendo comparaciones entre bases de datos. Todas estas consideraciones se subrayan en los objetivos de la Acción COST719 (The use of GIS in Meteorology and Climatology) en la que se participó activamente. Los datos contenidos en la base de datos reticular, así construida, son campos de precipitación y temperatura máxima y mínima a escala diaria en un periodo comprendido entre 1931 y 2006, ambos incluidos. Más información acerca de la base de datos reticular puede consultarse en MORATA et al. (2006) y en LUNA y ALMARZA (2007).

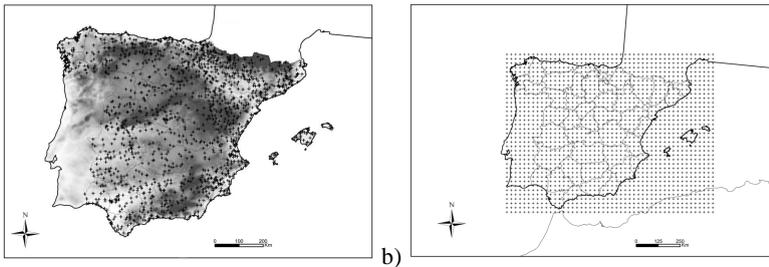


Fig. 2: a) Distribución espacial de las observaciones termométricas el 31 de diciembre de 2006; b) Disposición espacial de los puntos en la retícula de 25 km.

### 3. RESULTADOS DE LA VALIDACIÓN

En este apartado se mostraran los resultados de la validación de los datos de temperatura máxima y mínima contenidos en la base reticular, ya que dichos datos deben ofrecerse junto con una medida de su precisión. La validación se realizará mediante la eliminación del análisis espacial con el que se efectúa la interpolación unos pocos e independientes observatorios, que posteriormente serán utilizados para validar. Así, se mantienen la mayoría de las observaciones en la interpolación y se modifica muy poco el producto final. El estudio de validación se realizará mediante una comparación estadística exhaustiva entre los observatorios independientes y el punto de retícula más cercano, estimando además las medidas de error más habituales: el sesgo, el error absoluto medio, el error cuadrático medio y el coeficiente de correlación lineal, así como gráficos representativos de su distribución estadística tales como los Box-Whisker Plots.

#### 3. 1. Validación de los datos mensuales

En las Tablas 1 y 2 se muestran los valores de las medidas de error calculadas para los cinco observatorios de temperatura máxima y mínima mensual, respectivamente. Se puede notar como los errores, tanto absoluto medio como cuadrático medio, son bajos, inferiores o cercanos

a 1°C, en todos los casos salvo en el observatorio de Navacerrada. Igualmente el sesgo, calculado como los valores observados menos valores calculados en retícula, presenta valores bajos excepto en Navacerrada. Los valores de correlación son todos superiores a 0.95, aunque el más bajo es 0.96 se obtiene también para el mismo observatorio. Esta estación, situada a 1890m de altitud, se ha elegido en el estudio de validación por ser uno de los pocos observatorios situados por encima de los 1000m de los que se dispone de una serie climática relativamente larga.

Los errores más grandes se detectan en los valores de temperatura máxima en Navacerrada con un sesgo de -4.1°C, indicando que los datos de la base reticular sobrestiman los valores reales. La sobrestimación se repite en los valores de temperatura mínima pero en menor medida, -1.3°C. En los gráficos Box-Whisker (Figura 3a) se observa claramente esta sobrestimación en el caso de Navacerrada. A modo de ejemplo y por razones de brevedad, se ilustra en la Figura 3b el caso de San Sebastián, en el que se denota claramente como las distribuciones observada y calculada tanto para la mínima como para la máxima son prácticamente idénticas.

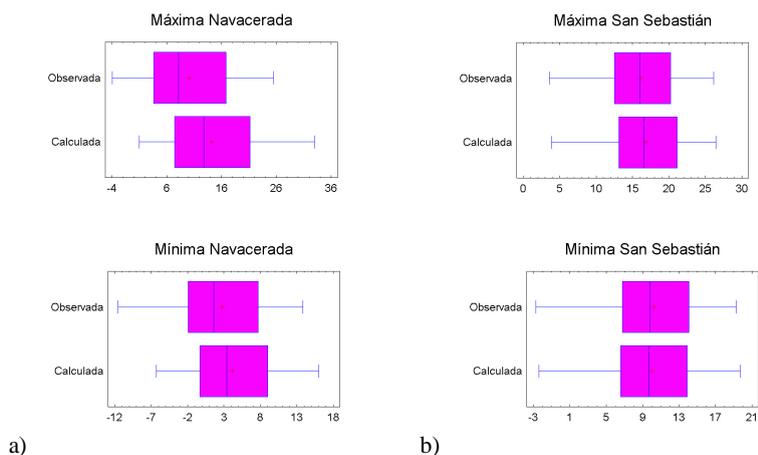


Fig. 3: Gráficos Box-Whisker para las series observadas y calculadas en el punto más cercano de temperaturas máxima y mínima diarias de a) Navacerrada y b) San Sebastián.

	A Coruña	Alicante	Navacerrada	Palma	San Sebastián
Sesgo	-0.1	0.3	-4.1	0.1	-0.7
EAM	0.8	0.5	4.1	0.2	0.8
ECM	1.0	0.7	4.6	0.3	0.9
R	0.97	0.99	0.96	0.99	0.99

Tabla 1: VALORES DE SESGO (EN °C), ERROR ABSOLUTO MEDIO (EAM, EN °C), ERROR CUADRÁTICO MEDIO (ECM, EN °C) Y COEFICIENTE DE CORRELACIÓN (R) PARA LA TEMPERATURA MÁXIMA MEDIA MENSUAL

	A Coruña	Alicante	Navacerrada	Palma	San Sebastián
Sesgo	1.2	-0.5	-1.3	-0.6	0.2
EAM	1.3	0.7	1.5	0.6	0.3
ECM	1.5	1.2	2.0	1.0	0.4
R	0.96	0.98	0.96	0.99	0.99

Tabla 2: IGUAL QUE LA TABLA 1 PERO PARA LA TEMPERATURA MÍNIMA MEDIA MENSUAL

Para evaluar si las funciones de distribución son iguales se realiza un análisis estadístico mediante los test no paramétricos de Kolmogorov-Smirnov y de Mann-Whitney. El primero de ellos compara las distribuciones de frecuencia de dos muestras calculando la máxima distancia entre las funciones de distribución acumulada, siendo su hipótesis nula que ambas distribuciones son idénticas, mientras que el segundo plantea la hipótesis nula de igualdad de la mediana de ambas muestras. Aplicando estos tests a los datos observados y calculados tanto de temperatura máxima y como de temperatura mínima, los resultados conducen a la aceptación de las hipótesis a un nivel de confianza del 95% en todos los casos, salvo en el observatorio de Navacerrada, por lo que se consideran que las series observadas y calculadas pertenecen a la misma muestra, no habiendo diferencias estadísticamente significativas.

### 3. 2. Validación de los datos diarios

Por razones de brevedad y dado los resultados obtenidos en el apartado anterior, en el estudio de validación del dato diario mostraremos solo los resultados de Navacerrada ya que ofrecía mayores errores. En la Tabla 3 se muestran los valores de los parámetros de la serie observada de Navacerrada y la calculada en el punto de retícula más cercano para la temperatura máxima diaria. Lo más destacable es el sesgo en los valores de forma que los pertenecientes a la base reticular sobreestima la media y la mediana alrededor de 4°C, de forma similar a los datos mensuales. Además, los valores extremos también están sobreestimados. Sin embargo, la varianza y, por ende, la desviación estándar son iguales en ambas series debido a que el procedimiento de interpolación espacial, Kriging, conserva estos parámetros, presentando la serie de diferencias muy poca varianza.

	Observada	Calculada	Observada - Calculada
Media	10.1	14.2	-4.1
Mediana	9.2	13.1	-3.6
Varianza	72.9	72.8	6.1
Desviación estándar	8.5	8.5	2.5
Mínimo	-14.4	-7.9	-18.6
Máximo	31.8	39.3	6.3

Tabla 3: VALORES DE DISTINTOS PARÁMETROS PARA LA TEMPERATURA MÁXIMA DIARIA OBSERVADA EN NAVACERRADA, CALCULADA EN EL PUNTO DE GRID MAS CERCANO Y SU DIFERENCIA.

El gráfico Box-Whisker (Figura 4a) nos muestra claramente la sobrestimación de los datos de retícula de la temperatura máxima diaria. Representando ambos conjuntos de datos en la Figura 4b se observa que existe una fuerte relación lineal entre ambos, siendo la ordenada en el origen de  $-3.6^{\circ}\text{C}$ , indicativa del sesgo, la pendiente es 0.96, muy próxima al valor ideal 1, y el coeficiente de correlación de 0.95.

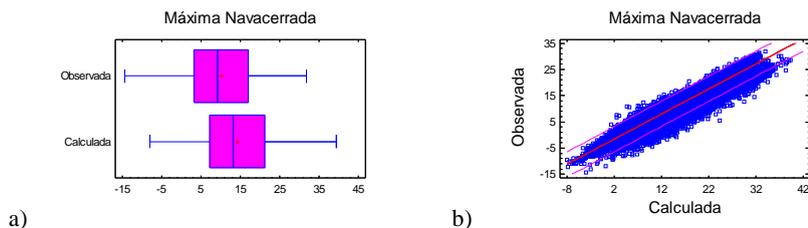


Fig. 4: a) Gráficos Box-Whisker para las series de temperatura máxima diaria observada en Navacerrada y calculada en el punto más cercano; b) Relación lineal entre ambas series.

	Observada	Calculada	Observada - Calculada
Media	2.7	4.1	-1.4
Mediana	2.2	3.6	-1.2
Varianza	42.9	36.3	5.6
Desviación estándar	6.5	6.0	2.4
Mínimo	-20.3	-17.3	-16.31
Máximo	20.6	22.9	9.7

Tabla 4: IGUAL A LA TABLA 1 PERO PARA LA TEMPERATURA MÍNIMA DIARIA

En la Tabla 4 se puede observar como el comportamiento se repite para la temperatura mínima diaria en Navacerrada, siendo los valores de sesgo mucho menores, del orden de  $-1.6^{\circ}\text{C}$ , al igual que en los datos mensuales. Igualmente se detecta como los valores de varianza se mantienen. El gráfico Box-Whisker (Figura 5a) muestra una ligera sobrestimación de los datos de retícula de la temperatura mínima diaria. Representando ambos conjuntos de datos en la Figura 4b se observa que existe una fuerte relación lineal entre ambos, siendo la ordenada en el origen de  $-1.4^{\circ}\text{C}$ , indicativa del sesgo, la pendiente es 1.01, muy próxima al valor ideal 1, y el coeficiente de correlación de 0.93. En este caso el coeficiente de correlación, aunque bastante bueno, es algo inferior a la de la temperatura máxima debido a que hay más datos anómalos fuera de las curvas de confianza en ambos paneles de la Figura 5.

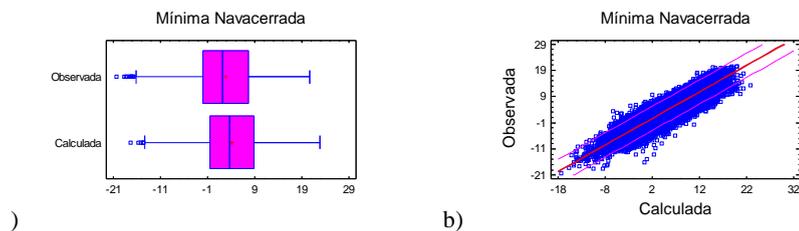


Fig. 5: Igual a la Figura 4 excepto para la temperatura mínima diaria.

### 3. 3. Comparación con el Atlas Climático Nacional

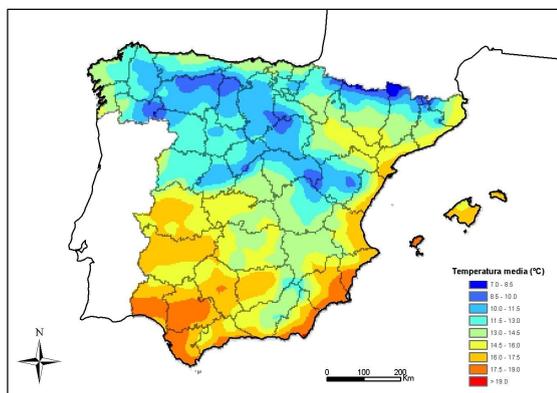
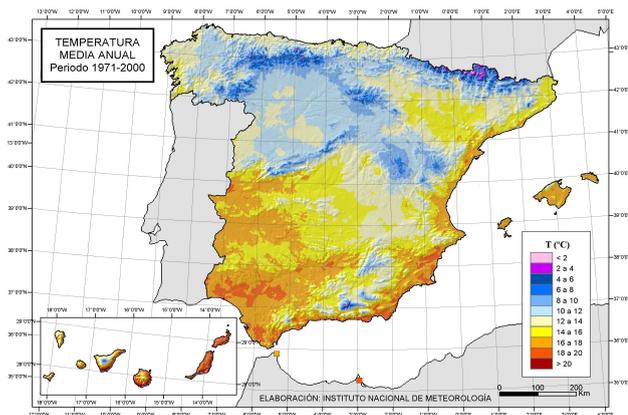
En la última edición del Atlas Nacional de España, se incluyeron mapas de los valores de los elementos climáticos para el periodo de 30 años 1971-2000, que aunque no corresponde al periodo estándar, permite dar una visión actual del clima en España (IGN, 2004). Para el caso de la temperatura, los mapas del atlas se han construido mediante interpolación espacial de todas las observaciones por el método de Kriging residual. Este método consiste en aplicar inicialmente un modelo de regresión lineal múltiple entre la temperatura y parámetros fisiográficos: altitud, latitud y distancia al mar. Los residuos resultantes se interpolan con Kriging ordinario, y al mapa de anomalías resultante se le suma un campo estimado por la ecuación de regresión y un modelo digital del terreno. El mapa obtenido para el Atlas (Figura 6a) es muy similar al que se obtiene a partir de los datos de temperatura media reticular (Figura 6b), las zonas más cálidas y más frías presentan una distribución espacial prácticamente idéntica. Las diferencias más notables se encuentran en las zonas situadas en las principales cadenas montañosas, ya que los datos reticulares no alcanzan el nivel de detalle que se obtiene del mapa del Atlas. Esto es claramente debido a que para el Atlas se han tenido en cuenta parámetros fisiográficos que no se han considerado en el método de interpolación usado en la construcción de la base reticular.

## 4. CONCLUSIONES

El estudio de validación de los datos reticulares de temperatura máxima y mínima diarias para el área Peninsular y Baleares ha mostrado como dichos datos presentan una calidad aceptable con errores promedios cercanos o inferiores a 1°C, excepto en el observatorio de Navacerrada. Para la temperatura máxima en Navacerrada se obtiene el máximo sesgo, -4.1°C, indicando que los datos de la base reticular sobrestiman los valores reales. Aunque este comportamiento se repite con la temperatura mínima, el sesgo es menor, -1.3°C. La peor calidad en la estimación en este observatorio es debida a que no se han considerado parámetros fisiográficos, tales como la altitud, en el método de interpolación usado en la construcción de la base reticular.

Esta base reticular ha dado como resultado un conjunto de datos de mucha utilidad, pero que muestran ciertos signos de inhomogeneidad tanto espacial como temporal. El siguiente paso sería por tanto, la mejora de la base de datos climatológicos reticulares de alta resolución espacial de temperatura y precipitación a escala diaria para su uso en la obtención de escenarios de cambio climático, en el estudio de la variabilidad climática y, en general, para la

evaluación de riesgos climáticos. Se trata de cubrir la necesidad de disponer de datos climáticos de buena calidad, homogéneos y fiables que proporcionen robustez a los estudios climáticos. Además, se deberá tener en cuenta la singularidad geográfica de nuestro territorio caracterizado por una orografía muy compleja y distinta influencia marítima atlántica o mediterránea.



b)

Fig. 6: Temperatura media (°C) para el periodo 1971-2000: a) Atlas Nacional de España; b) Datos reticulares.

## 9. AGRADECIMIENTOS

Este trabajo ha sido parcialmente financiado por el proyecto de investigación CGL2007-61328/CLI. Los autores agradecen a los integrantes del equipo de Área de Climatología su inestimable colaboración en la obtención de los datos y preparación de algunos gráficos.

## 10. REFERENCIAS

- ALMARZA C. y LUNA M. Y. (2006). "Homogeneidad y variabilidad de la precipitación y la temperatura en zonas climáticamente homogéneas de la Península Ibérica". Proceedings of XXIX Jornadas Científicas AME. Pamplona. Spain, 24-26 abril, 2006.
- IGN (2004). Atlas Nacional de España. Sección II Climatología. Ed. Centro Nacional de Información Geográfica.
- IPCC (2007). *Climate Change 2007. The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* Eds. Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp.
- LUNA M.Y. y ALMARZA C. (2007). Interpolation of 1961-2003 daily climatic data in Spain. En: Szalai, S.; Bihari, Z.; Szentimrey T. and Lakatos, M. (Eds.). *Spatial Interpolation in Climatology and Meteorology (ISBN 92-898-0033-X)*. COST Office, pp. 189-196.
- LUNA, M.Y., MORATA, A., AVELLO E. y PERAZA B. (2004). Estructuras Espaciales Coherentes de Precipitación en la España Peninsular: Régimen Atlántico. Proceedings of XXVIII Jornadas Científicas AME (ISBN 84-8320-261-1). Badajoz. Spain, 11-13 February, 2004.
- MORATA, A., MARTÍN, M. L., LUNA, M. Y. y VALERO F. (2006). "Self-Similarity patterns of precipitation in the Iberian Peninsula". *Theoretical and Applied Climatology*, 85, pp. 41-59.
- SHEN, S.P.; DZIKOWSKI, P.; LI G. y GRIFFITH D. (2001). "Interpolation of 1961-97 daily temperature and precipitation data onto Alberta Polygons of ecodistrict and soil landscapes of Canada". *J. Appl. Meteor.*, 40, pp. 2162-2177.
- SOTILLO M.G., MARTÍN M.L., VALERO F. y LUNA M.Y. (2006): "Validation of a homogeneous 41-year (1961-2001) winter precipitation hindcasted dataset over the Iberian Peninsula: assessment of the regional improvement of global reanalysis". *Climate Dynamics*, 27, pp. 627- 645.
- VALERO F., MARTÍN M. L., SOTILLO M. G., MORATA A. y LUNA M. Y. (2008). "Characterization of the autumn Iberian precipitation from long-term data sets: comparison between observed and hindcasted data". *International Journal of Climatology*. In press.