

## **ESTUDIO COMPARATIVO DE DISTINTAS BASES DE DATOS DE HUMEDAD DEL SUELO SUPERFICIAL EN LA PENÍNSULA IBÉRICA**

Emilio ROMERO JIMÉNEZ, Matilde GARCÍA-VALDECASAS OJEDA, Patricio YESTE DONAIRE, Sonia Raquel GÁQMIZ-FORTIS, Yolanda CASTRO-DÍEZ, María Jesús ESTEBAN-PARRA  
*Dpto Física Aplicada, Facultad de Ciencias, Universidad de Granada*  
**emiliorj@ugr.es**

### **RESUMEN**

Existen numerosas bases de datos con parámetros que permiten estudiar diversas variables, tanto climáticas como hidrológicas. Una posible aplicación del estudio de estas variables es la validación de datos provenientes de simulaciones, permitiendo establecer la habilidad que proporcionan los modelos de cara a realizar simulaciones de futuro. Esta validación puede ser especialmente complicada en lo referente a humedad del suelo debido a la dificultad que entraña la obtención de datos observacionales. Además, distintas bases de datos expresan la humedad del suelo de forma diferente, por ejemplo, como volumen de agua en un volumen de suelo o como capa de agua equivalente. Por ello, los datos de humedad del suelo no siempre son fácilmente comparables.

Así, el objetivo de este trabajo es la comparación de diversas bases de datos de humedad superficial del suelo, provenientes fundamentalmente de datos satelitales y de modelos, para la Península Ibérica (PI). Para ello, se han comparado los patrones de variabilidad espacio-temporal de humedad del suelo aportados por estas bases de datos y se han realizado test estadísticos para determinar si hay correlación entre ellos y conocer las limitaciones de cada conjunto de datos.

Se ha comprobado que, en general, todas las bases de datos presentan tendencias similares, si bien los valores absolutos de humedad del suelo son diferentes. En algunos casos, se aprecia que existe una importante correlación entre algunos conjuntos de datos. Esto puede ser debido al uso de modelos de suelo para reescalar y homogeneizar los datos observados.

**Palabras clave:** humedad del suelo, datos de satélite, modelo, simulación, Península Ibérica.

### **ABSTRACT**

The existence of several databases containing climate and hydrologic parameters allows the study of their variables. A possible application of the study of these variables is the validation of data obtained from simulations, which makes model skill evaluation possible. This is useful when simulating future scenarios. However, this validation can be especially problematic for soil moisture data due to the difficulty inherent in its acquisition. Moreover, databases use different magnitudes for this

variable, e.g. volume of water contained in the soil or thickness of equivalent water layer. Therefore, data cannot be easily compared.

The aim of this work is the comparison of several soil moisture databases, mainly from satellite observations and model reanalysis. Spatial and temporal variability of the data was compared, focusing on the Iberian Peninsula. Statistical tests were also performed, in order to determine whether databases are correlated and to know the limitations of the studied datasets.

Results of the analysis show that all databases present similar trends, even though their absolute values of soil moisture are different. In some cases, there are strong correlations between the datasets. This may be due to the use of land surface models to re-scale and harmonise observed soil moisture data.

**Key words:** soil moisture, satellite data, model, simulation, Iberian Peninsula.

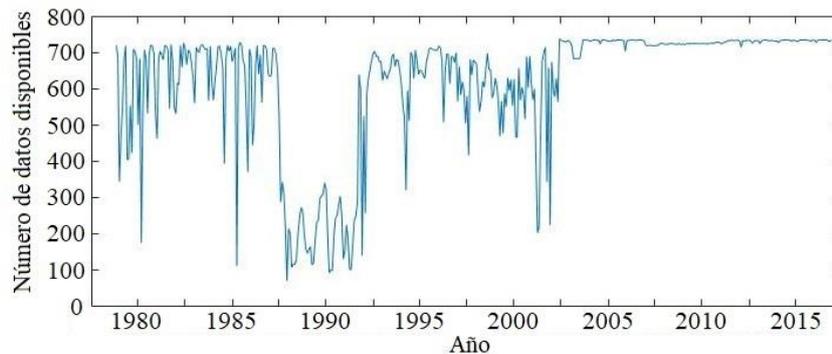
## 1. INTRODUCCIÓN

El estudio del cambio climático ha adquirido gran importancia en los últimos tiempos debido al impacto que puede producir a medio y largo plazo sobre diversos ámbitos naturales y socio-económicos. Se pueden estudiar sus efectos en el futuro mediante el uso de simulaciones realizadas con modelos globales y regionales, como el modelo *Weather Research and Forecasting* (WRF) o el *Consortium for Small-scale Modelling* (COSMO). Es necesario que los datos introducidos en el modelo sean adecuados, ya sean variables climáticas, usadas como condiciones iniciales o de contorno, o datos de tipos y usos del suelo. Por lo tanto, se hace necesario comprobar que los datos disponibles se ajustan a la realidad.

Una de las variables de particular relevancia es la humedad del suelo. Resulta especialmente interesante, por ejemplo, por su influencia en la precipitación, que puede ser positiva o negativa dependiendo del clima (Yang et al., 2018). Además, se ha demostrado que las condiciones iniciales de humedad del suelo en modelos climáticos tienen influencia en los resultados incluso tras el paso de varios años en la simulación (Khodayar et al., 2014). De ahí la importancia de un análisis correcto de la misma a la hora de abordar una simulación climática de futuro. La adecuada representación de la humedad del suelo por los modelos regionales es un aspecto fundamental para comprender y cuantificar procesos como la retroalimentación entre humedad del suelo y temperatura (Seneviratne et al., 2010), siendo una variable relevante en estudios relacionados con la sequía y otros eventos extremos.

Sin embargo, es difícil obtener datos fiables de humedad del suelo. Hay dos fuentes principales: datos de campo obtenidos *in situ* con sensores y datos de satélite. Estos últimos presentan como una de sus principales ventajas el hecho de ser fácilmente accesibles. Ejemplos de este tipo de bases de datos son las bases de datos ESA-CCI (Liu et al., 2012, Liu et al., 2011, Wagner et al., 2012) y GRACE Tellus (Swenson, 2012, Landerer y Swenson, 2012, Swenson y Wahr, 2006). No obstante, presentan una serie de inconvenientes importantes, como que sólo caracterizan la humedad en la capa más superficial del suelo, que su resolución espacial puede ser demasiado amplia (1° en el caso de GRACE Tellus) o que pueden presentarse huecos (como sucede, por ejemplo, en las primeras décadas de ESA-CCI, Figura 1). Estos huecos se

deben a que los satélites no siempre realizan sus mediciones en todo el territorio, por lo que pueden quedar áreas sobre la región de interés sin datos durante días o meses. En el caso de ESA-CCI, el número de datos disponibles depende en gran medida de los satélites utilizados (Su et al., 2016).



*Figura 1: Número de datos disponibles por mes en la base de datos ESA-CCI para la Península Ibérica.*

Por otra parte, se pueden encontrar bases de datos que combinan observaciones con un reanálisis realizado a través de un modelo. Sus principales ventajas son las siguientes: rellenan los huecos presentes en las bases de datos de satélite, pueden utilizarse resoluciones mayores mediante reescalados y permiten obtener un rango de datos mayor en el tiempo. Como contrapartida, se tienen datos cuya bondad depende de la habilidad del modelo empleado y de los datos asimilados en el mismo.

El objetivo del presente estudio es la comparación de distintas bases de datos de humedad del suelo superficial en la Península Ibérica (PI), entre ellas las bases de datos satelitales ESA-CCI y GRACE Tellus. Se compararán con otras bases de datos de reanálisis, actualmente usadas en diversos estudios como son los datos de GLEAM (Martens et al., 2017, Miralles et al., 2011), GLDAS (Rodell et al., 2004) y CPC Soil Moisture (van den Dool, Huang y Fan, 2003). También se compararán con los datos de humedad del suelo obtenidos mediante simulaciones del modelo WRF usando las condiciones de contorno ERA-Interim (García-Valdecasas Ojeda et al., 2017). Con ello, se pretende llevar a cabo una validación de las salidas del modelo WRF para la PI, por lo que en algunas de las comparaciones estos datos se tomarán como referencia.

## **2. DATOS Y MÉTODOS**

La Tabla 1 muestra algunas características de las bases de datos estudiadas, relativas a resolución espacial, cobertura temporal y procedencia de los datos de humedad del suelo superficial.

Para realizar el estudio comparativo de las bases de datos, primero se han ajustado todas a una misma escala espacial y temporal: datos de humedad del suelo superficial media mensual en  $\text{m}^3/\text{m}^3$  para toda la PI con una resolución de  $0.25^\circ \times 0.25^\circ$ . En el caso de la base de datos CPC, se ha realizado un remallado de los datos para adaptarlos a esta resolución. Para ello, se han interpolado los datos por el método del vecino más

cercano, ya que la humedad del suelo varía mucho espacialmente, y las variaciones no son necesariamente lineales. En cualquier caso, este método de interpolación genera unos resultados muy similares a los obtenidos mediante interpolación lineal al obtener el promedio para toda la PI.

Base de datos	Resolución	Periodo de tiempo	Tipo de dato
CPC Soil Moisture	0.5° x 0.5°	1948-2018	Modelo
ESA-CCI	0.25° x 0.25°	1978-2016	Satélite
GLDAS	0.25° x 0.25°	2000-2018	Reanálisis
GLEAM	0.25° x 0.25°	1980-2017	Reanálisis
GRACE Tellus	1° x 1°	2002-2017	Satélite
WRF-ERA	0.088° x 0.088°	1979-2014	Modelo

*Tabla 1: Características de las bases de datos estudiadas.*

La base de datos obtenida de WRF también ha necesitado un remallado, pero en este caso en sentido inverso a los datos CPC, por ser la resolución original más fina que la usada en la comparación. El método de interpolación es el mismo por las razones anteriormente descritas.

En el caso de los datos GRACE Tellus, también se hizo el remallado. Sin embargo, el resultado no fue satisfactorio porque el tamaño de celda original es demasiado grande. Teniendo en cuenta que estos datos están integrados en el reanálisis GLDAS, se decidió prescindir de los datos GRACE en los análisis posteriores.

Además, es necesario destacar otra modificación realizada sobre los datos. Los datos CPC, al provenir de un modelo, abarcan un periodo de tiempo muy amplio. Para evitar que los datos iniciales provocaran una desviación de los valores estadísticos con respecto a las demás bases de datos, se ha recortado el periodo utilizado. Así, la fecha de inicio de los datos CPC considerada es la misma que la de ESA, 1978.

Tras este tratamiento previo, la comparación entre las bases de datos se basó en los siguientes aspectos:

- Correlación cruzada entre las series de anomalías en sus periodos coincidentes para cada par de bases de datos.
- Comparación de medias mensuales para toda la Península Ibérica.
- Test skill score (Perkins et al., 2007).
- Autocorrelación de anomalías mensuales de cada serie para caracterizar la estructura temporal.

### 3. RESULTADOS

La Figura 2 muestra la serie de anomalías mensuales medias de humedad del suelo para toda la PI obtenidas a partir de cada base de datos. Se aprecia una evolución temporal semejante entre ellas. Sin embargo, los valores de humedad del suelo no son los mismos para todas las bases de datos, presentándose diferencias significativas en algunos casos. Esto puede apreciarse más claramente en la Figura 3. Además, en la Tabla 2 se muestran los coeficientes de correlación entre las series anuales de las distintas bases de datos. Los coeficientes presentados se han calculado tomando la serie de promedios anuales en toda la PI. Este análisis se ha repetido separando los

datos en estaciones del año (Tabla 3). Como puede verse, las correlaciones son elevadas entre todas las bases de datos, destacando el alto valor entre los datos de WRF y GLEAM. Por otra parte, las bases de datos ESA y GLDAS son las que muestran una menor correlación. En general, GLDAS es la base de datos que presenta menor correlación con las restantes. Estacionalmente, es en primavera donde las correlaciones entre las distintas series de humedad de suelo presentan en general correlaciones más altas, siendo todas significativas al 99%. Por otra parte, durante verano y otoño, las correlaciones tienden a ser inferiores, particularmente entre la base de datos de satélite ESA y la GLDAS. Para todas las estaciones del año, WRF presenta correlaciones significativas con el resto de las bases de datos, especialmente altas respecto a GLEAM, lo que indica un elevado grado de sincronía de las simulaciones de WRF respecto a los datos de satélite y reanálisis.

El estudio de las autocorrelaciones de cada base de datos arroja los resultados representados en la Figura 4: la autocorrelación disminuye rápidamente, de tal forma que, pasados dos meses, no se aprecian autocorrelaciones significativas. La única excepción es la base de datos CPC, con una mayor memoria, de tal forma que autocorrelaciones significativas pueden encontrarse para 4 o 5 meses de retardo.

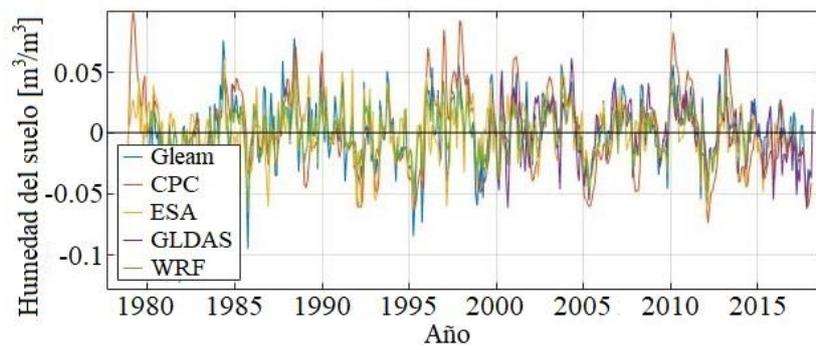


Figura 2: Valores de anomalías de humedad del suelo.

	CPC	ESA	GLDAS	GLEAM
ESA	0.55			
GLDAS	0.52*	0.81		
GLEAM	0.81	0.70	0.85	
WRF	0.89	0.73	0.94	0.91

Tabla 2: Correlación entre las series de promedios anuales de humedad del suelo para las bases de datos. Todas las correlaciones son significativas al nivel de confianza del 99% salvo la marcada con asterisco, que lo es al 95%.

Se ha realizado un análisis del solapamiento entre las distribuciones de probabilidad de valores de humedad del suelo mediante el test *skill score* de Perkins. A diferencia de la correlación temporal, que evalúa en qué medida hay sincronía entre las bases de datos, este análisis permite comprobar el grado de similitud de las funciones de distribución de la humedad del suelo entre las bases de datos. Los resultados se muestran en la Tabla 4.

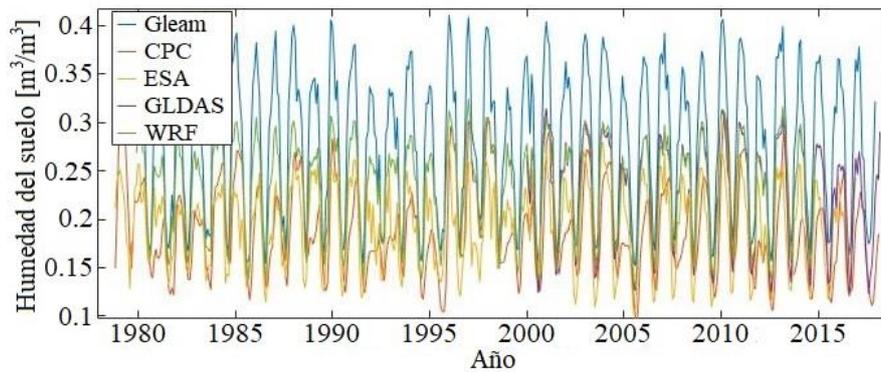


Figura 3: Series temporales completas de humedad del suelo.

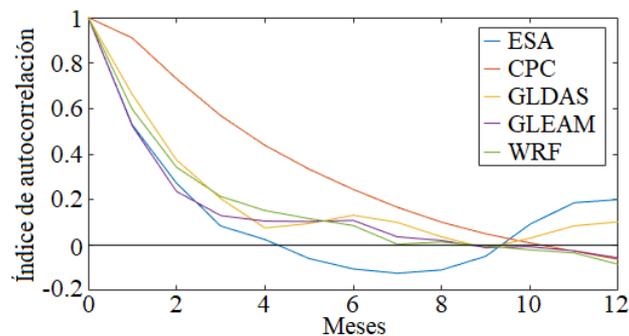


Figura 4: Índices de autocorrelación para distintos meses de retardo.

#### 4. DISCUSIÓN

En general, los coeficientes de correlación obtenidos tienen máximos entre 0.4 y 0.7, llegando en ocasiones a superar 0.9. Es necesario matizar que las series altamente correlacionadas entre sí suelen provenir de datos de reanálisis. Además, el modelo de suelo utilizado en la modelización influye en el comportamiento de la humedad del suelo ante los cambios del resto de variables del modelo. Por ejemplo, el rango de valores de los datos ESA viene corregido por el modelo de suelo Noah, mientras que los datos GLEAM utilizan el modelo de propiedades de suelo IGBP.

Para los datos promediados estacionalmente, se puede observar cómo la correlación disminuye en los meses de verano, cuando hay un incremento de temperatura y descenso en la precipitación, lo que provoca una mayor evaporación y un descenso acusado del contenido de humedad del suelo. Es probable que los datos modelados sean más dependientes de las parametrizaciones durante el verano. Así, por ejemplo, WRF tiende a subestimar las temperaturas durante todo el año, siendo la subestimación más pronunciada durante el verano (García-Valdecasas Ojeda, 2018). Resulta interesante observar cómo un alto coeficiente de correlación no conlleva necesariamente un alto solapamiento de los histogramas según el test *skill score*. De hecho, los valores de este *skill score* son relativamente bajos. Las diferencias pueden producirse por las diferentes correcciones aplicadas a los valores en bruto de humedad del suelo. Estas correcciones tienen en cuenta factores como agua en superficie o

corrientes subterráneas. Además, los datos de satélite parecen corresponder a valores de humedad del suelo superficial más pequeña frente a los 10 cm que considera los datos de reanálisis y modelos. GLDAS es la base de datos que presenta un *skill score* más bajo frente a las demás. Estas diferencias en las funciones de distribución pueden apreciarse en la Figura 5. Destaca que mientras que CPC presenta una distribución unimodal, el resto de bases de datos presenta distribuciones bimodales.

	CPC	ESA	GLDAS	GLEAM
<b>Invierno</b>				
ESA	0.68			
GLDAS	0.57*	0.65		
GLEAM	0.90	0.77	0.57*	
WRF	0.90	0.74	0.47**	0.94
<b>Primavera</b>				
ESA	0.53			
GLDAS	0.76	0.89		
GLEAM	0.80	0.76	0.91	
WRF	0.88	0.71	0.98	0.95
<b>Verano</b>				
ESA	0.54			
GLDAS	0.67	0.37*		
GLEAM	0.78	0.49	0.79	
WRF	0.82	0.52	0.90	0.95
<b>Otoño</b>				
ESA	0.56			
GLDAS	0.79	0.88		
GLEAM	0.67	0.77	0.90	
WRF	0.77	0.77	0.92	0.91

Tabla 3: Correlación entre las series de promedios estacionales de distintas bases de datos. Las correlaciones son significativas al nivel de confianza del 99% salvo las marcadas con \*, que lo son al 95%, y la marcada con \*\*, que no lo es al 95%.

	CPC	ESA	GLDAS	GLEAM
<b>ESA</b>	76.3%			
<b>GLDAS</b>	51.9%	46.4%		
<b>GLEAM</b>	48.1%	38.5%	46.6%	
<b>WRF</b>	61.1%	56.7%	63.0%	51.8%

Tabla 4: Resultados de los skill scores de Perkins.

Respecto a las diferencias espaciales entre las bases de datos, la Figura 6 muestra las diferencias porcentuales entre los promedios anuales de cada base de datos y WRF. Existe un adecuado acuerdo entre los valores de WRF y GLDAS, con un predominio de zonas en las que WRF presenta subestimaciones del orden de 20% respecto a GLDAS. Sin embargo, los valores de ESA tienden a quedar por debajo de los de WRF, excepto en el sur y algunas zonas dispersas del centro y norte, donde las diferencias

no son significativas. Por otro lado, los valores de GLEAM son claramente mayores que los de WRF en la mitad occidental de la PI, mostrando una mayor concordancia de valores en el este. Las diferencias más marcadas aparecen con la base de datos CPC, en las que WRF alcanza sobrestimaciones de más del 100% en el sureste y en zonas del Sistema Central y del valle del Ebro, y subestimaciones del orden del 30% en la mitad norte de Portugal y Galicia y en el este de la cornisa cantábrica.

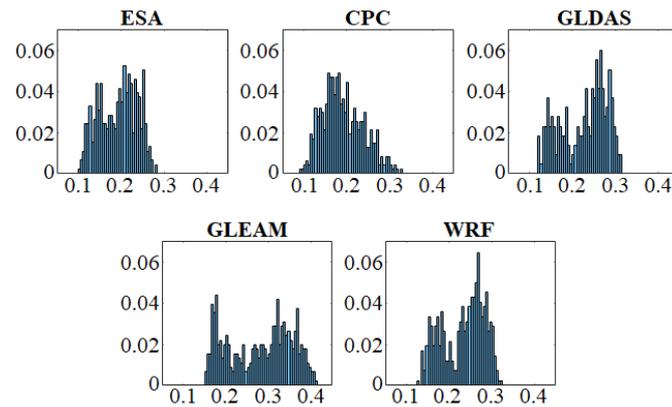


Figura 5: Histogramas de humedad del suelo media mensual en  $m^3/m^3$ .

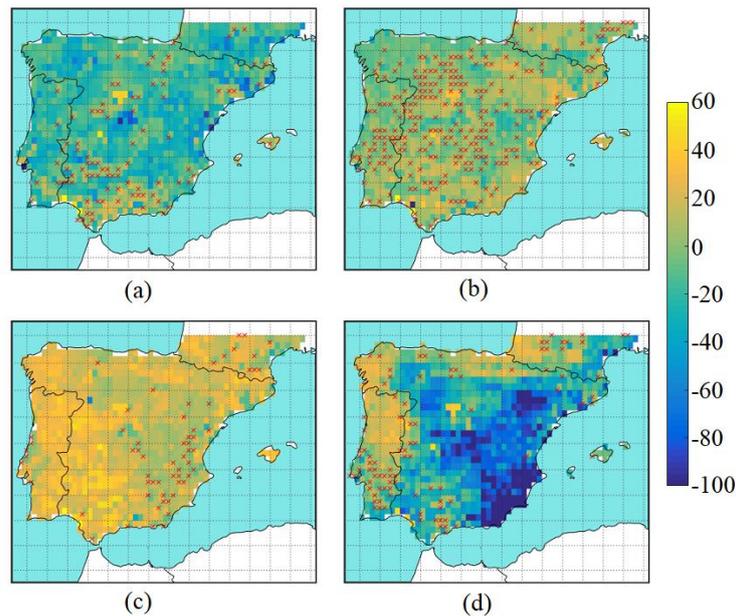


Figura 6: Diferencias en tanto por ciento de las bases de datos (a) ESA, (b) GLDAS, (c) GLEAM, (d) CPC, con respecto a WRF. Se marcan los puntos en los que el test de Wilcoxon no rechaza la hipótesis nula con un intervalo de confianza del 95%.

Respecto a las autocorrelaciones, su rápida disminución se justifica por el hecho de que los de datos de humedad considerados corresponden a la capa más superficial. El efecto de los procesos atmosféricos (precipitación, evapotranspiración) provoca

variaciones marcadas de un mes a otro. Sería interesante estudiar cómo cambia la humedad del suelo en capas más profundas, aunque en este caso, datos en zonas extensas sólo pueden obtenerse mediante modelización.

Con todo lo comentado, se puede afirmar que la diferencia en los valores de humedad del suelo viene dada principalmente por: el tipo de medición, que puede ser de modelo, satélite o campo; el modelo de suelo utilizado, y su habilidad en el ajuste a la región estudiada; así como la escala temporal considerada en el estudio (estacional o anual). En general, el análisis de tendencias de la humedad del suelo para la PI de forma global no se vería muy alterado por la elección de una base de datos en particular, ya que la evolución temporal entre los distintas bases de datos es similar. Sin embargo, otro tipo de usos (por ejemplo, modelización hidrológica) podría verse muy condicionado por la elección. Por ello sería deseable contar con mediciones de campo históricas, que permitieran reducir la incertidumbre al elegir la base de datos. Sin embargo, la ausencia de una base de datos para toda la PI limitaría este tipo de validación a zonas muy pequeñas.

#### **AGRADECIMIENTOS**

Las simulaciones de WRF se realizaron en el supercomputador Alhambra de la Universidad de Granada.

Agradecemos a Dr. Diego G. Miralles su ayuda con los datos de GLEAM.

Este trabajo fue financiado por los proyectos CGL2013-48539-R y CGL2017-89836-R (MINECO-Spain, FEDER).

#### **REFERENCIAS**

- García-Valdecasas Ojeda, M. (2018). Climate-change projections in the Iberian Peninsula: a study on the hydrological impacts. PhD Thesis. Universidad de Granada.
- García-Valdecasas Ojeda, M., Gámiz Fortis, S.R., Castro Díez, Y., Esteban Parra, M.J. (2017) Evaluation of WRF capability to detect dry and wet periods in Spain using drought indices. *J. of Geoph. Res.*, 122. doi: <http://dx.doi.org/10.1002/2016JD025683>
- Khodayar, S., Sehlinger, A., Feldmann, H., Kottmeier, Ch. (2014) Sensitivity of soil moisture initialization for decadal predictions under different regional climatic conditions in Europe. *Int. J. of Climat.*, 35. doi: <http://dx.doi.org/10.1002/joc.4096>
- Landerer, F.W., Swenson, S.C. (2012) Accuracy of scaled GRACE terrestrial water storage estimates. *Water Resources Research*, 48, W04531, 11 PP. doi: <http://dx.doi.org/10.1029/2011WR011453>
- Liu, Y.Y., Dorigo, W.A., Parinussa, R.M., de Jeu, R.A.M., Wagner, W., McCabe, M.F., Evans, J.P., van Dijk, A.I.J.M. (2012) Trend-preserving blending of passive and active microwave soil moisture retrievals. *Remote sensing of environment*, 123, 280-297. doi: <http://dx.doi.org/10.1016/j.rse.2012.03.014>
- Liu, Y.Y., Parinussa, R.M., Dorigo, W.A., de Jeu, R.A.M., Wagner, W., van Dijk, A.I.J.M., McCabe, M.F., Evans, J.P. (2011) Developing an improved soil

- moisture dataset by blending passive and active microwave satellite-based retrievals. *Hydr. and Earth System Sciences*, 15, 425-436. doi: <http://dx.doi.org/10.5194/hess-15-425-2011>
- Martens, B., Miralles, D.G., Lievens, H., van der Schalie, R., de Jeu, R., Fernández-Prieto, D., Beck, H.E., Dorigo, W.A., Verhoest, N.E.C. (2017) GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, 10, 1903–1925. doi: <http://dx.doi.org/10.5194/gmd-10-1903-2017>
- Miralles, D.G., Holmes, T., de Jeu, R., Gash, J.H., Meesters, A.G.C.A., Dolman, A.J. (2011) Global land-surface evaporation estimated from satellite-based observations, *Hydr. & Earth Sys. Sc.* 15, doi: <http://dx.doi.org/10.5194/hess-15-453-2011>
- Perkins, S.E., Pitman, A.J., Holbrook, N.J., McAneney, J. (2007) Evaluation of the AR4 Climate Models' Simulated Daily Max. Temp., Min. Temp., and Precip. over Australia Using PDFs. *J. of Clim.*, 20. doi: <http://dx.doi.org/10.1175/JCLI4253.1>
- Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.J., ... Toll D. (2004) The Global Land Data Assimilation System. *American Meteor. Society*, 85 (3), 381-394. doi: <http://dx.doi.org/10.1175/BAMS-85-3-381>
- Seneviratne, S., Corti, T., Davin, E., Hirschi, M., Jaeger, E., Lehner, I., Orlowsky, B., Teuling, A. (2010) Investigating soil moisture-climate interactions in a changing climate. *Earth-Sci. Rev.* 99. doi: <https://dx.doi.org/10.1016/j.earscirev.2010.02.004>
- Su, C.H., Ryu, D., Dorigo, W., Zwieback, S., Gruber, A., Albergel, C., Reichle, R.H., Wagner, W. (2016) Homogeneity of a global multisatellite soil moisture climate data record. *Geoph. res. letters*, 43. doi: <https://dx.doi.org/10.1002/2016GL070458>
- Swenson, S. C., Wahr, J. (2006) Post-processing removal of correlated errors in GRACE data. *Geoph. Res. Letters*, 33. doi: <http://dx.doi.org/10.1029/2005GL025285>
- Swenson, S.C. (2012) GRACE monthly land water mass grids NETCDF RELEASE 5.0. Ver. 5.0. PO.DAAC, CA, USA. doi: <http://dx.doi.org/10.5067/TELND-NC005>
- Van den Dool, H., Huang, J., Fan, Y. (2003) Performance and Analysis of the constructed analogue method applied to US soil moisture applied over 1981-2001. *J. of Geoph. Research*: 108 (D16), 1-16. doi: <http://dx.doi.org/10.1029/2002JD003114>
- Wagner, W., Dorigo, W., de Jeu, R., Fernández, D., Benveniste, J., Haas, E., Ertl, M. (2012) Fusion of active and passive microwave observations to create an Essential Climate Variable data record on soil moisture. *ISPRS Annals of the photogr. remote sensing and spat. inf. sciences*, I. Recuperado de <https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/I-7/315/2012/isprsannals-I-7-315-2012.pdf>

Yang, L., Sun, G., Zhi, L., Zhao, J. (2018) Negative soil moisture-precipitation feedback in dry and wet regions. *Scie. rep.* 8. doi: <http://dx.doi.org/10.1038/s41598-018-22394-7>