

García Ramajo, J.; Jiménez Romanillos, J.; Martínez García, F.; Fragoso Campón, L. Del grado de saturación del suelo al caudal: integración de EU-SoilHydroGrids y ERA5-Land en modelos hidrológicos

# Del grado de saturación del suelo al caudal: integración de EU-SoilHydroGrids y ERA5-Land en modelos hidrológicos

García Ramajo, Jorge Juan <sup>1</sup> Jiménez Romanillos, José <sup>2</sup> Martínez García, Francisco Manuel <sup>1</sup> Fragoso Campón, Laura <sup>1</sup>

<sup>1</sup> Universidad de Extremadura, España

<sup>2</sup> Junta de Extremadura

ORCID: García Ramajo 0009-0000-7296-3082 Jiménez Romanillos 0009-0009-0612-4390 Martínez García 0000-0002-0862-9933 Fragoso Campón 0000-0003-0397-6247

Correspondencia: [jjgarcia@unex.es](mailto:jjgarcia@unex.es) [jose.jimenez@juntaex.es](mailto:jose.jimenez@juntaex.es) [fmmgarcia@unex.es](mailto:fmmgarcia@unex.es) [laurafragoso@unex.es](mailto:laurafragoso@unex.es)

## RESUMEN

Las inundaciones constituyen uno de los desastres naturales más frecuentes y con mayor impacto socioeconómico, relacionadas no solo con eventos extremos de precipitación, sino también con las características hidrológicas del territorio. Este estudio se centra en la cuenca del río Gévora, en el entorno de Badajoz, que en los últimos años ha experimentado varios episodios de inundación que han requerido la activación de medidas de emergencia. Estos eventos ponen de manifiesto la elevada vulnerabilidad de la zona frente a fenómenos hidrológicos extremos y la necesidad de mejorar los sistemas de monitorización, modelización y predicción de avenidas. En este contexto, se evalúa la utilidad de integrar información meteorológica de alta resolución procedente de ERA5-Land con propiedades hidráulicas del suelo derivadas de la base de datos EU-SoilHydroGrids para su aplicación en la modelización hidrológica mediante aprendizaje automático basado en Random Forest. Los resultados muestran que el modelo que combina ambas fuentes de información presenta el mejor rendimiento, reproduciendo adecuadamente la dinámica temporal del caudal y la magnitud de los picos de avenida, incluso durante eventos extremos. El análisis de la importancia de los predictores destaca el papel del grado de saturación del suelo, especialmente en relación con la capacidad de campo poniendo de relieve el papel determinante del estado de saturación del suelo en la generación de escorrentía.

**Palabras clave:** inundaciones; propiedades hidráulicas del suelo; predicción de caudales


Fecha de recepción: 19 febrero 2026 · Fecha de aceptación: 19 febrero 2026

# Del grado de saturación del suelo al caudal: integración de EU-SoilHydroGrids y ERA5-Land en modelos hidrológicos


García Ramajo, Jorge Juan <sup>(1)</sup>, Jiménez Romanillos, José <sup>(2)</sup>, Martínez García, Francisco Manuel <sup>(1)</sup>, Fragoso Campón, Laura <sup>(1)</sup>

<sup>(1)</sup> Universidad de Extremadura, España.

 0009-0000-7296-3082, [jjgarcia@unex.es](mailto:jjgarcia@unex.es);  0000-0002-0862-9933, [fmmgarcia@unex.es](mailto:fmmgarcia@unex.es)

 0000-0003-0397-6247, [laurafragoso@unex.es](mailto:laurafragoso@unex.es).

<sup>(2)</sup> Junta de Extremadura

 0009-0009-0612-4390, [jose.jimenez@juntaex.es](mailto:jose.jimenez@juntaex.es).

**Resumen:** Las inundaciones constituyen uno de los desastres naturales más frecuentes y con mayor impacto socioeconómico, relacionadas no solo con eventos extremos de precipitación, sino también con las características hidrológicas del territorio. Este estudio se centra en la cuenca del río Gévora, en el entorno de Badajoz, que en los últimos años ha experimentado varios episodios de inundación que han requerido la activación de medidas de emergencia. Estos eventos ponen de manifiesto la elevada vulnerabilidad de la zona frente a fenómenos hidrológicos extremos y la necesidad de mejorar los sistemas de monitorización, modelización y predicción de avenidas. En este contexto, se evalúa la utilidad de integrar información meteorológica de alta resolución procedente de ERA5-Land con propiedades hidráulicas del suelo derivadas de la base de datos EU-SoilHydroGrids para su aplicación en la modelización hidrológica mediante aprendizaje automático basado en Random Forest. Los resultados muestran que el modelo que combina ambas fuentes de información presenta el mejor rendimiento, reproduciendo adecuadamente la dinámica temporal del caudal y la magnitud de los picos de avenida, incluso durante eventos extremos. El análisis de la importancia de los predictores destaca el papel del grado de saturación del suelo, especialmente en relación con la capacidad de campo poniendo de relieve el papel determinante del estado de saturación del suelo en la generación de escorrentía.

**Palabras clave:** inundaciones; propiedades hidráulicas del suelo; predicción de caudales.

## ***From Soil Saturation to Streamflow: Integration EU-SoilHydroGrids and ERA5-Land in Hydrological Models***

**Abstract:** *Floods are among the most frequent natural disasters and have some of the greatest socioeconomic impacts. They are associated not only with extreme precipitation events but also with the hydrological characteristics of a territory. This study focuses on the Gévora River basin, in the Badajoz area, which in recent years has experienced several flood episodes that have required the activation of emergency measures. These events highlight the high vulnerability of the area to extreme hydrological phenomena and the need to improve flood monitoring, modelling, and forecasting systems. In this context, the study evaluates the usefulness of integrating high-resolution meteorological information from ERA5-Land with soil hydraulic properties derived from the EU-SoilHydroGrids database for application in hydrological modelling using a Random Forest machine learning approach. The results show that the model combining both sources of information provides the best performance, accurately reproducing streamflow dynamics and flood peak magnitudes, even during extreme events. Predictor importance analysis highlights the role of soil saturation, particularly in relation to field capacity, underscoring the decisive role of soil saturation status in runoff generation.*

**Keywords:** *Flood modelling; soil hydraulic properties; streamflow prediction.*

## **1. INTRODUCCIÓN**

Las inundaciones son unos de los desastres naturales más frecuentes y con gran impacto socioeconómico, normalmente consecuencia de eventos extremos de precipitación. La respuesta hidrológica de un territorio

depende de varios factores que pueden ser monitorizados mediante técnicas de teledetección, consolidándose como fuente fundamental de información básica para el análisis y la modelización hidrológica. Estudios recientes confirman que el enfoque

de aprendizaje automático mejora la estimación de caudales respecto a los modelos conceptuales tradicionales (Mimeau *et al.*, 2024), pero es fundamental construir una base de datos completa de predictores que abarquen todas las variables que controlan la respuesta hidrológica.

El Copernicus Climate Change Service (C3S), como componente del programa europeo de observación de la Tierra Copernicus, proporciona el conjunto de datos de reanálisis climático ERA5-Land, que ofrece variables meteorológicas con alta resolución espacial y temporal. Por otro lado, la base de datos EU-SoilHydroGrids (EU-SHG) (Tóth *et al.*, 2017), facilitada por el European Soil Data Centre (ESDAC) (Panagos *et al.*, 2022) y desarrollada por el Joint Research Centre (JRC) de la Comisión Europea, suministra información edáfica armonizada a escala continental, con especial énfasis en propiedades hidráulicas del suelo.

El presente trabajo pretende fusionar los datos de la EU-SHG y ERA5-Land, para obtener nuevos predictores relacionados con el antecedente de humedad del suelo, y evaluar la eficacia de incorporar estos parámetros en la modelización hidrológica mediante aprendizaje automático.

## 2. MATERIAL Y MÉTODOS

### 2.1. Area de estudio

El caso de estudio corresponde a la cuenca del río Gévora (Badajoz), tomando como punto de cierre la estación de aforo en río de Valdeobótoa (CR 2-45). Los caudales registrados se han obtenido del SAIH de Confederación Hidrográfica del Guadiana y se ha procesado la serie temporal horaria desde 2018 a 2025. Para generar los predictores, se ha simulado la cuenca agregada en la estación de aforo, calculando las estadísticas zonales de: valor medio, desviación estándar y los cuantiles 5, 10, 25, 50, 75, 90 y 95.

### 2.2. Conjunto ERA5-Land

Los datos ERA5-Land suministra datos horarios con una resolución espacial de malla de 9 km (0.1° x 0.1°). Dentro de los productos disponibles, para la caracterización hidrológica del territorio, se han utilizado los conjuntos de variables de temperatura, de radiación-presión e hidrológicas (Tabla 1). Además, se han generado variables acumuladas de precipitación y evaporación, a 24h, 72h y 120h anteriores.

**Tabla 1.** Relación de variables utilizadas del ERA5-Land

Variable ERA5-Land	Notación
2 metre dewpoint temperatura	2d
2 metre temperatura	2t
Skin temperatura	skt
Soil temperature levels*	stl
Surface net short-wave (solar) radiation	ssr
Surface net long-wave (thermal)	str
Surface pressure	sp
Snowmelt	smlt
Volumetric soil water levels*	swvl
Skin reservoir content	src

Variable ERA5-Land	Notación
Evaporation	e
Potential evaporation	pev
Surface runoff	sro
Sub-surface runoff	ssro
Total precipitation	tp

\*Las variables "Levels" son datos multicapa ofreciendo un valor para cada en los siguientes estratos. Capa 1°: 0-7; 2°: 7-28; 3°: 28-100; 4°: 100-289 cm.

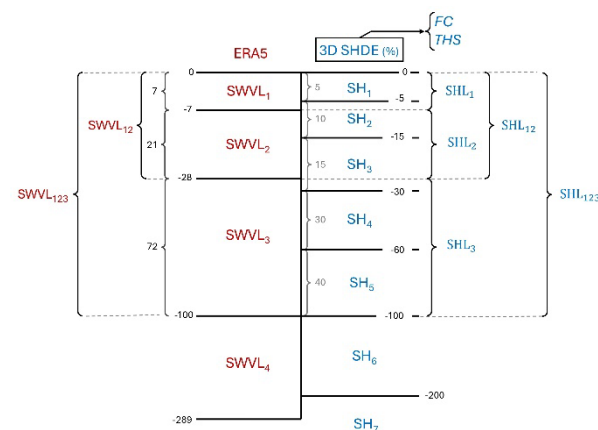
### 2.3. Conjunto EU-SoilHydroGrids (EU-SHG)

De todos los conjuntos de datos disponibles en EU-SHG, se han seleccionado, el contenido de agua de saturación (THS) y la capacidad de campo (FC), ya que ambos parámetros son fundamentales para caracterizar la capacidad de almacenamiento hídrico del suelo y la generación de escorrentía. La información se facilita a una resolución espacial de 250 m, para las que se suministran valores en siete profundidades: 0, 5, 15, 30, 60, 100 y 200 cm.

### 2.4. Armonización de las bases de datos

La información de las SWV, y EU-SHG se refieren a estratos de distintos espesores por lo que ha sido necesario llevar a cabo una armonización de ambas bases de datos (Fig. 1). Se ha efectuado una normalización por estratos de las 5 primeras capas del grupo EU-SHG (SH<sub>ij</sub>), respecto a las SWVL de ERA5-Land (SWVL<sub>i</sub>), generando las nuevas capas armonizadas SHL<sub>ij</sub> (soil hydro layer) según la media ponderada de los valores medidos en cada estrato.

Además, se han agregado los principales estratos en función de la profundidad, generando dos nuevas capas: SWVL<sub>12</sub> y SWVL<sub>123</sub>; junto con SHL<sub>12</sub> y SHL<sub>123</sub>, respectivamente, calculando de nuevo el promedio ponderado por el espesor.



**Figura 1.** Esquema de profundidad de capas de suelo de los grupos SWVL (ERA5) y EU-SHG (cotas en cm).

Una vez generadas todas las capas armonizadas se han obtenido los nuevos predictores relativos a la humedad del suelo, denominados SM<sub>ij</sub> (soil moisture) obteniendo la diferencia entre el valor registrado en las capas SWVL<sub>i</sub> y el valor de referencia de SHL<sub>ij</sub>, valores positivos indican exceso de humedad respecto el valor de referencia y los valores negativos déficit.

$$SM_{ji} = SWVL_i - SHL_{ji}$$

donde  $i$  se refiere al estrato: 1, 2, 3, 12, 123 y  $j$  se refiere al parámetro FC o THS.

## 2.5. Algoritmo de aprendizaje automático

Para el entrenamiento de los modelos se utilizó Random Forest (RF) de la librería *Scikit-learn* (Pedregosa *et al.*, 2011). RF permite interpretar fácilmente la importancia de los predictores y previamente ha sido validado para la estimación de caudales horarios (Cisterna-García *et al.*, 2025). La serie temporal se dividió en un conjunto de entrenamiento (80 %) y otro de test (20 %), sobre el cual se ha evaluado la eficiencia de la simulación. Para que el modelo pueda capturar el retardo en la respuesta hidrológica de la cuenca, para cada dato de medición de Q, se calculan los datos con lags horarios hasta 24h antes de la medición de Q en la estación de aforo.

## 2.6. Modelos

Se han planteado distintas combinaciones de los grupos de predictores (Tabla 2) para alizar la eficacia de incorporar estos parámetros en la modelización hidrológica.

**Tabla 2.** Modelos, grupo de predictores y número (n) de predictores en cada modelo.

Modelo	ERA5-L	SWWL	EU-SHG	n
QRF_M1	X			3975
QRF_M2	X	X		5325
QRF_M3	X	X	X	7525
QRF_M4	X		X	6225

## 2.7. Medida de la Eficiencia

La eficiencia de los modelos se ha evaluado mediante un conjunto de indicadores estadísticos ampliamente utilizados en la modelación hidrológica (Moriasi *et al.*, 2015). En particular, se han utilizado las métricas KGE (*Kling-Gupta Efficiency*) y NSE (*Nash-Sutcliffe Efficiency*), MAE (*Error Absoluto Medio*) y RMSE (*Raíz del Error Cuadrático Medio*), R2 y Pbias (sesgo porcentual).

## 3. RESULTADOS Y DISCUSIÓN

### 3.1. Eficiencia de los modelos

La Tabla 3 muestra los valores de eficiencia obtenidos para cada modelo evaluado.

**Tabla 3.** Evaluación de la eficiencia de los caudales simulados

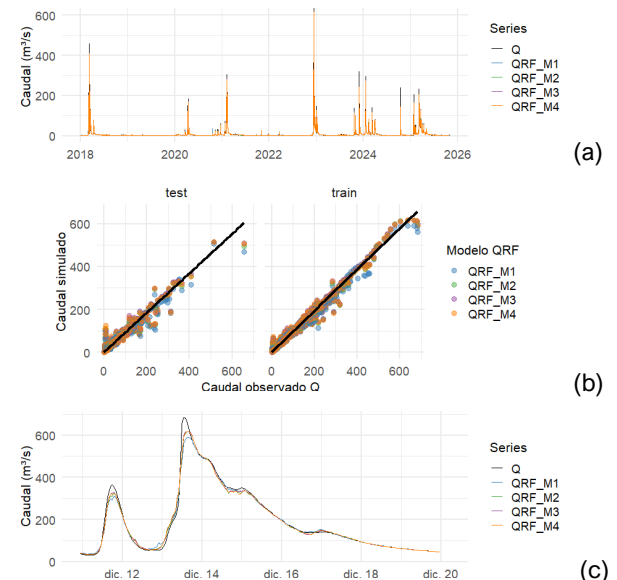
Modelo	NSE	KGE	RMSE	MAE	PBIAS
QRF_M1	0.960	0.908	2.048	0.303	0.10
QRF_M2	0.969	0.938	1.800	0.231	-0.10
<b>QRF_M3</b>	<b>0.970</b>	<b>0.943</b>	<b>1.781</b>	<b>0.227</b>	<b>-0.10</b>
QRF_M4	0.967	0.939	1.860	0.238	0.00

En general, todos los modelos presentan una eficiencia muy alta, con valores elevados de NSE y KGE, así como bajos errores (RMSE y MAE) y sesgos porcentuales

cercanos a cero. Entre ellos, el modelo QRF\_M3 es el que mejor rendimiento global tiene, seguido muy de cerca por QRF\_M2 y QRF\_M4, que también muestran un ajuste excelente. Por su parte, QRF\_M1, presenta valores ligeramente inferiores en comparación con los demás modelos.

De acuerdo con los criterios de evaluación propuestos por Moriasi *et al.* (2015), la eficiencia de los modelos es excelente, destacando el modelo QRF\_M3 como el de mejor rendimiento global, alcanzando eficiencias ligeramente superiores a los obtenidos en Cisterna-García *et al.* (2025) que obtuvieron valores del NSE de 0.921 y superiores a los resultados obtenidos en Mimeau *et al.* (2024) que obtuvieron KGE de 0.76 (ambos trabajos simulaban caudales horarios con RF)

La Figura 2 muestra la serie temporal completa (Fig. 2a) y el detalle del evento de diciembre del 2022 (Fig. 2c) correspondiente al paso de la borrasca Efraín. Las métricas de eficiencia y la capacidad de simular los picos de avenidas (Fig. 2b) son superiores a otros trabajos similares como Schoppa *et al.* (2020), si bien en este caso la escala temporal era diaria.



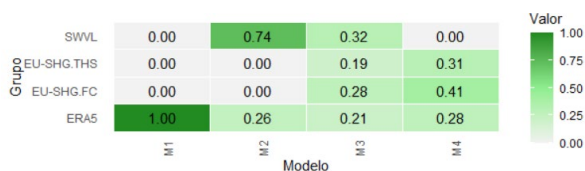
**Figura 2.** Hidrograma de caudal observado (Q) y caudales simulados (QRFi) (a) serie temporal 2018 a 2025; (b) diagrama de dispersión (c) evento de diciembre del 2022

### 3.2. Importancia de los predictores

La Figura 3 muestra la importancia de los predictores utilizados en cada modelo, agrupados por categorías. La simulación RF1, solo cuenta con predictores del grupo ERA5 por lo que es la única fuente de información predictiva. En el caso del RF2, en el que se incluyen además los predictores derivados del grado de saturación, se observa como los predictores del grupo SWWL presentan una importancia (0.74) muy superior al resto de los predictores del ERA5, que pasa a tener un grado de contribución secundario (0.26). El modelo RF3, los nuevos predictores SHG, adquieren en conjunto mayor representatividad alcanzado 0.47 frente a 0.32 de las SWV, destacando que son los predictores del grupo EU-SHG.FC los más importantes (0.28). En el último de

los modelos, los predictores, de los grupos EU-SHG, especialmente EU-SHG.FC (0.41) y EU-SHG.THS (0.31), predominan sobre los del ERA5 (0.28).

Los resultados muestran que los grupos vinculados a las condiciones hidrológicas del suelo y a su estado relativo de saturación (especialmente los asociados a EU-SHG y SWVL, según el modelo) concentran una parte sustancial de la importancia predictiva. En particular, los predictores asociados al exceso o déficit del grado de saturación del suelo respecto a la capacidad de campo (FC) son, en conjunto, los de mayor relevancia. Entre estos, las variables de las capas superficiales del terreno hasta los 21 cm de profundidad (smfc2 y smfc12) destacan por su contribución, no solo en el modelo EU-SHG (RF4) sino también en el modelo combinado con variables SWVL (RF3) donde presentan una importancia aún mayor. Este resultado es coherente con la base física del proceso hidrológico, donde el grado de saturación controla la generación de escorrentía.



**Figura 3.** Importancia relativa de los predictores agrupadas en cada categoría.

#### 4. CONCLUSIONES

La integración de variables meteorológicas de ERA5-Land y propiedades hidráulicas del suelo procedentes de EU-SoilHydroGrids permite mejorar la estimación hidrológica mediante modelos de aprendizaje automático, como Random Forest. Todos los modelos evaluados muestran un rendimiento excelente, resultando que el modelo que fusiona ambas bases de datos es el de mejor comportamiento global, con capacidad para reproducir adecuadamente la dinámica del caudal y los picos de avenida.

El análisis de importancia de predictores evidencia que las variables relacionadas la humedad y el grado de saturación del suelo, especialmente respecto de la capacidad de campo (FC), son los principales factores explicativos del modelo, en concordancia con el papel físico del estado de saturación en la generación de escorrentía.

En conjunto, el estudio demuestra que la incorporación explícita de propiedades hidráulicas del suelo como predictores mejora la representación de las condiciones de humedad y mejora la capacidad de la simulación hidrológica basada en aprendizaje automático con Random Forest.

#### 5. AGRADECIMIENTOS



Cofinanciado por la Unión Europea a través del Programa Interreg VI-A España-Portugal (POCTEP) 2021-2027.

“Redes de alertas tempranas, para la teledetección de riesgos derivados del cambio climático, por satélites de observación de la tierra para respuesta de protección civil (RAT\_EOS\_PC)”.

#### 6. REFERENCIAS

- Cisterna-García, A., González-Vidal, A., Martínez-Ibarra, A., Ye, Y., Guillén-Teruel, A., Bernal-Escobedo, L., & Skarmeta, A. F. (2025). Artificial intelligence for streamflow prediction in river basins: A use case in Mar Menor. *Scientific Reports*, 15, 19481. <https://doi.org/10.1038/s41598-025-04524-0>
- Mimeau, L., Künné, A., Branger, F., Kralisch, S., Devers, A., & Vidal, J.-P. (2024). Flow intermittence prediction using a hybrid hydrological modelling approach: Influence of observed intermittence data on the training of a random forest model. *Hydrology and Earth System Sciences*, 28, 851–871. <https://doi.org/10.5194/hess-28-851-2024>
- Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. *Transactions of the ASABE*, 58(6), 1763-1785. <https://doi.org/10.13031/trans.58.10715>
- Muñoz-Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., ... & Thépaut, J. N. (2021). ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth system science data*, 13(9), 4349-4383. <https://doi.org/10.5194/essd-13-4349-2021>
- Panagos, P., Van Liedekerke, M., Borrelli, P., Köninger, J., Ballabio, C., Orgiazzi, A., Lugato, E., Liakos, L., Hervás, J., Jones, A., & Montanarella, L. (2022). *European Soil Data Centre 2.0: Soil data and knowledge in support of EU policies*. *European Journal of Soil Science*, 73(6), e13315. <https://doi.org/10.1111/ejss.13315>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. y Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830
- Schoppa, L., Disse, M., & Bachmair, S. (2020). Evaluating the performance of random forest for large-scale flood discharge simulation. *Journal of Hydrology*, 590, 125531. <https://doi.org/10.1016/j.jhydrol.2020.125531>
- Tóth, B., Weynants, M., Pásztor, L., & Hengl, T. (2017). 3D soil hydraulic database of Europe at 250 m resolution. *Hydrological Processes*, 31(14), 2497–2666. <https://doi.org/10.1002/hyp.11203>
- Tripathy, K. P., & Mishra, A. K. (2024). Deep learning in hydrology and water resources disciplines: Concepts, methods, applications, and research directions. *Journal of Hydrology*, 628, 130458. <https://doi.org/10.1016/j.jhydrol.2023.130458>