Cuadernos de Investigación Geográfica *Geographical Research Letters*	2026	Nº 52 (1)	pp. 3-21	EISSN 1697-9540

http://doi.org/10.18172/cig.7283

SPATIAL MODELLING OF SOIL PROPERTIES WITH RANDOM FOREST: APPLICATION TO "LOS ALCORNOCALES" NATURAL PARK (ANDALUSIA, SOUTHERN SPAIN)

JOSÉ ANTONIO ORTEGA-PINO[1], JOSÉ ANTONIO SILLERO-MEDINA*, PALOMA HUESO-GONZÁLEZ, JOSÉ DAMIÁN RUIZ-SINOGA

Instituto de Hábitat, Territorio y Digitalización, Departamento de Geografía, Universidad de Málaga, España.

ABSTRACT. Soil is a key component for sustaining life and maintaining ecosystem balance, as it provides nutrients and participates in essential processes such as the water cycle and carbon sequestration. Therefore, assessing its condition and promoting its conservation have become priorities in the context of climate change. One of the main challenges in soil management is its spatial representation, as the heterogeneity of soil properties and environmental variability hinder an accurate characterization of the territory. This challenge is particularly relevant because precise spatial representation enables the identification of distribution patterns, vulnerable areas, and zones of high environmental quality, thereby supporting informed decision-making. In this context, the main objective of this study is to characterize the soils of Los Alcornocales Natural Park (Andalusia, southern Spain) from a physical, organic, and hydrological perspective using advanced spatial modelling techniques. To this end, the Random Forest (RF) algorithm, based on decision trees, was employed. This algorithm allows the integration of multiple environmental variables and enhances the accuracy of the results. The modelling approach was based on the analysis of 461 soil samples (0–10 cm depth), and validation was conducted using a dual approach: external validation with 10% of the data as an independent dataset, and the internal Out-of-Bag (OOB) validation inherent to the RF algorithm. Error and accuracy metrics demonstrated the robustness of the RF model for predictive mapping in a natural park characterized by high eco-geomorphological heterogeneity (CC, $R^2 = 0.937$; Factor K, R^2 0.935; COS, $R^2 = 0.918$). The results indicated that biotic factors, derived from spectral data, play a determining role in the spatial distribution of the analysed soil indicators (organic carbon, water retention capacity, and soil erodibility). This information constitutes a key tool for guiding land-use planning and sustainable management, with high potential for application in other similar Mediterranean contexts.

Modelización espacial de propiedades edáficas con Random Forest: aplicación al Parque Natural los Alcornocales (Andalucía, sur de España)

RESUMEN. El suelo es un elemento clave para sostener la vida y el equilibrio de los ecosistemas, pues provee nutrientes y participa en procesos como el ciclo del agua y la captura de carbono. Por ello, analizar su estado y promover su conservación es una prioridad ante el cambio climático. Uno de los principales retos en su gestión es la representación espacial, dado que la heterogeneidad de las propiedades edáficas y la variabilidad ambiental dificultan una caracterización precisa del territorio. Este desafío es especialmente relevante porque una representación precisa permite identificar patrones de distribución, áreas vulnerables y zonas de alta calidad ambiental, facilitando la toma de decisiones informadas. En este contexto, el objetivo principal de este estudio es caracterizar el suelo del Parque Natural Los Alcornocales (Andalucía, sur de España) desde una perspectiva física, orgánica e hídrica mediante técnicas avanzadas de modelización espacial. Para ello, se utilizó el algoritmo Random Forest (RF), basado en árboles de decisión. Este algoritmo permite integrar múltiples variables ambientales e incrementar la exactitud de los resultados. La modelización se fundamentó en el análisis de 461 muestras de suelo (0-10 cm de profundidad) y la validación se efectuó mediante un enfoque dual: una validación externa con un 10 % de los datos como conjunto independiente y la validación interna Out-of-Bag (OOB) del algoritmo RF. Los indicadores de error y precisión evidenciaron la solidez del modelo RF para la generación de cartografía predictiva en un parque natural con elevada heterogeneidad eco-geomorfológica (CC, R^2 = 0,937; Factor K, R^2 0,935; COS, R^2 = 0,918). Los resultados evidenciaron que los factores bióticos, derivados de datos espectrales, tienen un papel determinante en la distribución espacial de los indicadores edáficos analizados (carbono orgánico, capacidad de retención hídrica y erodabilidad del suelo). Esta información constituye una herramienta clave para orientar la planificación y gestión sostenible del territorio, presentando un elevado potencial de aplicabilidad en otros contextos mediterráneos similares.

Key words: Soil, Digital soil mapping, Modelling, Random Forest, Mediterranean ecosystem.

Palabras clave: Suelo, Cartografía digital de suelos, Modelización, Random Forest, Ecosistemas mediterráneos.

Received: 27 February 2026

Accepted: 10 June 2026

***Corresponding author:** José Antonio Sillero-Medina, Instituto de Hábitat, Territorio y Digitalización. Departamento de Geografía. Universidad de Málaga. España. Email: jasillero@uma.es

1. Introduction

Soil is an essential resource for life on Earth, not only for its intrinsic value but also for its crucial role in maintaining ecosystem health and supporting plants, animals, and human societies. Its importance extends across multiple dimensions of sustainability, including food, water, energy, and climate security (McBratney *et al.*, 2014).

Progressive soil degradation represents a major global threat, driven by both natural processes and intensive human activities such as deforestation, intensive agriculture, urbanization, and pollution (Dragović and Vulević, 2021; Durán-Sandoval *et al.*, 2023). These factors alter soil structure, composition, and fertility, compromising biodiversity, environmental quality, the balance of natural systems, and the integrity of protected natural areas (ENP), whose conservation largely depends on the health and functionality of the soils that sustain them (Mesele *et al.*, 2025).

Understanding soil quality is fundamental to addressing these challenges. According to Bünemann *et al.* (2018), soil quality is defined as the capacity of a soil to perform essential ecosystem functions with respect to its use, promoting biological productivity, maintaining environmental balance, and ensuring the well-being of plants and animals. Soil quality is closely associated with soil organic matter content, particularly soil organic carbon (SOC), which enhances soil fertility, structure, and water-holding capacity while acting as a carbon sink (Blanco Bernardeau *et al.*, 2014). SOC plays a key role in climate change mitigation by reducing atmospheric CO_2 concentrations. However, its degradation leads to carbon losses from the soil, thereby exacerbating global warming (Peri *et al.*, 2024).

In Mediterranean regions, water erosion is the primary driver of soil degradation. In areas such as Andalusia, pronounced rainfall gradients create ecological contrasts that determine soil vulnerability. Although Mediterranean regions are characterized by relatively low annual precipitation (approximately 500 mm), they also exhibit highly irregular rainfall patterns, with prolonged dry periods interrupted by short, intense, and torrential rainfall events. These conditions increase susceptibility to soil degradation processes, particularly in areas receiving less than 500 mm of annual rainfall, which are especially vulnerable to desertification. This phenomenon is further intensified by human activities such as deforestation and intensive agriculture (Ruiz-Sinoga and Romero-Díaz, 2010). Assessing soil susceptibility to erosion is therefore essential for sustainable land management. Tools such as the Revised Universal Soil Loss Equation (RUSLE) integrate factors including slope, vegetation cover, and soil type (Morgan and Nearing, 2011). Water erosion caused by intense rainfall destabilizes soils, reduces fertility, and releases CO_2 into the atmosphere (Gaitán *et al.*, 2017). Consequently, preserving soil structure and water-holding capacity is critical for maintaining ecosystem functioning and mitigating environmental impacts (Sillero-Medina, 2022).

Soil texture and structure determine the soil's ability to retain water, a critical factor for vegetation development and ecosystem functioning (Panagea *et al.*, 2021). Water-holding capacity is considered a key indicator of soil environmental quality because it reflects interactions among multiple soil properties. Field capacity (FC) represents the amount of water retained after gravitational drainage; the permanent wilting point (WP) defines the soil moisture level at which plants can no longer extract sufficient water, resulting in irreversible wilting (Tamara Polo *et al.*, 2016); and available water content (AWC) corresponds to the difference between FC and WP, representing the amount of water available for plant uptake. These parameters are essential for evaluating soil functioning under water-stress conditions, characteristic of Mediterranean climates (Pérez-De Los Reyes *et al.*, 2018; Sillero-Medina *et al.*, 2021).

Within this context, soil mapping constitutes a strategic tool for assessing the quality and health of natural areas, particularly ENP. Soil maps enable the identification of areas of high ecological value, zones vulnerable to degradation, and spatial patterns in soil properties, thereby facilitating sustainable land-use planning and management (Varvaris *et al.*, 2025). Digital Soil Mapping (DSM), which integrates geographic information systems (GIS), remote sensing, and environmental sensors, has significantly advanced the spatial representation of soil properties (Valera, 2023). This approach is based on the quantitative modelling of relationships between soil observations and environmental covariates to generate reproducible spatial predictions, often within the SCORPAN framework (soil, climate, organisms, relief, parent material, age, and spatial position), although some studies employ only subsets of these variables or alternative approaches depending on data availability and the soil properties being predicted (McBratney *et al.*, 2003; Minasny and McBratney, 2016; Wadoux *et al.*, 2020). Traditional approaches, such as kriging, have been widely used for soil property interpolation, although they exhibit limitations in representing non-linear relationships and highly heterogeneous environmental conditions (Worsham *et al.*, 2010; Zhu *et al.*, 2022; Jeong and Koo, 2025). These limitations have encouraged the increasing adoption of machine learning techniques, which can capture complex relationships between environmental variables and soil properties and improve the accuracy of spatial predictions (Subburayalu and Slater, 2013; Wadoux *et al.*, 2020).

Among these techniques, Random Forest (hereafter RF) has become one of the most widely used algorithms in DSM due to its ability to model complex relationships and reduce overfitting through ensemble learning approaches (Breiman, 2001; Belgiu and Drăguț, 2016). These models use soil properties—such as texture, organic matter content, water-holding capacity, and pH—as input variables to generate accurate predictions of soil quality and condition, enabling the production of detailed maps that integrate soil, climatic, and topographic factors. Such maps facilitate the identification of degraded areas and support informed conservation and land management decisions (Pouladi *et al.*, 2019). Furthermore, integrating DSM with machine learning techniques enhances the delineation of degraded areas and the prioritization of conservation actions under climate change and increasing anthropogenic pressure (Alaminos-Fernández, 2022).

Within this framework, the main objective of the present study is to produce detailed maps of selected soil indicators and properties—specifically SOC, FC, and K Factor (RUSLE)—in Los Alcornocales Natural Park (Andalusia, southern Spain) in order to assess soil quality and health within the context of ENP management. To achieve this objective, the following specific aims are proposed: (i) to identify the most relevant soil and environmental variables for predicting soil quality indicators in ENP of southern Spain, using Los Alcornocales Natural Park as a case study with potential transferability to regions with similar eco-geomorphological conditions; (ii) to evaluate the effectiveness of RF as a spatial modelling technique for representing soil quality in natural environments; (iii) to analyse the spatial variability of surface soil properties across the study area; and (iv) to identify areas with differing levels of soil quality within Los Alcornocales Natural Park in order to support conservation strategies and sustainable management practices.

2. Materials y methods

2.1. Study area

Los Alcornocales Natural Park, with an area of 1,677.38 km², is the third-largest protected area in Andalusia. Located at the western end of the Betic Cordillera and near to the Strait of Gibraltar, it occupies a geostrategic position connecting the Atlantic and Mediterranean oceans, as well as the European and African continents. The Natural Park covers approximately 80 km in length and 35 km in width, characterized by predominantly mountainous relief with steep slopes, a mean slope of 27%, and a mean elevation of 295 m a.s.l., with its highest point at Pico del Aljibe (1,091 m a.s.l.; Fig. 1) (Rubio, 2010).

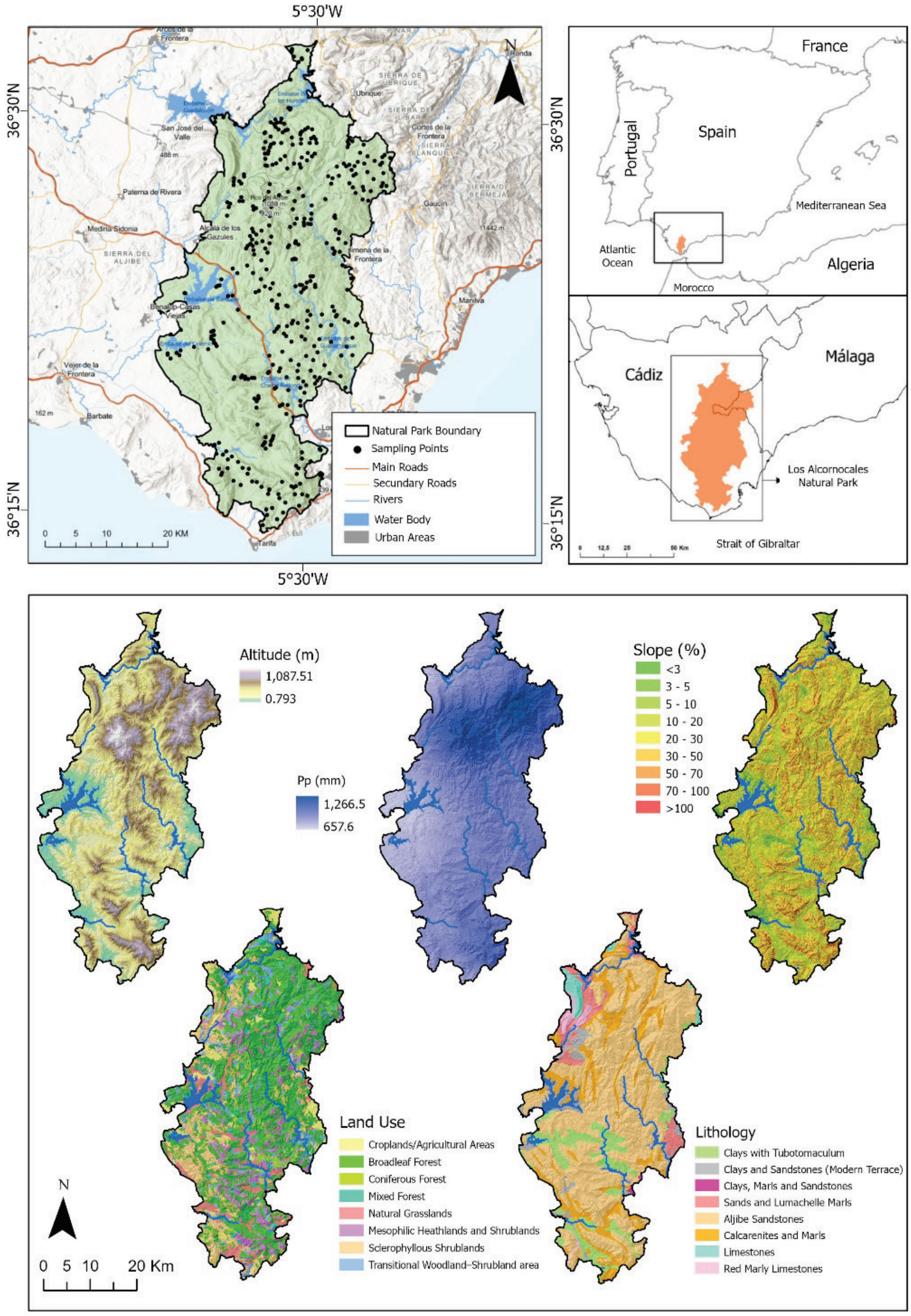

Figure 1. Location of the study area and the analyzed soil sampling sites.

From a lithological perspective, the ENP is situated within the Campo de Gibraltar Flysch, associated with the Aljibe Unit (IGME, 2003; Gutiérrez-Mas, 2016). This geological formation includes Eocene clays, which mainly give rise to Vertisols and Vertic Cambisols, and Oligocene siliceous sandstones, according to the FAO-WRB classification (IUSS Working Group WRB, 2022), responsible for the presence of Cambisols, Luvisols, and Leptosols (Rubio, 2010). These characteristics determine soil texture, fertility, and water-holding capacity, which are key factors governing ecosystem structure and functionality.

The climate is typically Mediterranean, with a significant influence from the proximity to the sea, which moderates temperatures and promotes high precipitation. According to the Köppen classification, the climate corresponds to types *Csa* and *Csb* (AEMET, 2024), while the classification of Rivas-Martínez (1987) places it within the subhumid thermomediterranean belt. Mean annual temperatures are around 17 °C, indicating altitude-related thermal gradients. Average annual precipitation is approximately 1,000 mm, rising to 2,000 mm in higher-altitude areas such as the Sierra del Aljibe, with a marked summer drought. This high rainfall is strongly influenced by the area's geographic location and its exposure to moist Atlantic winds, which produce more frequent and sustained precipitation.

Vegetation is dominated by cork oak forests (*Quercus suber* L.), forming the largest continuous extent of this species in the Iberian Peninsula. This combination of lithological, climatic, and structural heterogeneity makes the Los Alcornocales Natural Park an ideal setting for studying the spatial variability of soil properties and their relationship with soil quality in protected natural areas (Rubio, 2010).

2.2. Field sampling and analysis of soil properties

The soil data used in this study come from the EnBiC2-Lab project, an initiative aimed at assessing the effects of climate change on biodiversity through databases and e-science services provided by LifeWatch ERIC, the European e-Science Infrastructure for Biodiversity and Ecosystem Research, established as a European Research Infrastructure Consortium (ERIC). The project, completed in 2023, integrated information on air, water, soil, flora, and fauna across various Protected Natural Areas of Andalusia to support scientific research and promote the sustainable management of ecosystems.

During this period, a total of 461 soil samples from Los Alcornocales Natural Park were collected and analysed, processed at the Geomorphology and Soils Laboratory of the University of Málaga. For each location, one disturbed sample (~1 kg) was collected, and three undisturbed samples were obtained using 100 cm^3 cylinders, all from the topsoil horizon (0–10 cm depth).

The properties selected for spatial representation included SOC, FC, and the K Factor (RUSLE), determined following the methodology of Sillero-Medina (2022). SOC was determined using the loss-on-ignition method (550ºC) (Guitián and Carballas, 1976), while FC was measured using a pressure plate apparatus (Richards pressure chamber) (Cassel and Nielsen, 1986; Townend *et al.*, 2001). The K Factor was calculated according to the model of Sharpley and Williams (1990) (Eq. 1).

$$\mathrm{K} = \mathrm{Fcsand} \ * \ \mathrm{Fsi} - \mathrm{cl} * \ \mathrm{Forgc} \ * \ \mathrm{Fhisand} * \ 0.1317 \qquad \text{(Eq. 1)}$$

Where K is expressed in units of t·ha·h / (ha·MJ·mm), and:

$$\mathrm{Fcsand} = [0.2 + 0.3 \exp\left(-0.0256\ \mathrm{SAN}\ \left(1 - \frac{SIL}{100}\right)\right)] \qquad \text{(Eq. 2)}$$

$$\mathrm{Fsi} - \mathrm{cl} = [\frac{SIL}{CLA+SIL}^{0.3}] \qquad \text{(Eq. 3)}$$

$$\text{Forgc} = [1.0 - \frac{0.25C}{C+\exp(3.72-2.95C)}] \quad \text{(Eq. 4)}$$

$$\text{Fhisand} = [1.0 - \frac{0.70SN1}{SN1+\exp(-5.51+22.9SN1)}] \quad \text{(Eq. 5)}$$

Where SAND, SILT, and CLAY are the percentages of sand, silt, and clay, respectively, determined using a laser diffraction analyser (Mastersizer 3000) (Marañés *et al.*, 1994); C is organic carbon content; and SN1 is calculated as $(1 - \text{SAND}/100)$.

These point data were used as training variables for the machine learning model used to map the selected properties. The selection of SOC, FC, and the K factor is justified by their representativeness of the organic, physical, and hydrological dimensions of soil, as well as by their usefulness as key indicators of soil quality and ecosystem health (Sillero-Medina *et al.*, 2020, 2020b).

2.3. Spatial modeling methodology

2.3.1. Selection of explanatory variables

For the implementation of the RF model, a diverse set of explanatory variables was selected and harmonized to a spatial resolution of 10 × 10 m, consistent with the spatial resolution of Sentinel-2 imagery from the Copernicus program (operated by the European Space Agency, ESA). These variables include topographic, climatic, spectral, and categorical information, obtained from official sources (Table 1) and processed using GIS tools (ArcGIS Pro) and remote sensing platforms (Google Earth Engine). After compilation, all variables were processed in raster format and prepared for joint analysis, ensuring spatial consistency by unifying the coordinate system (EPSG:25830 – ETRS89 / UTM zone 30N) and standardizing spatial resolution (Table 1). Subsequently, they were classified within the model by nature, distinguishing between continuous variables, such as those derived from topography, climate, and spectral indices, and categorical variables, such as lithology and land use.

Table 1. *Explanatory variables used for soil property prediction.*

Variable	Category	Definition and remarks	Source
Altitude	Topographic	Elevation of a location above sea level, measured in meters (m a.s.l.). Resolution: 5 m.	IGN
Slope	Topographic	Slope angle or inclination of the terrain surface relative to the horizontal, measured in percentage (%). Resolution: 5 m.	IGN
Aspect	Topographic	Downslope direction of a hillside, expressed in degrees (°). Resolution: 5 m.	IGN
LS Factor	Topographic	Key component of the RUSLE model representing the combined effect of slope length (m) and slope steepness (%) (dimensionless variable). Resolution: 5 m.	IGN
Annual Rainfall	Climatic	Average value of accumulated total precipitation in annual millimeters between 1991 and 2020 (mm). Resolution: 500 m.	REDIAM. Junta de Andalucía
Tmean	Climatic	Mean temperature recorded between 1991 and 2020, expressed in degrees Celsius (°C). Resolution: 500 m.	REDIAM. Junta de Andalucía
Tmax	Climatic	Average maximum temperature recorded between 1991 and 2020, expressed in degrees Celsius (°C). Resolution: 500 m.	REDIAM. Junta de Andalucía
Tmin	Climatic	Average minimum temperature recorded between 1991 and 2020, expressed in degrees Celsius (°C). Resolution: 500 m.	REDIAM. Junta de Andalucía

NDVI	Spectral	Normalized Difference Vegetation Index NDVI=(NIR−Red) /(NIR+Red). Period: 2017-2024. Resolution: 10 m.	Sentinel-2
NDMI	Spectral	Normalized Difference Moisture Index NDMI=(NIR-SWIR1) / (NIR + SWIR) Period: 2017-2024. Resolution: 10 m.	Sentinel-2
EVI	Spectral	Enhanced Vegetation Index 2,5*((NIR-R) / (NIR + 6*R-7,5*B + 1)). Period: 2017-2024. Resolution: 10 m	Sentinel-2
B3	Spectral	Green band in the visible spectrum. Period: 2017–2024. Resolution: 10 m.	Sentinel-2
B8	Spectral	Near-infrared (NIR) band. Period: 2017–2024. Resolution: 10 m.	Sentinel-2
Lithology	Categorical	Lithological characteristics. Scale: 1:50,000.	IGME 2003. MAGNA 50
Land use	Categorical	Land use and land cover distribution. Scale: 1:100,000.	Corine Land Cover 2018

2.3.2. Random Forest

The RF algorithm is an ensemble machine learning method that constructs multiple decision trees to improve predictive accuracy and reduce the overfitting characteristic of individual trees. Each tree is trained on a bootstrap sample generated through resampling with replacement, so that approximately 63% of the observations are included in the training set. The remaining 37% constitute the Out-of-Bag (OOB) samples, which are used to estimate the model's prediction error (Breiman, 2001).

This procedure thus makes RF a non-parametric algorithm that is highly robust to non-linear relationships, complex interactions among variables, and noise and outliers. Furthermore, it allows for the simultaneous integration of continuous and categorical explanatory variables (Table 1), which is particularly advantageous in spatial analyses that combine multiple raster data sources. At each tree node, the algorithm randomly selects a subset of explanatory variables (mtry), whose default size corresponds to the square root of the total number of predictors. This additional randomness reduces correlation among trees within the forest, thereby increasing model stability and generalization ability (Heung *et al.*, 2014).

During the training phase, the key model hyperparameters were tuned, including the number of trees (*ntree*), the number of explanatory variables sampled at each node (*mtry*), and the minimum leaf size, with the combination that yielded the lowest prediction error and highest model stability selected in each case.

Furthermore, the model estimates the relative importance of each explanatory variable based on the mean decrease in accuracy, which reflects the increase in OOB error after randomly permuting that variable's values. This procedure enables identification of the explanatory variables that most strongly influence the response variable, providing key information for the ecological and spatial interpretation of the model (Heung *et al.*, 2014).

2.3.3. Statistical metrics for model evaluation

Model validation was performed using two complementary procedures. First, the RF tool in ArcGIS Pro version 3.4 (Esri, 2024) randomly reserved 10% of the data as an independent validation set, using the remaining 90% for model training. This proportion maximizes the information available for model fitting and is consistent with the strong performance of the algorithm when large training datasets are available (Belgiu & Drăguţ, 2016). This partition allows the evaluation of the model's ability to predict unseen data. Second, the OOB error generated internally by the algorithm from observations not included in the bootstrap samples of each tree was used. This approach provides an additional, robust estimate of predictive performance without requiring further data partitioning.

Model performance was evaluated using metrics widely used in spatial modeling and non-parametric regression, including mean squared error (MSE), root mean squared error (RMSE), the coefficient of determination (R^2), and the percentage of explained variance (Yang *et al.*, 2016; Wang *et al.*, 2018), all computed using the validation dataset. Based on this procedure, the final model was run with 500 trees, an *mtry* value of 4, and a minimum leaf size of 5. This configuration was selected based on the training phase results and is consistent with values recommended in the literature to prevent overfitting in RF regression models (Heung *et al.*, 2014).

All processing, modeling, and result generation phases were performed in ArcGIS Pro version 3.4 (Esri, 2024).

3. Results

3.1. Evaluation of predictive model performance and parameter tuning

The RF model was applied to estimate SOC, FC, and K Factor in Los Alcornocales Natural Park. To this end, multiple model runs were performed to identify the configuration that maximized predictive accuracy. In an initial phase, model performance was analysed as a function of the number of trees included in the forest. Table 2 presents the MSE values for each soil property and for each evaluated configuration. The results show a clear trend: as the number of trees increases, MSE progressively decreases, indicating greater model stability and reliability. Although increasing the number of trees entails a higher computational cost, the configuration with 500 trees yielded the best results (Table 2), thus representing the best balance between predictive accuracy and computational efficiency.

Table 2. Mean squared error (MSE) as a function of the number of trees in the RF model for each soil property (training data).

SOC (%)				
Ntree	**100**	**200**	**250**	**500**
MSE	3.486	3.298	3.249	3.149
FC (%)				
Ntree	**150**	**250**	**300**	**500**
MSE	59.19	57.22	56.74	55.67
K Factor (Mg/J)				
Ntree	**150**	**250**	**300**	**500**
MSE	0,.02	0.002	0.002	0.002

Once the optimal configuration was identified, the final RF model—configured with 500 trees, an *mtry* of 4, and a minimum leaf size of 5—was applied to evaluate the agreement between predicted and observed values. Model validation was performed using two complementary procedures: (i) an independent validation set obtained through random partitioning, and (ii) the OOB error, used as an additional internal metric derived from observations not included in the bootstrap samples of each tree. However, random data partitioning may be affected by spatial autocorrelation; therefore, combining the independent dataset with the OOB approach provides a more robust evaluation of model performance.

In this regard, the results indicate that R^2 reaches high values across all three soil properties—FC (R^2= 0.937), the K Factor (R^2 = 0.935), and SOC (R^2 = 0.918) (Table 3)—confirming an excellent level of model fit. Likewise, the low RMSE values reflect a small discrepancy between predicted and measured values, consistent with the overall variability observed in these soil properties. This is particularly notable for the K Factor, where the model achieves an RMSE of 0.018 (Table 3). Similarly, the low standard errors of estimate for the three properties (ranging from 0.008 to 0.010) indicate high model stability and an adequate representation of soil variability in this natural area.

Table 3. RF model accuracy metrics for the three soil properties (500 trees)

	SOC (%)	FC (%)	K Factor (Mg/J)
R-Squared (R^2)	0.918	0.937	0.935
RMSE	0.784	3.137	0.018
Standard error	0.010	0.009	0.008

3.2. Importance of explanatory variables

Figure 2 shows a heatmap representing the relative importance (%) of the explanatory variables used in the RF models for SOC, FC, and K Factor estimation. Additionally, hierarchical clustering analysis, presented through dendrograms, is included to identify structural similarities in the behavior of the modeled soil properties and among the explanatory variables.

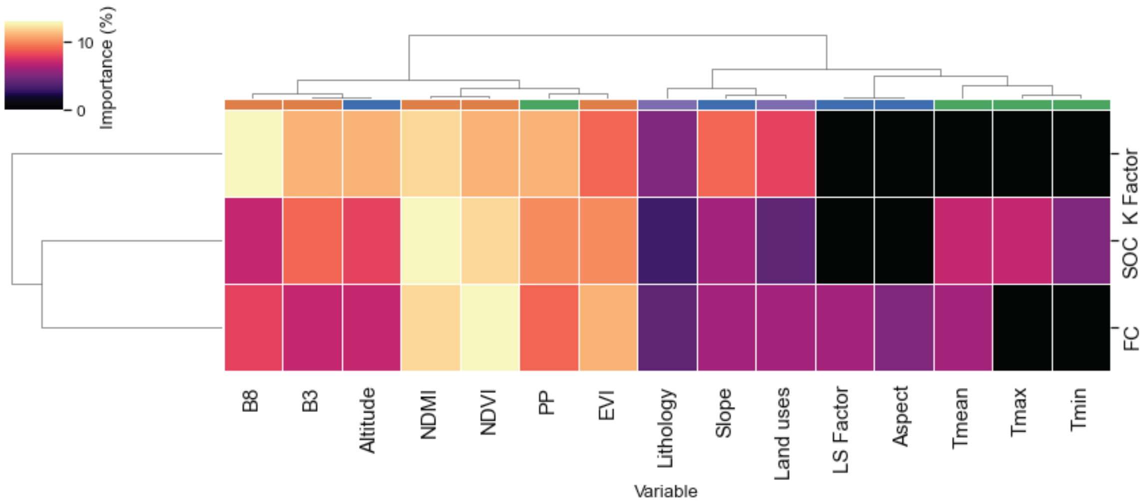

Figure 2. Summary of the importance of the explanatory variables selected after applying the RF model for the three soil properties. Legend: PP, annual precipitation; SOC, soil organic carbon; FC, field capacity; K Factor, soil erodibility.

On the one hand, the dendrogram associated with the models indicates that SOC and FC share the most similar variable-importance patterns, clustering together in a single group, whereas the K Factor forms an independent branch with a distinct structural pattern. The main cluster of spectral bands, vegetation and moisture indices (B8, B3, NDMI, NDVI, EVI), together with elevation and precipitation, exerts a dominant and consistent influence across the three soil properties studied. The main differences between the models are observed in the secondary variables. Specifically, the K factor shows an exclusive dependence on lithology, slope, and land use, with negligible importance of temperature variables (Tmean, Tmax, Tmin), the LS Factor, or aspect. In contrast, SOC is the only property influenced by temperature variables, whereas FC exhibits a specific response to topographic factors such as the LS Factor and aspect.

Regarding the importance of the spectral variables derived from Sentinel-2, they are highly relevant across all three models, particularly NDMI and NDVI, followed by B8. NDMI stands out as the main predictor of surface water content and soil moisture. NDVI and EVI play a strong role in SOC and FC, as they are closely linked to vegetation biomass and water retention. Their high relevance in predictive models confirms that vegetation is the primary regulator of soil processes in this Mediterranean ecosystem, reinforcing the value of multispectral remote sensing as a functional indicator of soil properties.

Likewise, topographic variables (elevation, slope, and the LS Factor) show moderate but consistent contributions, particularly relevant to the K Factor and, to a lesser extent, to SOC. Annual precipitation also has a significant influence across all three models, especially for the K Factor. In contrast, temperature variables (Tmean, Tmax, and Tmin) have a low overall contribution, with greater relevance only in the SOC model.

Finally, categorical variables such as land use and lithology exhibit moderate-to-low importance in the models. Their effect is slightly higher for SOC and FC than for the K factor. This is likely due to the relative homogeneity of the landscape, which is dominated by broadleaf forests as the main land use and by Aljibe sandstone formations across most of the Natural Park.

Overall, these results indicate that, in Los Alcornocales, vegetation presence and condition, together with the climatic regime, are the main drivers of soil behavior. Biotic components not only better explain the distribution of SOC and FC but also play a key role in regulating soil erodibility.

3.3. Spatial prediction results

3.3.1. Soil organic carbon

The RF model applied to Los Alcornocales Natural Park generated a predictive map of SOC content, identifying two main areas with values exceeding 5% (Fig. 3). The first is located in the northern half of the Natural Park, and the second is in the extreme south, both associated with higher elevations and high annual rainfall. These conditions favor dense cork oak forests with a well-developed understory, promoting organic matter accumulation.

The central zone of the Natural Park shows slightly lower values, associated with less dense vegetation dominated by sclerophyllous shrubland and scattered cork oak stands, which limit carbon inputs (Fig. 3). The western and eastern edges show SOC values below 3.5% and 4%, respectively, reflecting transitional areas where cork oak forests become fragmented and agricultural land use emerges—particularly in the west, where the presence of reservoirs has driven land-use change.

In summary, the results indicate that the highest SOC values are associated with areas characterized by higher precipitation, denser vegetation, and greater moisture availability in Los Alcornocales Natural Park. This makes it possible to identify areas of high ecological value, which should be prioritized for soil conservation.

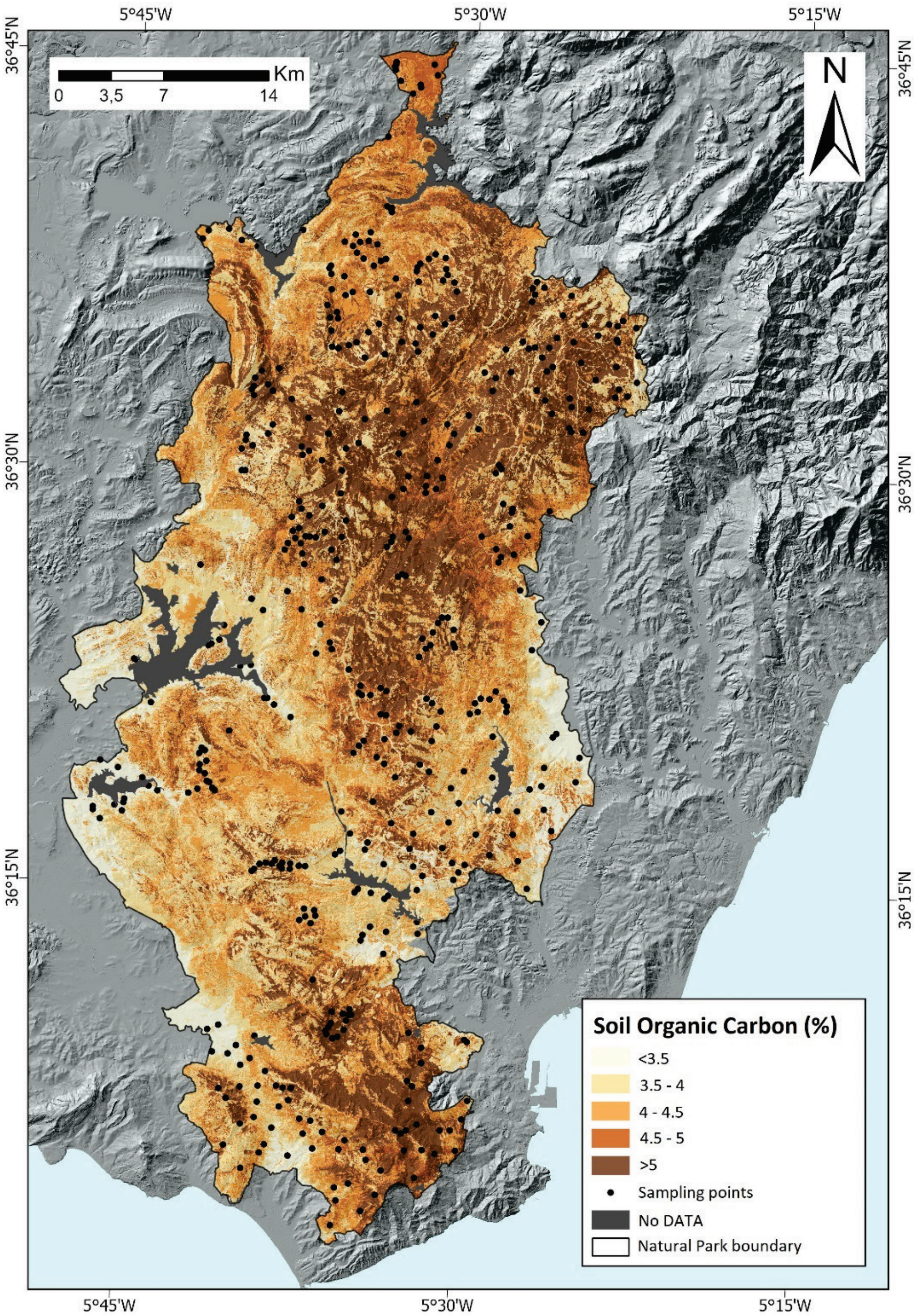

Figure 3. Predictive map of soil organic carbon (SOC) in Los Alcornocales Natural Park.

3.3.2. Field capacity

The spatial prediction of FC in Los Alcornocales Natural Park (Fig. 4) indicates that most of the area exhibits values close to 30%, indicating an optimal level of water retention for vegetation. Two areas with values exceeding 33% are identified in the northern half and in the southernmost part of the Park. These areas coincide with those showing higher SOC values, suggesting a direct relationship between water retention and vegetation density, as biotic factors that regulate soil porosity and the formation of stable aggregates.

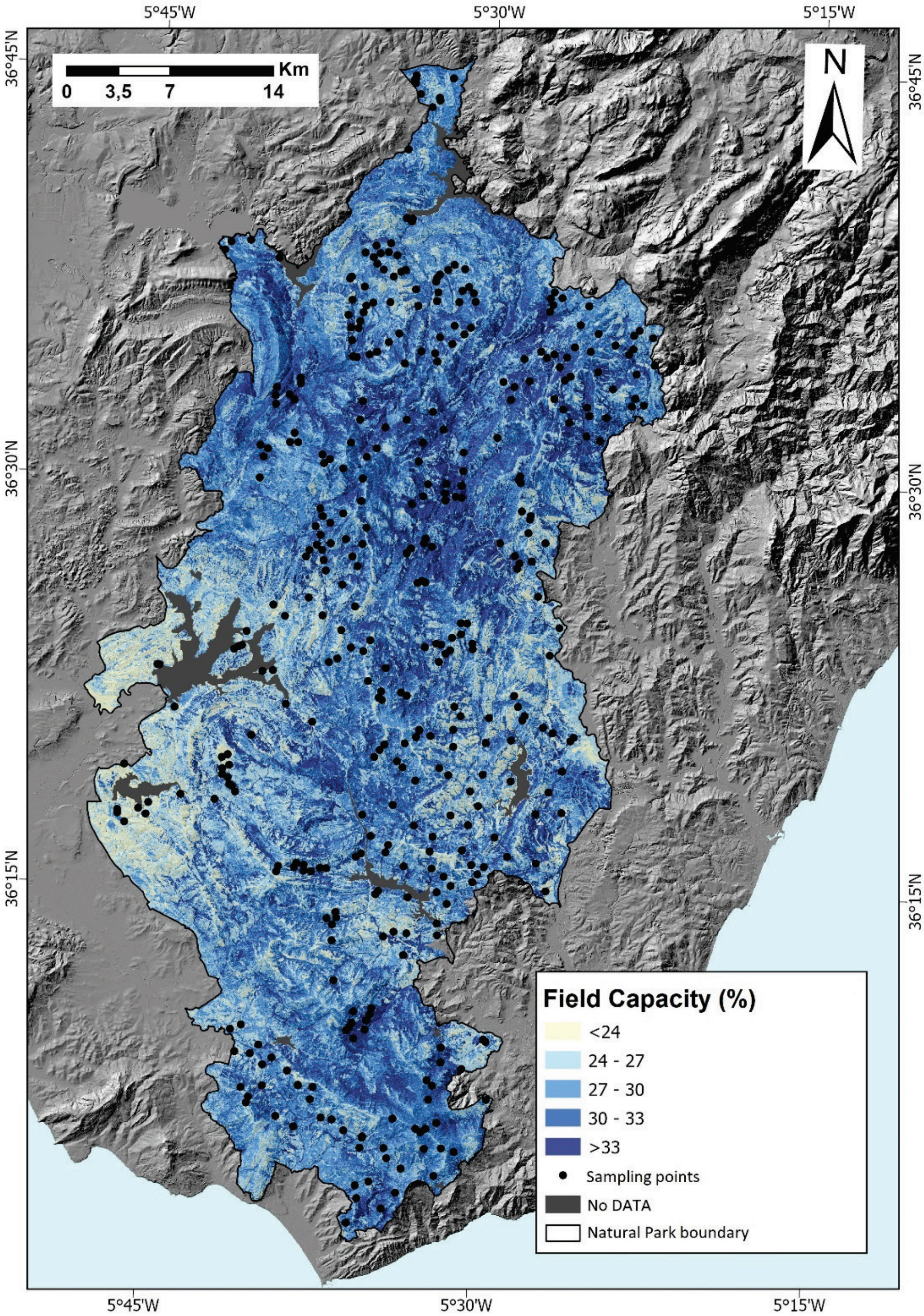

Figure 4. Predictive map of field capacity (FC) in Los Alcornocales Natural Park.

In areas with lower FC, mainly in the western sector and some eastern parts of the Natural Park, values drop below 24%. This reduction is associated with lower vegetation cover, land-use conversion to agriculture, and the presence of sparse sclerophyllous shrubland, which limits water infiltration and promotes the formation of surface crusts, thereby reducing water retention efficiency.

The results demonstrate that the spatial distribution of FC is primarily driven by biotic factors, particulary vegetation cover and its dynamics. This confirms that areas with higher vegetation density exhibit greater water retention capacity, with direct implications for water resource management and soil conservation.

3.3.3. K Factor (RUSLE)

The predictive map of the RUSLE K Factor reveals the spatial distribution of soil susceptibility to water erosion in Los Alcornocales Natural Park (Fig. 5). The results indicate that areas with lower erodibility are concentrated in the northern half and the southernmost part of the Park, where K Factor values remain below 0.055 Mg/J. These areas coincide with higher elevations, high precipitation levels, and dense vegetation—mainly well-developed cork oak forests—indicating that biotic factors dominate the regulation of erodibility. Furthermore, root density and canopy cover significantly contribute to soil stability, minimizing water erosion and promoting optimal soil functioning.

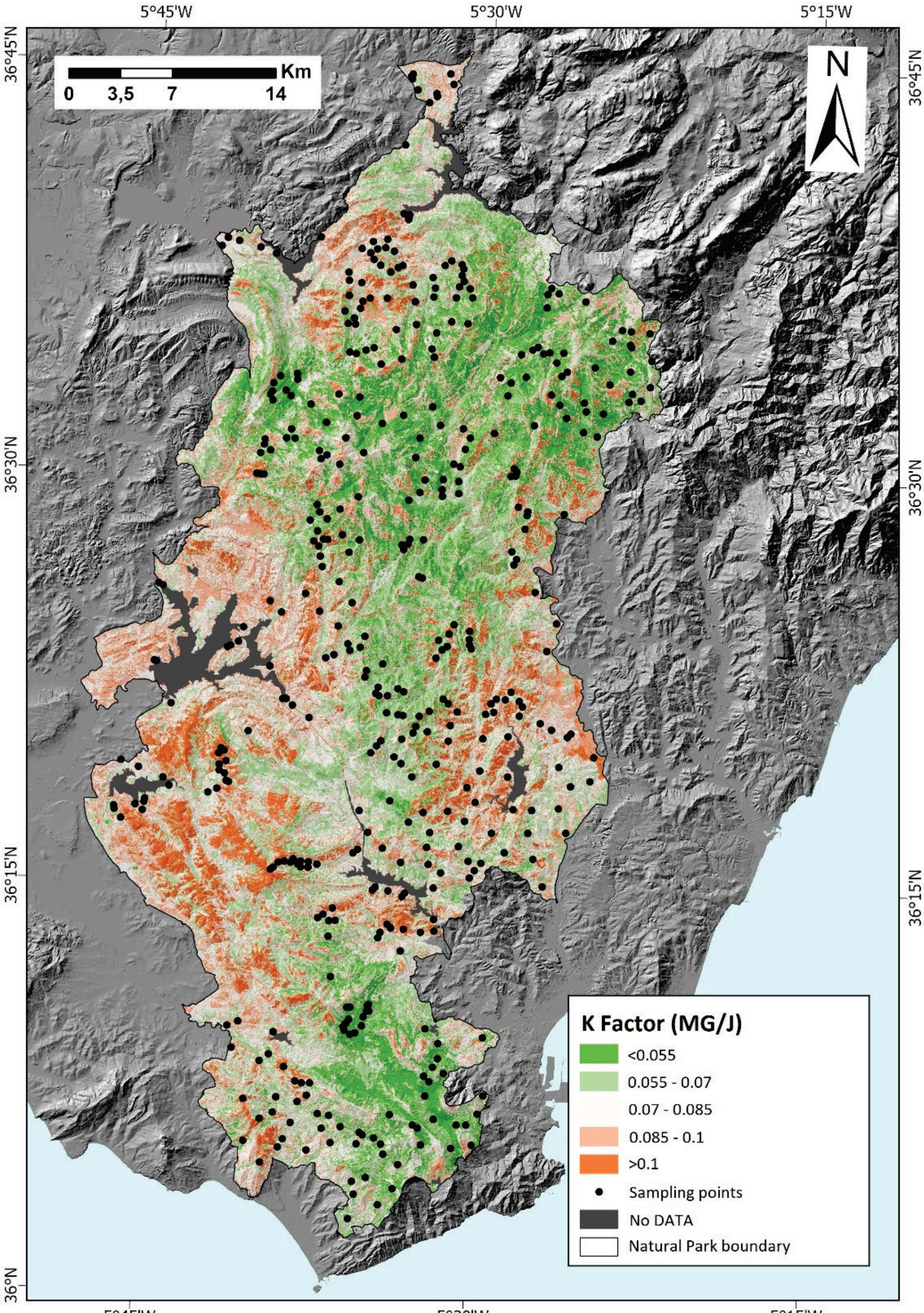

Figure 5. Predictive map of K Factor in Los Alcornocales Natural Park.

In contrast, areas with higher susceptibility to erosion (indicated by reddish tones) are located in the western and eastern sectors, as well as in a small northern portion of the Park, where the K Factor reaches approximately 0.1 Mg/J. These areas are characterized by sparser vegetation—including sclerophyllous shrubland and agricultural land uses—which increases the influence of abiotic factors, such as slope and lithology, on erosion processes.

Overall, the results indicate that most of the territory exhibits low susceptibility to erosion, highlighting the key role of cork oak forests in regulating soil health and quality in this ecosystem.

4. Discussion

DSM remains constrained by significant limitations, primarily due to the scarcity of high-resolution ground-truth data and the marked spatial heterogeneity inherent to many soil properties. These constraints hinder the generation of accurate and reliable predictive soil maps for environmental management (Helfenstein *et al.*, 2024; De Caires *et al.*, 2025). In this context, machine learning techniques—and RF in particular—have emerged as highly effective approaches to address these limitations. These algorithms excel at modeling complex, nonlinear relationships, thereby improving predictive performance by integrating biotic, climatic, and topographic covariates (Breiman, 2001; Zeraatpisheh *et al.*, 2019).

In this study, the modeling of the three assessed soil properties in Los Alcornocales Natural Park (SOC, FC, and the K Factor) demonstrated high predictive performance. Specifically, FC reached an R^2 of 0.937, the K Factor achieved 0.935, and SOC yielded 0.918, indicating a robust model fit and a strong capacity to explain the observed soil variability. Although these values fall within the upper range—or even exceed—those reported in recent literature on DSM for these properties (Trinco *et al.*, 2025; Szabó *et al.*, 2019; Tian *et al.*, 2022), such comparisons must be interpreted with caution given the methodological differences involved. In this regard, the outstanding predictive accuracy obtained here is intrinsically linked to the spatial scale of the study, the high density and strategic distribution of the sampling points (461 samples within a well-defined area), the specific selection of covariates, and the environmental characteristics of the study area (McBratney *et al.*, 2003).

Taken together, these findings reinforce the robustness of the adopted methodological framework and support the suitability of RF as a benchmark algorithm for the spatial prediction of soil properties. Furthermore, they align with previous studies—such as those by Blanco-Bernardeau *et al.* (2014) and Wang *et al.* (2018)—highlighting the superiority of RF over alternative predictive methods for estimating variables such as SOC, particularly in environments characterized by high ecological complexity and spatial variability.

Regarding the importance of the explanatory variables, the results indicate that biotic factors play a predominant role over abiotic ones in regulating soil dynamics within Los Alcornocales Natural Park. Specifically, vegetation-related variables—reflecting abundance, condition, and vigor—together with moisture availability, represented by total precipitation, emerge as the primary drivers of SOC accumulation, water retention capacity, and low soil erodibility.

This pattern highlights the close interdependence between the organic, physical, and hydrological properties of the soil, whereby the input and transformation of organic matter condition soil structure and porosity and, consequently, hydrological behavior and susceptibility to erosion. Thus, these results are consistent with those reported by López-Senespleda *et al.* (2022), who identified tree cover, litter quality, and precipitation as key variables for estimating SOC at the regional scale. Similarly, Ren and Cai (2025) demonstrate that climatic variables, along with vegetation presence and productivity, consistently rank among the most influential predictors in RF models for soil property estimation.

In subhumid Mediterranean environments—such as the climatic context of Los Alcornocales—Sillero-Medina *et al.* (2020) indicate that biotic factors are the main drivers of soil dynamics. Complementarily, studies conducted in similar Mediterranean settings, such as that by Armas *et al.* (2017), highlight that variables including topography, precipitation, and the EVI vegetation index are key predictors of SOC when using RF. The mapping produced in this study supports the notion that biotic factors—particularly vegetation—play a fundamental role in the spatial distribution of SOC.

The mapping results for FC in Los Alcornocales Natural Park, estimated with RF, indicate very high accuracy. These findings are consistent with those reported by Szabó *et al.* (2019) and Nolet *et al.* (2021), who highlight that RF not only enables reliable estimation of soil moisture and other hydrological properties but also identifies the variables controlling their variability, even under limitations associated with model extrapolation.

In turn, the mapping of the soil erodibility factor (K Factor) shows that most of the Park exhibits low susceptibility to erosion, associated with areas of high cork oak density and greater precipitation, whereas erosion increases in areas with more sparse vegetation. This suggests that, in this humid Mediterranean environment, biotic factors play a predominant role in soil dynamics, as reported by Gaitán *et al.* (2017) and Sillero-Medina (2022), who emphasize the importance of vegetation cover and precipitation variability in regulating soil erodibility.

In summary, the results obtained in Los Alcornocales indicate that biotic factors, particularly vegetation, predominantly regulate the spatial distribution of the three analysed soil properties. The presence of dense forests promotes organic matter accumulation and reduces susceptibility to erosion, whereas areas with more sparse vegetation exhibit higher erosion rates. This pattern highlights the close interdependence between the organic, physical, and hydrological properties of the soil and supports previous findings in Mediterranean ecosystems (Armas *et al.*, 2017; Gaitán *et al.*, 2017).

Conclusions

This study confirms that soil mapping based on the RF machine learning algorithm constitutes an effective tool for assessing soil quality and functioning in protected natural areas such as Los Alcornocales Natural Park. The integration of field-derived soil data with environmental, topographic, lithological, climatic, and spectral covariates successfully captures the spatial distribution of key soil properties essential for the conservation and functioning of these ecosystems.

The results highlight the high predictive capacity of the model and underscore the overriding importance of vegetation- and moisture-related explanatory variables—many derived from multispectral data—in driving the spatial variability of soil quality across the study area. In contrast, abiotic factors such as temperature, lithology, and slope exert a limited influence, likely due to the relative geomorphological homogeneity that characterizes the landscape of this Natural Park.

However, the predictive model showed a high degree of fit, characterized by low error rates and high accuracy, suggesting the strong capability of RF and its reliability in spatially representing soil properties in a complex territory, both physically and ecologically. Nevertheless, these results should be interpreted with caution, as validation performed on the same dataset may not fully account for spatial dependence, potentially leading to overfitting and spatial autocorrelation in model performance estimates. To mitigate this limitation, future studies should implement spatially explicit validation schemes that ensure spatial independence between training and validation datasets. Model parameter selection was performed by evaluating combinations of ntree, mtry, and minimum node size, selecting the configuration that minimized prediction error and maximized model stability. Specifically, using 500 trees optimized predictive accuracy, highlighting the model's ability to capture spatial heterogeneity.

In conclusion, the generated mapping clearly identified areas with high soil quality as well as zones more vulnerable to erosion, providing valuable insights for the planning, management, and conservation of the Natural Park. The methodological approach developed shows strong potential for replication in other Mediterranean natural areas with similar environmental characteristics. Nevertheless, key areas for improvement include optimizing the sampling design and the resolution of the explanatory variables, as these factors directly influence model accuracy and generalizability, thereby enhancing the applicability of DSM for environmental management.

Acknowledgements

This work was made possible by the results generated within the research project “Laboratorio de Cambio Climático Ambiental y Biodiversidad (ENVIRONMENTAL BIODIVERSITY CLIMATE CHANGE LAB – LifeWatch EnBi2Lab)” (LW-2019-UMA-01-SU), co-funded by the European Union and the University of Málaga through the European Regional Development Fund (ERDF), within the framework of a call by the Ministry of Economy, Industry, and Competitiveness. Additionally, part of the activities were carried out in the context of the project PPRO-RNM279-G-2023 (RNM279-G-FEDER).

The authors gratefully acknowledge Marina López-Claros, PhD, specialist technician at the Laboratory of Geomorphology and Soils, University of Málaga (IGSUMA), whose contribution was essential to the analysis of the soil properties evaluated in this study.

References

AEMET. Agencia Estatal de Meteorología (2024). *Mapas climáticos de España (1991–2020) y ETo (1996–2020)*. Ministerio para la Transición Ecológica y el Reto Demográfico.

Alaminos-Fernández, A. (2022). *Árboles de decisión en R con Random Forest.* Limencop, 2022. https://hdl.handle.net/10045/133067

Armas, D., Guevara, M., Alcaraz-Segura, D., Vargas, R., Soriano-Luna, M.A., Durante, P., Oyonarte, C. (2017). Mapa digital del perfil del carbono orgánico en los suelos de Andalucía, España. *Ecosistemas*, 26(3), 80–88. https://doi.org/10.7818/ECOS.2017.26-3.10

Belgiu, M., Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Blanco Bernardeau, A., Alonso-Sarria, F., Gomáriz-Castillo, F.G. (2014). Elaboración de un mapa de carbono orgánico del suelo en la Región de Murcia. In: *Tecnologías de la información para nuevas formas de ver el territorio: XVI Congreso Nacional de Tecnologías de la Información Geográfica* (pp. 284–292). https://doi.org/10.13140/2.1.2817.0568

Breiman, L. (2001). Random forests. Machine learning. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bünemann, E.K., Bongiorno, G., Bai, Z., Creamer, R.E., De Deyn, G., de Goede, R., Fleskens, L., Geissen, V., Kuyper, T.W., Mäder, P., Pulleman, M., Sukkel, W., van Groenigen, J.W., Brussaard, L. (2018). Soil quality – A critical review. *Soil Biology and Biochemistry*, 120, 105–125. https://doi.org/10.1016/j.soilbio.2018.01.030

Cassel, D.K., Nielsen, D.R. (1986). Field capacity and available water capacity. In A. Klute (Ed.), *Methods of Soil Analysis. Part 1. Physical and Mineralogical Methods* (2nd ed., Agronomy Monograph No. 9, pp. 901–926). American Society of Agronomy / Soil Science Society of America. https://doi.org/10.2136/sssabookser5.1.2ed.c36

De Caires, S.A., St Martin, C., Atwell, M.A., Kaya, F., Wuddivira, G.A., Wuddivira, M.N. (2025). Advancing soil mapping and management using geostatistics and integrated machine learning and remote sensing techniques: a synoptic review. *Discover Soil.*, *2*(1). https://doi.org/10.1007/s44378-025-00082-z

Dragović, N., Vulević, T. (2021). Procesos de degradación del suelo, causas y enfoques de evaluación. En: Leal Filho, W., Azul, A.M., Brandli, L., Lange Salvia, A., Wall, T. (eds.) *Vida en la tierra. Enciclopedia de los Objetivos de Desarrollo Sostenible de la ONU*. Springer, Cham. https://doi.org/10.1007/978-3-319-95981-8_86

Durán-Sandoval, D., Uleri, F., Durán-Romero, G., López, A.M. (2023). Food, Climate Change, and the Challenge of Innovation. *Encyclopedia,* 3(3): 839–852. https://doi.org/10.3390/encyclopedia3030060.

Esri. (2024). ArcGIS Pro 3.4 [Software]-

Gaitan, J., Navarro, M.F., Tenti, L., Carfagno, P., Pizarro, M.J., Rigo, S. (2017). *Estimación de la pérdida de suelo por erosión hídrica en la República Argentina*. INTA. https://www.researchgate.net/publication/321758306

Guitián, F., Carballas, T. (1976). *Técnicas de análisis de suelos*. Pico Sacro, Santiago de Compostela.

Gutiérrez-Mas, J.M., Gracia Prieto, J., Luján Martínez, M., Sánchez Bellón, Á. (2016). *Geología del Campo de Gibraltar*. En Geología 16: Geolodía Cádiz. Universidad de Cádiz; Ilustre Colegio Oficial de Geólogos de Andalucía.

Helfenstein, A., Mulder, V.L., Hack-ten Broeke, M.J.D., van Doorn, M., Teuling, K., Walvoort, D.J.J., Heuvelink, G.B.M. (2024). BIS-4D: Mapping soil properties and their uncertainties at 25 m resolution in the Netherlands. *Earth System Science Data*, 16(6), 2941–2970. https://doi.org/10.5194/essd-16-2941-2024

Heung, B., Bulmer, C.E., Schmidt, M.G. (2014). Predictive soil parent material mapping at a regional-scale: A Random Forest approach. *Geoderma*, 214–215, 141–154. https://doi.org/10.1016/j.geoderma.2013.09.016

IGME. Instituto Geologico y Minero de España (2003). *MAGNA 50 - Mapa Geológico de España a escala 1:50.000* (2ª Serie).

IUSS Working Group WRB. (2022). *World reference base for soil resources. International soil classification system for naming soils and creating legends for soil maps* (4th ed.). International Union of Soil Sciences (IUSS).

Jeong, M., Koo, H. (2025). Evaluating Spatio-Temporal Kriging with Machine Learning Considering the Sources of Spatio-Temporal Variation. *ISPRS International Journal of Geo-Information*, 14(6), 224. https://doi.org/10.3390/ijgi14060224

López-Senespleda, E., Calama Sainz, R.A., Menéndez-Miguélez, M., del Río, M., Montero, G., Ruiz-Peinado, R. (2022). Estimación del carbono acumulado en la capa orgánica del suelo de las masas forestales en la España peninsular y Baleares. In: *La Ciencia Forestal y su contribución a los Objetivos de Desarrollo Sostenible, 8ª Congreso Forestal Español.* Sociedad Española de Ciencias Forestales. http://hdl.handle.net/10261/279824

Marañés Corbacho, A., Sánchez Garrido, J.A., De Haro Lozaño, S., Sánchez Gomez, S.T., Lozano Cantarero, F.J. (1994). *Análisis de suelo, metodología e interpretación*. Servicio de Publicaciones de la Universidad de Almería.

McBratney, A., Santos, M.M., Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1-2), 3-52. https://doi.org/10.1016/s0016-7061(03)00223-4

McBratney, A., Field, D.J., Koch, A. (2014). The dimensions of soil security. *Geoderma*, 213, 203–213. https://doi.org/10.1016/j.geoderma.2013.08.013

Mesele, S.A., Mechri, M., Okon, M.A., Isimikalu, T.O., Wassif, O.M., Asamoah, E., Ahmad, H.A., Moepi, P.I., Gabasawa, A.I., Bello, S.K., Ayamba, B.E., Owonubi, A., Olayiwola, V.A., Soremi, P.A.S., Khurshid, C. (2025). Current Problems Leading to Soil Degradation in Africa: Raising Awareness and Finding Potential Solutions. *European Journal Of Soil Science*, 76(1). https://doi.org/10.1111/ejss.70069

Minasny, B., McBratney, A.B. (2016). Digital soil mapping: A brief history and some lessons. *Geoderma,* 264(Part B), 301–311. https://doi.org/10.1016/j.geoderma.2015.07.017

Morgan, R.P.C., Nearing, M.A. (Eds.). (2011). *Handbook of erosion modelling*. Wiley-Blackwell.

Nolet, C., Carranza, C., Pezij, M., van der Ploeg, M. (2021). Root zone soil moisture estimation with Random Forest. *Journal of Hydrology*, 593, 125840. https://doi.org/10.1016/j.jhydrol.2020.125840

Panagea, I.S., Berti, A., Čermak, P., Diels, J., Elsen, A., Kusá, H., Piccoli, I., Poesen, J., Stoate, C., Tits, M., Toth, Z., Wyseure, G. (2021). Soil water retention as affected by management induced changes of soil organic carbon: Analysis of long-term experiments in Europe. *Land*, 10(12). https://doi.org/10.3390/land10121362

Pérez-De-Los-Reyes, A.C., Pérez-De-Los-Reyes, M.L., Chocano, D., Sánchez-Ormeño, M., Bravo, S., Ángel, J.A., García, N.F.J. (2018). Estudio de las propiedades de retención de humedad de suelos vitícolas en Castilla-La Mancha (España). *E3S Web Of Conferences*, 50, 01034. https://doi.org/10.1051/e3sconf/20185001034

Peri, P.L., Gaitán, J., Mastrangelo, M., Nosetto, M., Villagra, P.E., Balducci, E., Pinazo, M., Eclesia, R.P., Von Wallis, A., Villarino, S., Alaggia, F., Polo, M.G., Manrique, S., Meglioli, P.A., Rodríguez-Souilla, J., Mónaco, M., Chaves, J.E., Medina, A., Gasparri, I., Pastur, G.M. (2024). Soil organic carbon stocks in native forest of Argentina: a useful surrogate for mitigation and conservation planning under climate variability. *Ecological Processes*, 13(1), 1. https://doi.org/10.1186/s13717-023-00474-5

Pouladi, N., Møller, A.B., Tabatabai, S., Greve, M.H. (2019). Mapping soil organic matter contents at field level with Cubist, Random Forest and kriging. *Geoderma*, 342, 85–92. https://doi.org/10.1016/j.geoderma.2019.02.019

Ren, T., Cai, A. (2025). Global patterns and drivers of soil dissolved organic carbon concentrations. *Earth Syst. Sci. Data*, 17, 2873–2885, https://doi.org/10.5194/essd-17-2873-2025

Rivas-Martínez, S. (1987). *Memoria del Mapa de Series de Vegetación de España. Escala 1:400.000*. Instituto para la Conservación de la Naturaleza (ICONA), Ministerio de Agricultura.

Rubio, A.M. (2010). La densidad aparente en suelos forestales del Parque Natural Los Alcornocales. Proyecto Fin de Carrera. https://digital.csic.es/handle/10261/57951

Ruiz-Sinoga, J.D., Romero-Diaz, A. (2010). Soil degradation factors along a Mediterranean pluviometric gradient in Southern Spain. *Geomorphology*, 118(3–4), 359–368. https://doi.org/10.1016/j.geomorph.2010.02.003

Sharpley, N.A., Williams, R.J. (1990). *EPIC—Erosion/Productivity Impact Calculator I: Model documentation.* U.S. Department of Agriculture, Technical Bulletin. Beltsville, MD, United States.

Sillero-Medina, J.A., Pérez-González, M.E., Martínez-Murillo, J.F., Ruiz-Sinoga, J.D. (2020). Factors affecting eco-geomorphological dynamics in two contrasting Mediterranean environments. *Geomorphology*, 352, 106996. https://doi.org/10.1016/j.geomorph.2019.106996

Sillero-Medina, J.A., Hueso-González, P., Ruiz-Sinoga, J.D. (2020b). Differences in the Soil Quality Index for Two Contrasting Mediterranean Landscapes in Southern Spain. *Land,* 9(11), 405. https://doi.org/10.3390/land9110405

Sillero-Medina, J.A., Rodrigo-Comino, J. Ruiz-Sinoga, J.D. (2021) Factors determining the soil available water during the last two decades (1997–2019) in southern Spain. *Arab J Geosci*, 14, 1971. https://doi.org/10.1007/s12517-021-08265-y

Sillero-Medina, J.A. (2022). *Repercusiones eco-geomorfológicas de la dinámica paisajística reciente, en ambientes mediterráneos contrastados*. [Tesis doctoral, Universidad de Málaga]. http://hdl.handle.net/10630/24375

Subburayalu, S.K., Slater, B.K. (2013). Soil Series Mapping by Knowledge Discovery from an Ohio County Soil Map. *Soil Science Society of America Journal*, 77(4), 1254-1268. https://doi.org/10.2136/sssaj2012.0321

Szabó, B., Szatmári, G., Takács, K., Laborczi, A., Makó, A., Rajkai, K., Pásztor, L. (2019). Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics, *Hydrol. Earth Syst. Sci.*, 23, 2615–2635, https://doi.org/10.5194/hess-23-2615-2019.

Tamara Polo, L.A., Ducuara, J. (2016). *Capacidad de campo y punto de marchitez permanente*. [Informe de prácticas, Universidad de Sucre].

Tian, Z., Liu, F., Liang, Y., Zhu, X. (2022). Mapping soil erodibility in southeast China at 250 m resolution: Using environmental variables and random forest regression with limited samples. *International Soil and Water Conservation Research*, 10(1), 62-74. https://doi.org/10.1016/j.iswcr.2021.06.005

Townend, J., Reeve, M.J., Carter, A. (2001). Water release characteristic. In K. A. Smith and C. E. Mullins (Eds.), *Soil and Environmental Analysis: Physical Methods* (2nd ed., pp. 107–152). CRC Press / Marcel Dekker. https://doi.org/10.1201/9780203908600-8

Trinco, F.D., Zeraatpisheh, M., Turner, H.C., El Mujtar, V., Tittonell, P.A., Galford, G.L. (2025). High-resolution soil organic carbon mapping for enhancing predictive accuracy of environmental drivers in heterogeneous and mountainous landscapes in Patagonia. *Catena*, 259, 109353. https://doi.org/10.1016/j.catena.2025.109353

Valera, A. (2023). Técnicas de cartografía digital para la evaluación del carbono orgánico del suelo. Digital Mapping Techniques for Soil Organic Carbon Assessment. In: *XXIII Congreso Venezolano de la Ciencia del Suelo*, pp. 274-282. https://doi.org/10.13141/RG.2.2.36528.02668

Varvaris, I., Pittaki, Z., Themistokleous, G., Koumoulidis, D., Ouerfelli, D., Eliades, M., Themistocleous, K., Hadjimitsis, D. (2025). Remote Sensing-Based Mapping of Soil Health Descriptors Across Cyprus. *Environments*, 12(8), 283. https://doi.org/10.3390/environments12080283

Wadoux, A.M., Minasny, B., McBratney, A.B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*, 210, 103359. https://doi.org/10.1016/j.earscirev.2020.103359

Wang, B., Waters, C., Orgill, S., Cowie, A., Clark, A., Li Liu, D., Simpson, M., McGowen, I., Sides, T. (2018). Estimating soil organic carbon stocks using different modelling techniques in the semi-arid rangelands of eastern Australia. *Ecological Indicators*, 88, 425–438. https://doi.org/10.1016/j.ecolind.2018.01.049

Worsham, L., Markewitz, D., Nibbelink, N. (2010). Incorporating Spatial Dependence into Estimates of Soil Carbon Contents under Different Land Covers. *Soil Science Society of America Journal*, 74(2), 635-646. https://doi.org/10.2136/sssaj2008.0412

Yang, R.-M., Zhang, G.-L., Liu, F., Lu, Y.-Y., Yang, F., Yang, F., Yang, M., Zhao, Y.-G., Li, D.-C. (2016). Comparison of boosted regression tree and random forest models for mapping topsoil organic carbon concentration in an alpine ecosystem. *Ecological Indicators*, 60, 870–878. https://doi.org/10.1016/j.ecolind.2015.08.036

Zeraatpisheh, M., Ayoubi, S., Jafari, A., Tajik, S., Finke, P. (2019). Digital mapping of soil properties using multiple machine learning in a semi-arid region, central Iran. *Geoderma*, 338, 445–452. https://doi.org/10.1016/j.geoderma.2018.09.006

Zhu, C., Wei, Y., Zhu, F., Lu, W., Fang, Z., Li, Z., Pan, J. (2022). Digital Mapping of Soil Organic Carbon Based on Machine Learning and Regression Kriging. *Sensors*, 22(22), 8997. https://doi.org/10.3390/s22228997