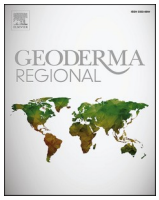
Geoderma Regional 37 (2024) e00780

Contents lists available at ScienceDirect

Geoderma Regional

journal homepage: www.elsevier.com/locate/geodrs



Predicting soil organic carbon with different approaches and spatial resolutions for the southern Iberian Peninsula, Spain

Pilar Durante [a,b], Mario Guevara [c], Rodrigo Vargas [d,*], Cecilio Oyonarte [b,e]

[a] Agresta Cooperative Society, c/Duque Fernán Núñez 2, 28012 Madrid, Spain,
[b] Department of Agronomy, University of Almeria, 04120 La Cañada, Almería, Spain
[c] Centro de Geociencias, UNAM, Campus Juriquilla, Boulevard Juriquilla 3001, Juriquilla 76230, Querétaro, Mexico
[d] Department of Plant and Soil Sciences, University of Delaware, United States
[e] Andalusian Centre for the Assessment and Monitoring of the Global Change (CAESCG), University of Almería, 04120 La Cañda, Alemría, Spain

ARTICLE INFO

Keywords:
Digital soil mapping
SOC concentration
SOC stock
Quantile regression forest (QRF)

ABSTRACT

Quantification and monitoring of soil organic carbon (SOC) stocks across local-to-global scales are needed to assess soil resource management practices and adapt environmental policies. Multiple SOC estimates are available worldwide; however, verification and validation are required to quantify the discrepancies and provide improved estimates. Here, we evaluated four different digital soil mapping (DSM) approaches (i.e., linear models, support vector machine, random forest, and quantile regression forest) to estimate SOC concentration (SOCc) and SOC stocks (SOCs) in the Region of Murcia (11,313 km^2), a complex topographic and climatic area in the southern Iberian Peninsula, at three spatial resolutions (100, 250, and 1000 m). We estimated SOC spatially using a local harmonized database of 255 soil profiles for modeling and 1100 topsoils for external validation. We found that the quantile regression forest (QRF) approach had the best data-model agreement at 100 m spatial resolution, with the highest accuracy percentage (79%), external validation (correlation coefficient of 52%), and spatial interpretability of patterns, especially for SOCc. The QRF model showed a mean SOCc of 12.18 g/kg with an overall uncertainty of 10.54 g/kg and an accuracy percentage of 79%, whereas the total SOCs was 27,572 GgC with an uncertainty of 0.016 GgC. Our results showed that using local environmental covariates and local soil information to predict SOC within this region resulted in a relative improvement in the prediction accuracy of ~40% for SOCc and ~ 65% for SOCs compared to the SOC products derived from national and global databases. Our results showed a large discrepancy between the national and global estimates for reporting SOC locally. Consequently, local-to-regional efforts are needed to describe SOC spatial variability better to reduce uncertainty and improve the assessment of soil resources. We provide the resulting SOC maps with associated spatial uncertainty on the public Environmental Data Initiative Repository at https://portal.edirepository.org/nis/mapbrowse?packageid=edi.1238.2.

1. Introduction

Global environmental changes disrupt biodiversity, structure, and function of terrestrial ecosystems (Pecl et al., 2017). Sustainable land-use management, specifically soil carbon management, is crucial for adaptation to global change and climate regulation (Jobbágy et al., 2000; Wiesmeier et al., 2019). Therefore, quantifying and monitoring soil organic carbon (SOC) across scales is essential for soil management, adaptation of local policies, and assessment of potential impacts (Richer-de-Forges et al., 2019; Vargas-Rojas et al., 2019). Unfortunately, there is still a need to address local-to-regional knowledge gaps in SOC dynamics to inform management practices at an appropriate spatial scale (Cash and Moser, 2000; Wiesmeier et al., 2019).

A current research challenge is to accurately predict SOC stocks at a high spatial resolution, including the whole soil profile (e.g., >30 cm soil depth). Owing to operational complexity, costs, and lack of temporal replication, research efforts are challenged to reproduce the potential high spatial variability of SOC and other soil-related variables (Smith et al., 2020; Vargas et al., 2017). To upscale information from soil surveys, soil mapping has traditionally included a framework considering soil forming processes assessed from soil-landscape and vegetation associations (Hiederer and Köchy, 2012), and, in the last decade, digital

* Corresponding author at: R. Vargas, Department of Plant and Soil Sciences, University of Delaware, 19716, USA.
E-mail address: rvargas@udel.edu (R. Vargas).

https://doi.org/10.1016/j.geodrs.2024.e00780
Received 10 November 2023; Received in revised form 14 February 2024; Accepted 4 March 2024
Available online 6 March 2024

soil mapping (DSM) has enabled large-scale implementation of this framework while providing local information on soil properties (Brus et al., 2011; McBratney et al., 2003; Savin et al., 2019; Searle et al., 2021). Recent DSM efforts have combined data-driven models relying on direct measurements (i.e., in situ pedon information) to provide local (Filippi et al., 2021), national (Vitharana et al., 2019), continental (Guevara et al., 2018), and global (Hengl et al., 2017) estimates of different soil properties (Chen et al., 2022). Moreover, these approaches have shown that the integration of single or multitemporal remote sensing information can improve the prediction of SOC, even when there is limited information to parameterize the models (Fathololoumi et al., 2020; Liang et al., 2020; Schillaci et al., 2017; Zhou et al., 2023). Consequently, DSM offers the possibility of using different digital resources to enhance the spatial representation of SOC from the local to global scales.

Critical regional-to-global efforts have been made to compile soil information worldwide, such as the Profile Analytical Database for Europe (SPADE), Harmonized World Soil Database (HWSD), World Soil Information Service (WoSIS), and the International Soil Carbon Network (ISCN; Harden et al., 2017). However, these efforts have three critical limitations: a) differences in data density or structured information (e.g., map units or point/pedon information) that contribute to biases in representing spatial variability (Kibblewhite et al., 2008; Smith et al., 2020; Trnka et al., 2011; Willaarts et al., 2016), b) low interoperability (Vargas et al., 2017), and b) lack of information on concomitant soil properties (Poeplau et al., 2017). The latter is required because the estimation of SOC stocks (SOCs) is dependent on information on SOC concentration (SOCc), bulk density (BD), and coarse fragment content (CRF) of the target soil depth. While SOCc is usually measured with precision in elemental analyzers, BD and CRF are often missing, which results in added uncertainty in predictions (Durante et al., 2020; Poeplau et al., 2017). Because of these limitations in most available soil databases, comparing and validating derived products (e.g., SOC maps) are needed to better understand their limitations and properly interpret them (Han et al., 2022; Lemercier et al., 2022).

Soil spatial inference is a common approach for generating continuous maps from point data and estimating SOCs across spatial scales (Wang et al., 2018). The SCORPAN approach infers SOC as a function of soil-forming environmental factors, such as climate, topography, vegetation, or land use (McBratney et al., 2003). Numerous strategies have been developed for statistical prediction models to correlate SOC with these forming factors (Kravchenko and Bullock, 1999; Omran, 2012; Robinson and Metternicht, 2006). For example, linear regression approaches are popular because of their computational simplicity and interpretability (Thompson et al., 2006); however, the relationships between soil properties and environmental variables are usually complex and nonlinear (Manning et al., 2015; Moni et al., 2010; Wiesmeier et al., 2019). Recent studies have proposed alternative techniques adapted from data mining, machine learning, and multi-model ensemble methods to account for these nonlinear relationships and improve the predictive capacity of the DSM (Shangguan et al., 2017; Wang et al., 2018). That said, there is no unique or perfect empirical approach as there are multiple limitations with model assumptions, data availability, and spatial scale of the soil predictions (Arrouays et al., 2020; Guevara et al., 2018).

One practical challenge is that the spatial resolution of the SOC variable must be consistent with the spatial scales of both the input covariates and land management (Hartemink, 2006). Therefore, SOC variability must be represented differently across spatial resolutions and management strategies must interpret this information carefully at reliable scales (Lark, 2006; Vargas et al., 2017). Arguably, scales from 1:50,000 to 1:500,000 are recommended to ensure consistency in national and local management strategies (Montanarella, 2015; Pásztor et al., 2019), where the corresponding pixel resolution ranges from 25 m to 250 m (Tobler, 1988). Several efforts have been made to map soil properties worldwide using different spatial domains. For example, the Global SOC Map (GSOC, at approximately 1 km resolution) of the Global Soil Partnership is derived from a country-driven approach to produce the final product as part of the Global Soil Information System (GLOSIS) (Yigini et al., 2018). An alternative global initiative (i.e., SoilGrids250m; at 250 m spatial resolution) was derived using WoSIS-standardized data (Hengl et al., 2017). Finally, a third global project is GlobalSoilMap, a consortium conducted by the International Union of Soil Sciences (IUSS) (Arrouays et al., 2014) to create global digital maps of crucial soil properties at a finer spatial resolution (approximately 100 m). These efforts are valuable for their contribution at the global level and for deriving information across areas with limited soil data. However, the application of SOC estimates performed for a broad spatial level to a smaller area (i.e., global or continental information at the national or local level) may not be able to capture SOC heterogeneity, especially for soil databases collected over different time periods and using a variety of soil analytical methods (Vitharana et al., 2019). Hence, locally derived benchmark information is required for the evaluation, applicability, and interpretability of these global efforts (Han et al., 2022, Villarreal et al., 2018).

Our overarching goal was to evaluate different DSM approaches and environmental covariates derived from remote sensing data to improve local SOC predictions. We focused on a local Mediterranean area (Region of Murcia, 11,313 km^2) with complex climatology and topography in the southeastern Iberian Peninsula, and used a local SOC database to derive SOC estimates. Because of the spatial heterogeneity of SOC in the study area (Conant et al., 2011; Minasny et al., 2017; Xiong et al., 2016), we hypothesized that data-driven models parameterized using local soil information and environmental covariates would capture SOC spatial variation better than available global estimates. To do this, we (1) compared SOC from six available products for the study area (at the national and local levels) and validated them with an independent dataset, and (2) produced a local SOC map testing different data, statistical models, and spatial (i.e., pixel) resolution (100, 250, and 1000 m).

2. Materials and methods

2.1. Study area

The study area is located in the southeastern Iberian Peninsula in the Region of Murcia (Spain). This region is about 11,313 km^2 and presents a complex topography including mountains (reaching 2000 m altitude), high plateaus (500–1000 m), and advanced degradation zones or badlands (> 14% of the territory). This topographic diversity results in contrasting climatic zones. For example, the southeastern area is influenced by the hot, dry winds of the Sahara Desert, which causes a NW to SE line of aridity. Overall, the study area has a mean annual temperature of 18 °C, annual rainfall of 300–350 mm/year distributed in torrential events, and mean annual evapotranspiration of about 900 mm (Albaladejo et al., 2009).

Most of the Region of Murcia (70%) is influenced by human activity occupied by cultivated areas; approximately 20% is covered by shrubland and 10% by pine forest. The soil typology is represented by underdeveloped soils with a wide variety of soil-landscape patterns and a predominance of codominant or associated soils (Alias and Ortiz, 1986). According to the World Reference Base (WRB-IUSS, 2014) the dominant soils are: Calcisols (43%), Leptosols (23%), Regosols (17%), and Fluvisols (9%), followed by Gipsisols, Solonchaks and Kastanozems (Alias and Ortiz, 1986).

2.2. Available SOC products across the study area

A methodological scheme that outlines each step involved in this study is depicted in Fig. 2.

In this study, we included six SOC products derived from DSM frameworks as benchmarks for comparison. Four products correspond to

Global and European approaches; one is a national map, and the last is a local SOC estimate for the Region of Murcia (Table 1). These products provide one or both organic carbon variables: soil organic carbon concentration (SOCc, % or g/kg) and soil organic carbon stock (SOCs, tC/ha). The global approach included the following products: (1) Global Soil Organic Carbon (Hiederer and Köchy, 2012), where the SOC for the topsoil (0 - 30 cm) and the subsoil layer (30–100 cm) were obtained using information from the HWSD and generalized linear techniques. (2) SoilGrids250m system of WoSIS (SG; Hengl et al., 2017), which used random forest-kriging and a global compilation of soil profile data (WoSIS) and environmental layers.

The European approach included the following products. (3) Organic Carbon Content in Topsoils for Europe (OCTOP; Jones et al., 2004), which uses information from the European Soil Database and combines refined pedo-transfer rules and spatially continuous data layers to generate a digital soil mapping by regression kriging approach. (4) The Top Soil Organic Carbon map in EU-25 is based on the Land Use/Cover Area frame statistical Survey 2009 (LUCAS) (ocCont-LUCAS**;** de Brogniez et al., 2015), which uses a generalized additive model between organic carbon measurements from the LUCAS survey (dependent variable) and selected environmental covariates. Urban areas, large water bodies, and areas above 1000 m altitude were masked.

The national-level effort includes the (5) Soil Organic Carbon Stock in Spain (SCSS, Rodríguez Martín et al., 2016), which uses topsoil samples from 4401 locations and ordinary kriging for spatial interpolation in Spain. Finally, the local effort is the (6) Soil Organic Carbon Map in the Region of Murcia (OCMRM, Blanco, 2015), which predicts SOC using random forest and support vector machine algorithms with bootstrapping methods for validation. This last product used over 1000 topsoil samples for the analysis from the LUCDEME (Fight against desertification in the Mediterranean, by its initials in Spanish) database.

To generate comparable information among available DSM products, we extracted the provided values of SOCc and/or SOCs at 30 cm depth for the target area. We excluded the ocCont product, which only has data for the uppermost 20 cm, and urban areas, large water bodies, and areas above 1000 m altitude were masked. For SOCc estimates in SG, we weighted the average of the generated predictions within the depth interval (i.e., 0–30 cm) using the trapezoidal rule for the numerical integration described by Hengl et al., 2017.

First, we performed a qualitative description and a quantitative comparison of the SOC estimates of the six available products at 30 cm. To evaluate the influence of applying SOC estimates performed for the wide spatial area to a smaller area, we compared the available products at two different levels of the spatial domain, from now on referred to as the national spatial domain (Spain) and local spatial domain (Murcia). We computed descriptive statistical parameters of SOCs and SOCc for comparisons: sum (in GgC) for SOCs and mean (in g/kg) for SOCc. To analyze the statistical significance of the mean values of the SOC products, we performed the Tukey plot interval using the 'graphics' package in R (Copenhaver and Holland, 1988).

Second, we tested the product estimates with an independent local database of 255 soil profiles (see section 2.3.12.3) using general statistical comparisons ('graphics' and 'stats' packages in R) and the determination criteria (R^2) and root mean squared error (RMSE) as information criteria.

2.3. Generated local spatial prediction of SOC and its uncertainty

2.3.1. Local soil database and SOC calculation

The independent local database was derived from the LUCDEME Project generated between the years 1986–2004 by the "Ministerio de Medio Ambiente de España" and the support of "Dirección General de Medio Ambiente de la Región de Murcia" (Alias and Ortiz, 1986). This legacy database consists of 255 soil profiles representative of soil typologies over a topographic range of 0–1700 m in altitude. Each profile has morphological and analytical data for each horizon (903 horizons). In addition, 1100 topsoils (0–20/30 cm soil depth) were sampled on a regular 3 × 3 km grid, except the southeast quadrant, distributed in an altitudinal range of 0–1950 m.

We harmonized and revised this soil database following published guidelines (Dobos et al., 2010). To validate the benchmark products, we generated synthetic profiles of 0–30 cm depth from the soil local database for SOCc and SOCs data. The aggregation of horizons depth was done using the equal-area spline technique through the mass-preserving spline ('mpspline') function to generate the synthetic profiles. This technique is based on fitting continuous depth functions for modeling the variability of carbon soil (Bishop et al., 1999). The estimation of SOCs in each soil profile was calculated as follows:

$$SOCs\left(Kg \bullet m^2\right) = SOC\,(g/Kg) \bullet BD\left(Kg \bullet m^3\right) \bullet \left[1 - \left(\frac{CRFVOL}{100}\right)\right] \bullet HSIZE(cm) \quad (1)$$

were BD is bulk density, CRFVOL is the percentage of coarse fragments (above 2 mm in diameter), and HSIZE is the thickness of the horizons. Due to data gaps, BD was estimated using a pedotransfer function

Table 1
Description of available SOC products for the study area.

Acronym	Product	Publisher	Publication date	Soil Database	Spatial resolution	Data	Units
GLOBAL PRODUCTS							
GSOC	Global soil organic carbon	JRC	2012	HWSDB	1 km	2 horizons: 0–30 cm 30–100 m	tC/ha
SG	SoilGrids250m: 7 horizons and Individuals soil layers	ISRIC	2017	WoSIS	250 m	7 horizons: 0–200 cm Top soil: 0–30 cm	g/kg tC/ha
EUROPEAN PRODUCTS							
OCTOP	Organic Carbon Content in Topsoils	JRC	2004	ESDB	1 km	Top soil: 0–30 cm	%
ocCont (LUCAS)	Topsoil Soil Organic Carbon Content	JRC	2014	LUCAS (Land Use/Cover)	500 m	Top soil: 0–20 cm	g/kg
NATIONAL PRODUCTS (Spain)							
SCSS	Soil Organic Carbon and soil organic carbon stock in Spain	INIA	2015	Spanish topsoil database	100 m	Top soil (0_30 cm)	% tC/ha
LOCAL PRODUCTS (Murcia)							
OCMRM	Soil Organic Carbon Map in Region of Murcia	University of Murcia	2014	LUCDEME	25 m	top soil: 0–30 cm	g/kg

adapted from a regional study (Barahona and Santos, 1981). We used the R package GSTAT for the stock estimates and 'mpspline' function, where the propagated error was estimated by the Taylor Series Method (Hengl and Mendes de Jesus, 2016; Heuvelink, 1998; Malone et al., 2009).

To evaluate the SOC spatial distribution of the local predicted products, we log-transformed the SOC original values of the local soil database to generate a normal distribution (Yigini et al., 2018). We tested the correlation between log SOC values with the prediction factors and compared them to SOC original ones. We also provided statistical data from the semivariograms of SOCc and SOCs for 30 cm depth, log-transformed and original ones.

2.3.2. Local spatial covariates

Our DSM approach was based on the SCORPAN conceptual model using the soil forming environmental factors as soil spatial prediction function (McBratney et al., 2003). We generated a covariate stack based on 34 environmental factors to predict SOC (Table 2).

We used dynamic and static variables as predictors for SOC. The static variables were 16 topographic parameters derived from a local digital elevation model (DEM) using the Terrain Analyst functions in the SAGA GIS software (Conrad et al., 2015). The DEM is available from the Geographic Information National Centre (Spain), resulting from interpolating LiDAR national images with a 25 m spatial resolution. We resampled the DEM into 100, 250, and 1000 m pixel sizes and calculated the basic terrain parameters to perform the local model at different spatial resolutions.

The dynamic variables included climatic variables (precipitation and temperature) (Ninyerola et al., 2005); land cover (IGN, 2012) reclassified into 13 classes; forest structural variables, aboveground biomass of forest trees cover from LiDAR data (Durante et al., 2019); and vegetation indexes (VIs). The calculated VIs were The Normalized Difference Vegetation Index (NDVI) and Enhanced Vegetation Index (EVI), which are associated with ecosystem functional attributes related to seasonal dynamics of net primary productivity. These indices were derived from mean annual time series images (2001–2016) of MODIS-Terra images satellite using Google Earth Engine as described in (Arenas-Castro et al., 2019).

These covariates were layer-stacked to build three different harmonized covariate stacks (at 100, 250, and 1000 m spatial resolutions) with the same projection, extent, and pixel size. The covariates were re-scale, re-projected, or, in the case of categorical covariates, rasterized when appropriate. All the statistical and geo-information analyses in this section were performed using R Statistical Software (v4.1.0, R Core Team, 2021), the raster (v3.4–13; Hijmans, 2021), the rgeos (v0.5–3, Bivand and Rundel, 2020), the rgdal (v1.5–22, Bivand, 2021) and the GISTools (v 0.7–4) (Brunsdon and Chen, 2014) R packages.

Table 2
Description of prediction factors used in statistical modeling of soil organic carbon (SOC).

Variables*	Source	Spatial resolution	Description
DEM	IGN (Spain)	25 m	Terrain altitude variability, basis of topographic variables.
PP	ACDPI University of Barcelona	200 m	Mean annual precipitation (mm), period 1951–1999.
TP	ACDPI University of Barcelona	200 m	Mean, minimum and maximum annual temperature (°C), period 1951–1999.
NDVI and CV_NDVI	MODIS-Terra (MOD13Q1)	230 m	Mean annual and coefficient of variation of Normalized Difference Vegetation Index (NDVI), period 2001–2016.
EVI and CV_EVI	MODIS-Terra (MOD13Q1)	230 m	Mean annual and coefficient of variation of Enhanced Vegetation Index (EVI), period 2001–2016.
Lithology	IGME (Spain)	1:200000	Lithological units and their associations.
Soil types	SEIS.net Project (MIMAM- CSIC)	1:100000	Digitized soil map from the National Atlas of Spain (1: 2000,000), IGN 1992).
Land cover	IGN-Corine Land Cover (Spain)	1:100000	Inventory of land covers.
MFE	MAPAMA (Spain)	1:25000	Forest structural types and cover canopy area.
LiDAR (Tree biomass)	IGN (Spain)	5 m (0.5 pt./ m^2)	High-precision of vegetation cover altitude from Light detection and ranging (LiDAR) data.

* DEM: Digital elevation model. PP: Precipitation. TP: Temperature. NDVI: Normalized difference vegetation index. CV_NDVI: Coefficient of variation of NDVI. EVI: Enhanced vegetation index. CV_EVI: Coefficient of variation of EVI. MFE: Map of forests in Spain. LiDAR: Light detection and ranging.

2.3.3. Model fitting, spatial prediction, and uncertainty assessment

To analyze the influence of different methodological criteria for estimating SOC spatial variability, we tested different empirical models at three different spatial (i.e., pixel) resolutions: 100, 250, and 1000 m.

Before model building, a regression matrix included the best correlated environmental factors with SOC local data as covariates. To select them, we considered a balance among higher Pearson coefficient of multiple linear regression, lower error (RMSE), and lower variance inflation factor (VIF) to identify the statistical redundancy (Heiberger et al., 2005). We used the Akaike information criterion (AIC) to determine the best compromise between model accuracy and model parsimony (Rossel and Behrens, 2010).

We tested different models to predict SOCc and SOCs. We fitted linear models (LM,(Chambers et al., 1990) using SOC as the response variable and the regression matrix of covariates as predictors. We used the 'stats' (R Core Team, 2021) R package to perform LM. The support vector machine (SVM) (Weston and Watkins, 1999) was also performed. This algorithm creates a line or a hyperplane, which separates the data into classes. Before performing this model, the qualitative variables were transformed into factors. We used the SVM with the linear kernel method (svmLinear) Kernel since a non-linear decision surface can be converted into a linear equation in a higher dimensional space. This method was implemented in the train function of the 'caret' (v6.0–88; (Kuhn, 2019) R package in R software. The tested random forest (Breiman, 2001) is an ensemble learning method (bagging) of decision trees. Decision trees learn how to best split the dataset into smaller subsets based on different conditions (or nodes) to predict the target value. The RF algorithm operates by constructing many decision trees at training time and outputting the mean of prediction of the individual trees. The number of variables available for splitting at each tree node (mtry) was set 1/3 of the total variables used in the model, and the total number of trees to grow (ntree) was 500. We implemented this method in 'randomForest' (Liaw and Wiener, 2002) package in R software. Finally, the quantile regression forest (qrf) model was performed; since it estimates an approximation of the full conditional distribution of the response variable, the inferred conditional quantiles to build prediction intervals were estimated as surrogates of the value of uncertainty associated with the response variable (Meinshausen, 2006). We used the 'qrf' algorithm implemented in R software for statistical computing in two different packages: 'quantregForest' (QRF) (Meinshausen, 2006) and' 'GSIF' (QRF_G) (Hengl and MacMillan, 2019). QRF validation was calculated from out-of-bag error, and QRF_G model validation was calculated from n-fold cross-validation. In addition, the latter combines predictions by qrf regression and interpolation of residuals (kriging) via the Regression-Kriging (RK) techniques. The information criteria to assess the fit of the different models were RMSE and R^2.

The observed values of the LUCDEME topsoil database were graphically compared with the estimate of the qrf generated local models and

available SOC products of the study area using Taylor diagrams (Carslaw and Ropkins, 2012; Taylor, 2001). In these diagrams, the similarity between two patterns is quantified in terms of their correlation, their centered root-mean-square difference, and the amplitude of their variations (represented by their standard deviations).

2.3.4. Validation and local map selection

An independent, external database was used for model validation. The validation data were based on the 1100 topsoil legacy local dataset described in 2.3.1 section (Fig. 1). Due to the local validation database presented large gaps in bulk density and coarse fragments data, so only the SOCc maps were validated. The agreement between predicted and observed data was measured by the accuracy percentage within the interval of SOC prediction (i.e., the interval corresponding to predicted values and their associated uncertainty). The balance between the predictive model performance and the validation determined the model selection for the final maps of the SOCc and SOCs.

We calculated the relative improvement (RI, Eq. (2) of prediction accuracy of generated SOCc and SOCs maps relying on OCMRM and SG stock maps (i.e., the maps with the most accurate balance of model validation criteria in the external validation).

$$RI = RMSE_{AP} - RMSE_{QRF}/RMSE_{AP} \quad (2)$$

where $RMSE_{AP}$ and $RMSE_{QRF}$ are the root mean square errors of a given available product and the maps generated in this study (SOCc and SOCs), respectively.

Once the model was selected, a scatter plot of the predicted SOC values versus their associated relative uncertainty was used to visualize the spatial distribution of SOC uncertainty related to their values.

3. Results

3.1. Available SOC products, comparison, and validation

We compared the available SOC products at two different spatial domains (national and local) to evaluate their influence on predictions in a smaller area. Our results showed substantial differences on the predictions (Table 3).

The SOCs estimates derived from available SOC products applied at the national spatial domain showed a large diversity among stock values (coefficient of variation CV = 0.35), with a range from 1892 Tg C (by LUCAS map) to 5068 Tg C (by SG map), representing a difference of 63%. At the local spatial domain, the results showed slightly smaller differences than at the national spatial domain (CV =0.28), ranging from 39,984 Gg C to 73,364 Gg C (46% difference), according to LUCAS and SG maps, respectively. We found the opposite pattern in SOCc, where the values remained less variable at the national than at the local domain (CV = 0.07 and CV = 0.22, respectively). The OCTOP map showed the lowest concentration value (22.71 g/kg at the national domain and 11.6 g/kg at the local domain) versus the highest values in the SG map (26.62 g/kg) at the national domain (14%), and the SCSS map (19.51 g/kg) local domain (40%). Therefore, despite the high data variability at the local spatial domain, the CV values are smaller in SOCc than in SOCs data.

The qualitative comparison of the maps at the local level domain derived from the available SOC products depicted different spatial patterns (Fig. 3, Fig. 4). The SOCc maps showed better agreement in the distribution of carbon values than SOCc maps, especially in the higher carbon values across the northeast of the area. The SCSS product showed the least detailed spatial distribution of SOC.

The independent validation of the available SOC products with the local soil dataset revealed a lower data-model agreement for SOCs than for SOCc (Table 4). The best data-model agreement in SOCs corresponded to the SG map with $R^2 = 0.06$ and RMSE = 25.73 GgC, and the best data-model agreement in SOCc values compared to the local OCMRM map with $R^2 = 0.53$ and RMSE = 14.74 GgC. For the Region of Murcia area, a pairwise comparison with a Tukey's test, using a studentized range distribution at a 95% confidence interval, indicated that the differences between the means of stock available products and the profile samples were statistically significant. Nevertheless, we found no statistically significant differences in the carbon concentration values in the SG and OCMRM maps (Fig. S 1).

3.2. Generated local SOC map: model fitting, spatial prediction, and uncertainty assessment

3.2.1. Local SOC values and spatial covariates

The statistical description of the SOC average profile (Fig. S 2) from the local database (Alias and Ortiz, 1986) showed that most of the SOCc is in upper horizons (0–30 cm) and decreases with soil depth. The total mean SOCc was 8.22 g/kg (10.49 SD) and 12.22 g/kg (12.52 SD) for the 0–30 cm depth. The mean SOCs was 26.71 kg/m^2 (16.96 SD) for the upper horizons (0–30 cm). Regarding the probability distribution, both SOCc and SOCs revealed a log-normal distribution with a right-skew.

Fig. 1. Spatial distribution of samples derived from the LUCDEME Project in the Region of Murcia (Spain). Dots represent 255 soil profiles and "+" represent 1100 topsoil samples (0–20 or 0–30 cm depth).

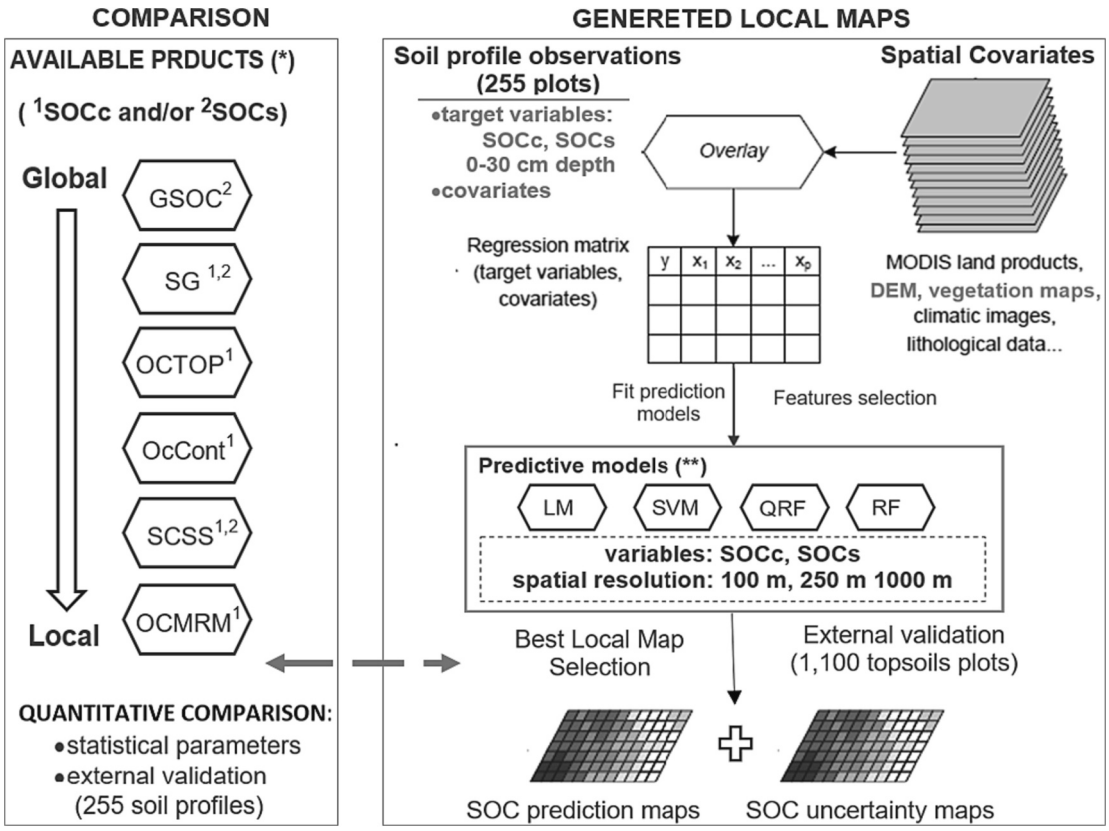


Fig. 2. Workflow describing the steps to produce the local map of soil organic carbon concentration (SOCc) and soil organic carbon stock (SOCs).
Note: (*) Global soil organic carbon (GSOC), SoilGrids250m (SG), Organic Carbon Content in Topsoils (OTCOP), Topsoil Soil Organic Carbon Content (ocCont), Soil Organic Carbon Stock in Spain (SCSS), Soil Organic Carbon Map in Region of Murcia (OCMRM). (**) Linear model (LM); support vector machine (SVM). quantile regression forest, random forest (RF).

However, their spatial autocorrelation behavior differed from SOCc and SOCs at 0–30 cm depth with a nugget-to-sill ratio (NSR) <25% for SOCc (Fig. S 3), which indicates a strong spatial structure. Conversely, SOCs presented no significant spatial autocorrelation structure.

The selected environmental drivers to describe the spatial variability of SOC at 30 cm depth and the three spatial resolutions (100, 250, and 1000 m) were mainly related to topographic and vegetation variables (Table S 1). Specifically, the most consistent variables for predicting SOCs and SOCc were slope, plane curvature, and vegetation index (EVI) at 100 m pixel size. For the middle spatial resolution (250 m), the vertical distance to the channel network (related to vertical valley depth) and vegetation index (NDVI) were the most consistent. The coarsest spatial resolution (1000 m) showed that relative slope position, vegetation indexes (NDVI and EVI), and canopy cover (shrubland and total vegetation) were the most consistent variables. Moreover, precipitation, vegetation indexes, tree biomass, and different topographic variables are the most correlated covariates common to the three spatial resolutions in SOCs. Vertical distance to channel network, topographic variables, and vegetation indexes were for SOCc.

3.2.2. *Evaluation of SOC model fitting*

The comparison among the generated local models of SOC spatial prediction based on the R^2 and RMSE as model information criteria revealed large differences in accuracy (Fig. 5; Table S1). The QRF_G model showed the most accurate information criteria values for both SOCc and SOCs models. In SOCc, the R^2 ranged from 0.90 to 0.93 and RMSE 3.38–3.93 g/kg. In SOCs, the values varied from 0.91 to 0.94 and 4.08–5.16 GgC for R^2 and RMSE, respectively. The rest of the models showed a prediction accuracy much lower in stock than concentration values at the three spatial resolutions (100 m, 250 m, and 1000 m). Both information criteria (R^2 and RMSE) varied from 0.26 to 0.41 and 9.65–11.76 g/kg for SOCc and 0.03–0.17 and 15.45–17.66 GgC for SOCs. The LM and QRF showed the best model performances, with the lower accuracy values at 1000 m spatial resolution.

The SOCc qrf approaches (QRF_G and QRF) had the best balance between R^2 and RMSE values and had the advantage of reporting model uncertainty.

The comparison of SOCc estimates (from both generated models and available products) with the local topsoil samples (LUCDEME database) representing the "observed" values showed that the 25 m spatial resolution map (OCMRM product) had the best model-data agreement and the lowest RMSE, but a greater standard deviation (Fig. 6). Next in order were the QRF models, specifically those at 100 m spatial resolution. The QRF_G models showed lower accuracy of the amplitude of the variations (i.e., the standard deviation), lower correlation, and higher RMSE than QRF models (Fig. 5). Regarding model performance, SCSSc had the lowest correlation and the highest RMSE models.

Estimates of the covariate importance in the QRF models revealed the slope as the highest value, followed by maximum temperature and plane curvature for the 100 m spatial resolution model. The most important covariates at 250 m spatial resolution were also linked to topography (vertical distance to channel network, plane curvature, and Ls-factor, listed in decreasing order), followed by climatic variables (maximum temperature) (Fig. S 4).

3.2.3. *Local SOC spatial prediction and uncertainty assessment*

The summary of data prediction and uncertainty associated with SOCc and SOCs showed small differences across spatial resolutions between the quantregForest (QRF) and GSIF (QRF_G) algorithms. The QRF predictions and uncertainty generally had lower values than those from

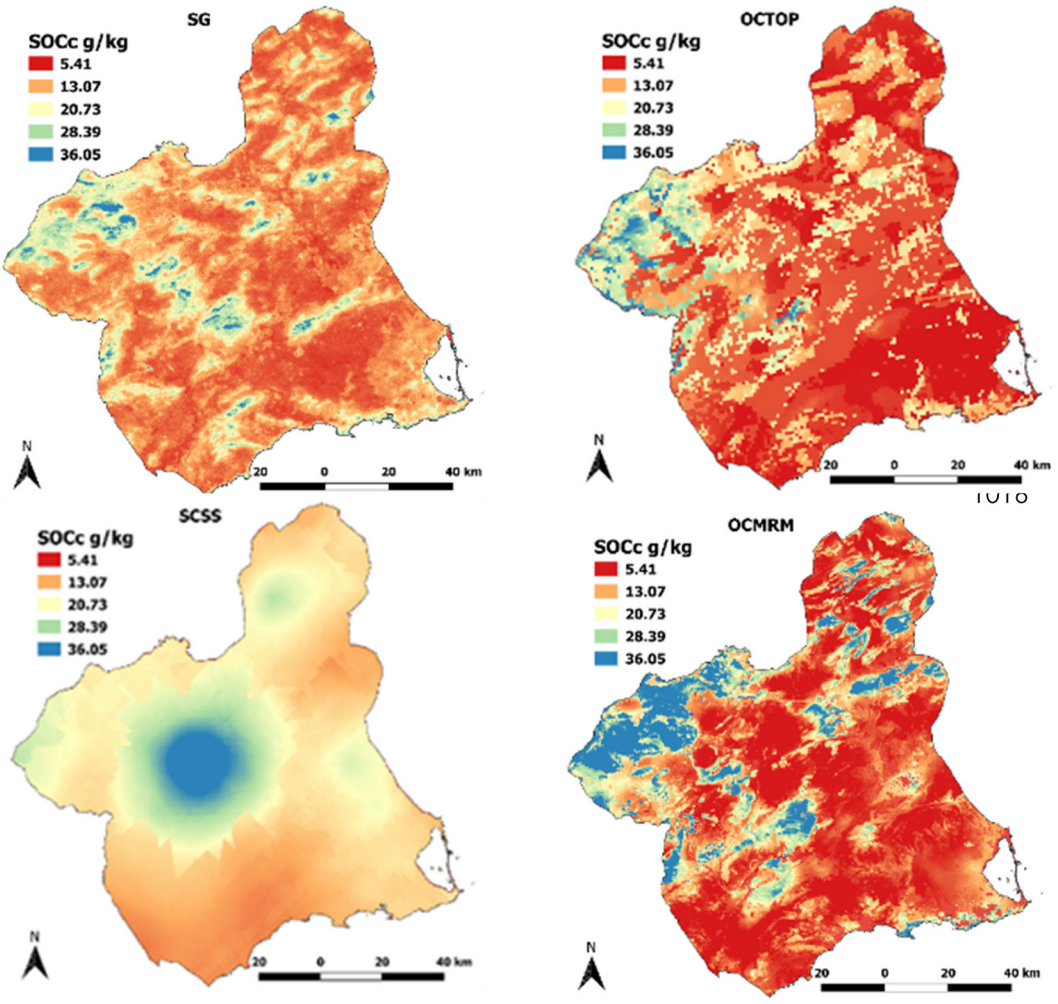


Fig. 3. Predictions of soil organic carbon concentration (SOCc) for the region of interest including SoilGrids250m (SG), Organic Carbon Content in Topsoils (OTCOP), Soil Organic Carbon Stock in Spain (SCSS), and Soil Organic Carbon Map in Region of Murcia (OCMRM). Data display was stretched by the cumulative pixel count cut method (default range 2%–98%).

QRF_G for SOCs along the three spatial resolutions. However, the predicted values for SOCc were slightly higher in QRF than QRF_G across spatial resolutions (Table 5). All available SOC products derived from global databases generally showed higher SOC values than those predicted by the local models. Specifically, the SCSS national map showed about 1.5 times of SOCc (19.50 vs. 12.18 and 11.20 g/kg C in QRF and QRF_G, respectively) and the SG global map about 2.6 times in SOCs: 73,364 vs. 27,646 and 28,435 Gg C in QRF and QRF_G, respectively.

3.2.4. Validation and final model selection

Overall, the QRF had a better data-model agreement than QRF_G using the external-independent topsoil data validation (Table S2). The largest prediction errors of both RMSE and mean absolute error (MAE) were at coarsest spatial resolution (12,19 g/kg and 7.85 g/kg for QRF; and 12,21 g/kg and 8.01 g/kg for QRF_G at 1000 m spatial resolution, for RMSE and MAE respectively). The accuracy percentage (i.e., the interval corresponding to predicted values and their associated uncertainty) was substantially higher in QRF than in QRF_G, especially at 100 m spatial resolution (79% in QRF versus 49% in QRF_G, Table 6).

The QRF model at 100 m spatial resolution achieved a better balance considering model performance, agreement with the reference soil information (i.e., topsoil local database), and independent external validation. The analysis of the residuals of the QRF model at 100 m spatial resolution confirmed the absence of spatial autocorrelation structure in both SOCc and SOCs. In general, the values of the QRF maps ranged from 5.1 to 21.8 g/kg and 4.1 to 31.1 tC /ha for SOCc and SOCs, respectively (Fig. 7). Both maps were consistent in the areas with low SOC located at low elevation (0 to 300 m), gentle slopes (<2%), and cultivated land. This area corresponds to the driest part of the Region with less mean annual rainfall (< 200 mm) and high mean annual temperature (>15 °C). The highest SOC values were found in coniferous forests in steeper and humid areas. Overall, the available SOC products derived from national-to-global databases showed higher ranges with a considerable overestimation of values, especially for SOCs.

The relative uncertainty analysis of the QRF model revealed an inversely proportional relationship with SOC-predicted data, emphasizing the extreme values. This pattern was more pronounced for SOCc, where the lower values presented very high uncertainty (0). The relative improvement in the prediction accuracy of the SOCc and SOCs maps produced in this study from the QRF model at 100 m spatial resolution compared to the available SOC products of the OCMRM and SG maps were 40.8% and 63.8%, respectively.

The final selected QRF model at 100 m spatial resolution for SOCs and SOCc and their associated uncertainties maps will be available on the public Environmental Data Initiative Repository at https://portal.edirepository.org/nis/mapbrowse?packageid=edi.1238.2

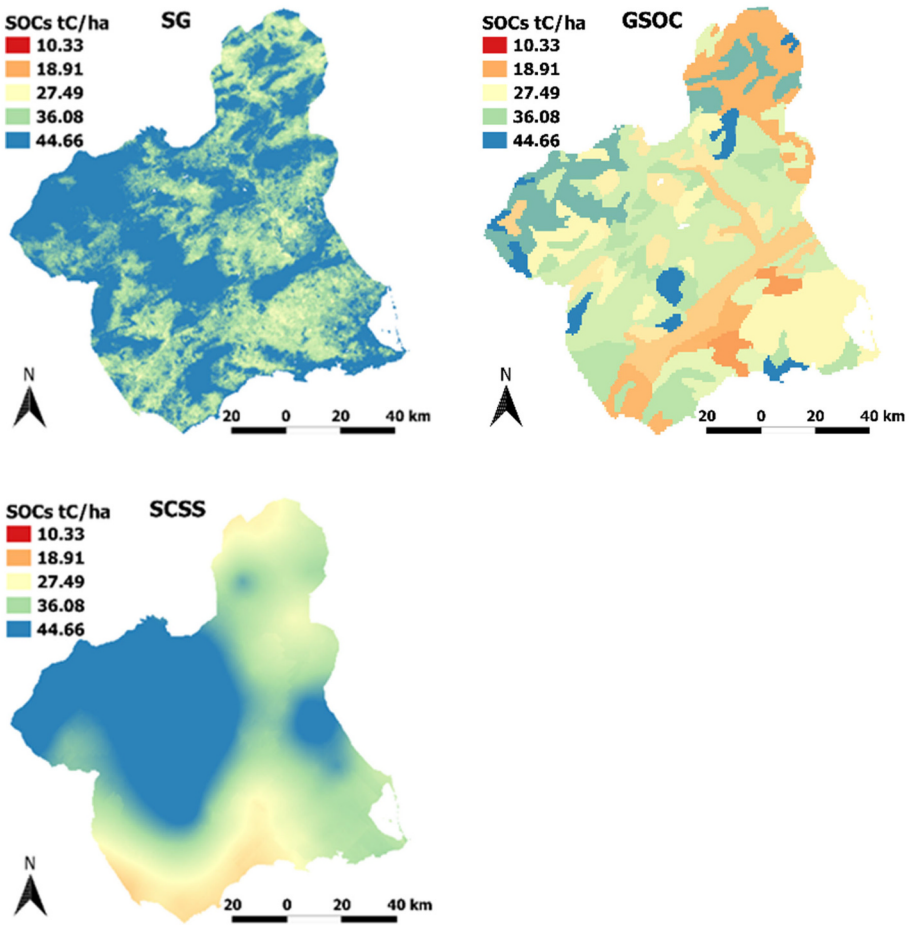


Fig. 4. Predictions of soil organic carbon stocks (SOCs) for the region of interest including SoilGrids250m (SG), Global soil organic carbon (GSOC), Soil Organic Carbon Stock in Spain (SCSS). Data display was stretched by the cumulative pixel count cut method (default range 2%–98%).

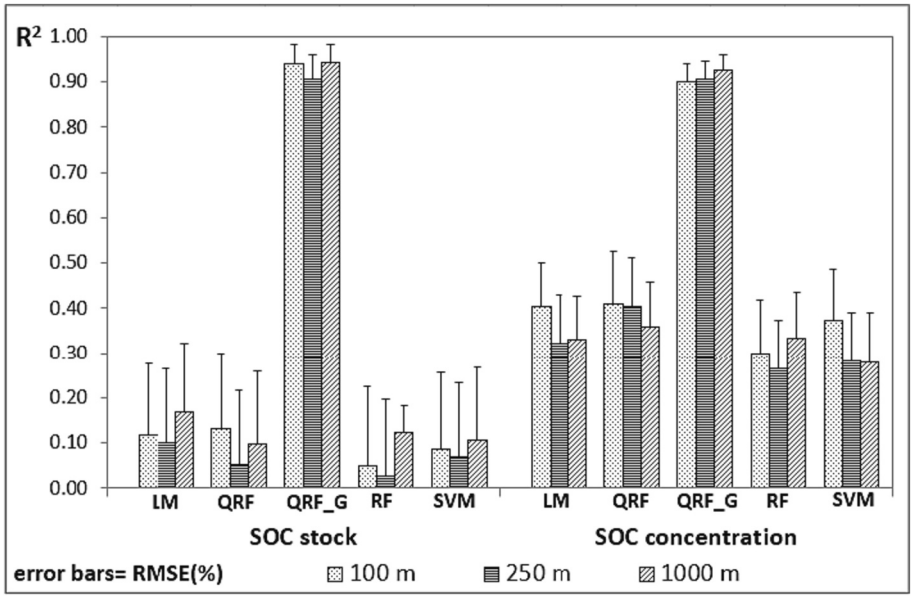


Fig. 5. Bar graph of R-squared and residual error (i.e., error bars) of the statistical local modeling of SOCs and SOCc (g/kg) at 30 cm depth and different spatial resolutions (100 m, 250 m and 1000 m).
Note. Predictive models are: linear model (LM); quantile regression forest (QRF = QuantregForest R package; QRF_G = Gstat R package); random forest (RF = Caret R package); support vector machine (SVM = Caret R package).

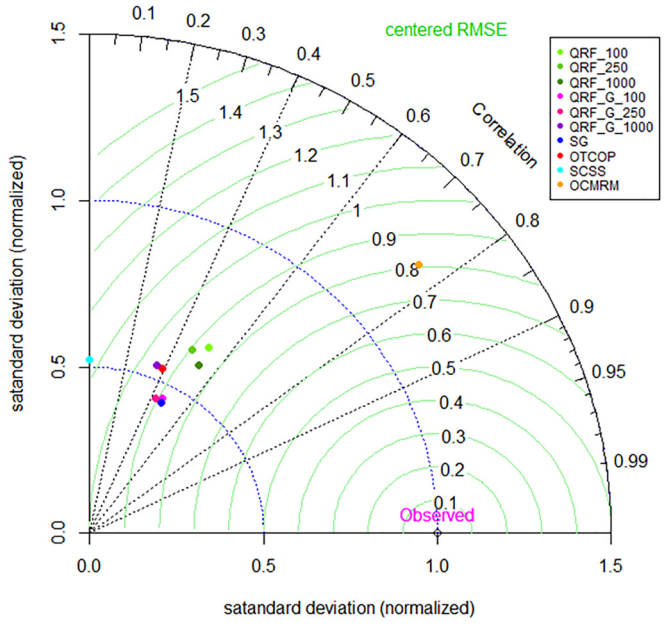


Fig. 6. Taylor diagram illustrating performance of generated local models and available soil organic carbon concentration (SOCc) products compared with the local topsoil samples (LUCDEME database) representing the "observed" values. Note: SoilGrids250m (SG), Organic Carbon Content in Topsoils (OTCOP), Topsoil Soil Organic Carbon Content (ocCont), Soil Organic Carbon Stock in Spain (SCSS), Soil Organic Carbon Map in Region of Murcia (OCMRM), quantile regression forest (QRF = QuantregForest R package; QRF_G = Gstat R package.

4. Discussion

4.1. Comparison of available SOC products

Different available soil dataset collections, spatial resolutions, and spatial predictive modeling for SOC estimates led to varying interpretations of SOCc and SOCs across the study area. These differences highlight the challenges associated with predicting SOC across heterogeneous semiarid regions. Discrepancies in SOCs were larger than those in SOCc among the products analyzed. This is probably due to the generalized absence of bulk density and coarse soil material (> 2 mm) in most soil databases, which are the main parameters required for estimating SOCs. These important variables are often derived from pedotransfer or extrapolation functions (Jalabert et al., 2010), resulting in additional bias and systematic errors in the calculation of SOCs (Durante et al., 2020; Poeplau et al., 2017). Therefore, we echo the call for including bulk density and coarse fragment information in future soil surveys or recovering it from legacy datasets (Hendriks et al., 2019).

There was no clear consistency among the tested products in terms of spatial domain and spatial resolution. Our results do not support the expectation that local products and a higher spatial resolution will better agree with locally derived SOC information. We highlight that the SOCc maps from SG (global domain, 250 m spatial resolution) and OCMRM (local domain, 25 m spatial resolution) had the most similar estimates (g/kg) and balance of model validation criteria (R^2 and RMSE) with the independent data for external validation. However, the estimates for the SG map were extrapolated from global information, as none of the 43 soil samples used for peninsular Spain in SG were from our study area. Both maps were modeled using a similar machine learning technique (i.e., random forest); however, while the selected covariates in the OCMRM map referred to climate, land cover, soil types, and terrain morphology, the SG prediction model included environmental layers from remote sensing data. Therefore, our results suggest that applying machine learning approaches combined with single and/or multitemporal remotely sensed satellite indices can result in comparable predictions of the spatial distribution of SOC. Therefore, we emphasize that there must be a balance between soil training data, statistical models, and covariates (Fathololoumi et al., 2020; Lemercier et al., 2022; Liang et al., 2020; Schillaci et al., 2017), and adding new information in regions with sparse training data could result in higher model uncertainty until the spatial bias of information is reduced (Stell et al., 2021; Smith et al., 2022).

We postulate that the SG map could be considered the best available SOC product for the study area, with a moderate spatial pixel resolution (i.e., 250 m). However, we clarified that this product has some limitations, as several studies that compared estimated SOCs showed wide discrepancies and overestimation (Han et al., 2022), especially in areas with low SOCs (Silatsa et al., 2020; Vitharana et al., 2019). This overestimation may be due to the absence of profile observations (i.e., training data), resulting in biased training data and spatial predictions (Lombardo et al., 2018; Stell et al., 2021; Vitharana et al., 2019). In our results, the SG map estimated the highest SOCc compared with the available SOC products. These observations confirm the need to reassess local-domain estimates of soil SOCs to gain insights into the effects of different methodological criteria in estimating SOC spatial variability.

4.2. Generated local prediction of SOC

We tested the available SOC products using locally derived information and empirical models at three spatial resolutions to provide a critical assessment. The QRF approach had the best data-model agreement at a spatial resolution of 100 m, with the best accuracy and external validation. The relative improvement in the prediction accuracy of the local SOCc and SOCs maps produced in this study was approximately 40% and 65%, respectively, in comparison to the available SOC products with the most accurate balance of model validation criteria in the independent, external validation (i.e., OCMRM for SOCc and SG for SOCs).

The QRF approach approximates the full conditional distribution of the SOC values. Therefore, it has the advantage of reporting the model uncertainty. This allows the building of a spatially explicit uncertainty map associated with the SOC values. The QRF performs the uncertainty and interpretation of its patterns better than other modeling methods, especially in Mediterranean areas with low SOC values and sparse spatial sampling, such as in the present study (Lombardo et al., 2018). However, the interpretability of the relationships between covariates and target variables must be improved to provide a better understanding of the model.

Consistent with other studies, local sampling distribution was more relevant than density in capturing SOC variation (Zeraatpisheh et al., 2019). The two maps derived using the local database showed low differences in prediction accuracy despite their different sampling densities (i.e., 1922 systematic sampling vs. 255 representative profiles of soil taxonomic for OCMRM and QRF_100 maps, respectively). The accuracy of the generated local models was higher for SOCc than for SOCs, and this difference can also be attributed to the error propagation of extrapolated parameters (i.e., bulk density and coarse fragments) that may have a stronger influence on predicting SOCs (Poeplau et al., 2017).

Our results showed that the regression-kriging technique (i.e., predictions plus kriging of the residuals) can lead to overfitting when variogram modeling is undersampled. It has been reported that a well-represented multivariate feature space should preferably have 300 sample points and at least ten observations per covariate (Hengl et al., 2017; Webster and Oliver, 2001). This could explain the discrepancies in the QRF_G models, which showed the best model performance but the lowest accuracy in external validation at the three spatial resolutions tested (100, 250, and 1000 m). Our results highlight the importance of model external validation to better evaluate model performance and avoid spurious data-model agreements influenced by over-parameterization (Zeraatpisheh et al., 2019).

The errors of the local models produced in this study were relatively

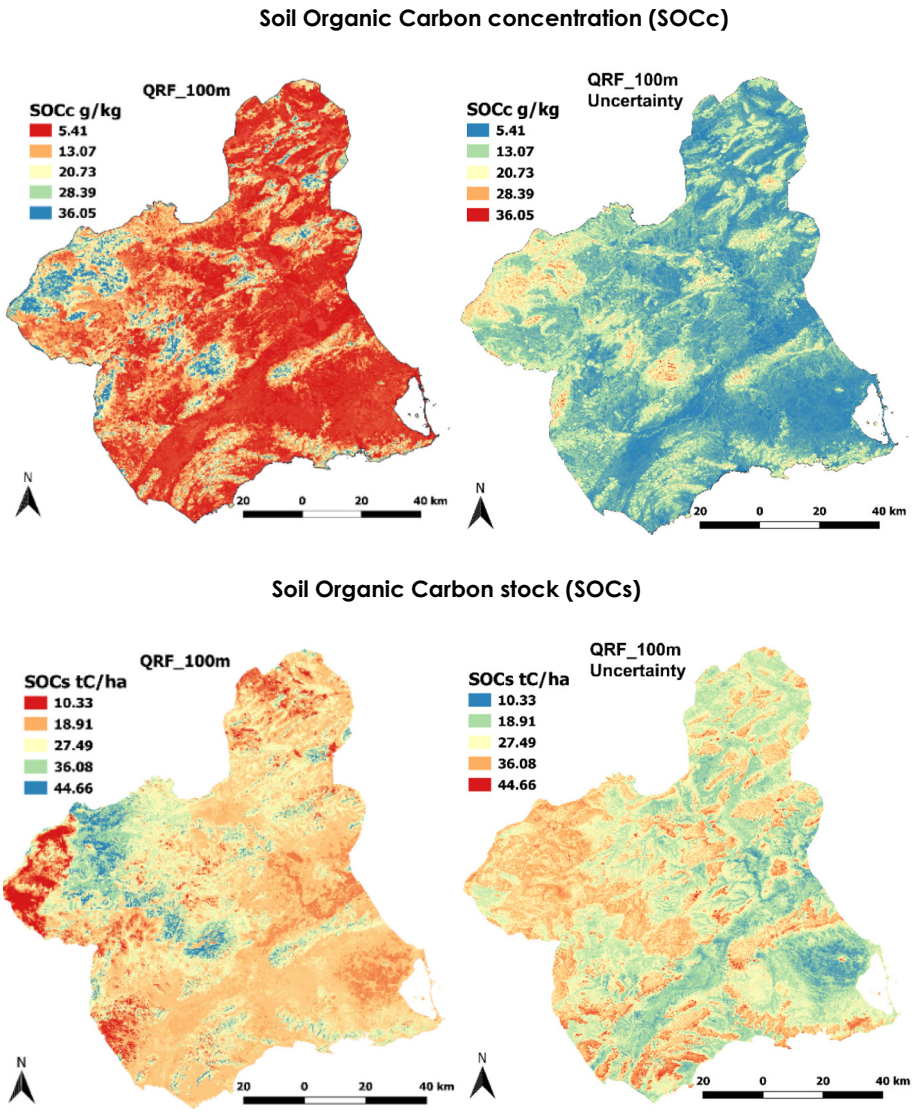


Fig. 7. Generated prediction maps of soil organic carbon concentration (SOCc) and SOC stocks (SOCs), and their uncertainties at 100 m.
Note: QRF = quantile regression forest models estimate by QuantregForest R package. Data display was stretched by the cumulative pixel count cut method (default range 2%–98%).

Table 3
Comparison of available products of soil organic carbon concentration (SOCc) and soil organic carbon stock (SOCs) at national (Spain) and local (Región de Murcia) domains.

AVAILABLE PRODUCTS*			GSOC	SG	OTCOP	ocCont (LUCAS)	SCSS	OCMRM
Spatial domain			(Global)	(Global)	(Europe)	(Europe)	(Spain)	(Murcia)
National (Spain)	SOCs (GgC)	Sum	$3.19{\cdot}10^6$	$5.07{\cdot}10^6$	$3.50{\cdot}10^6$	$1.89{\cdot}10^6$	$2.82{\cdot}10^6$	–
		Mean ± SD	0.06 ± 0.02	0.10 ± 0.03	0.07	0.04 ± 0.03	0.06 ± 0.04	–
	SOCc (g/kg)	Mean	–	26.62	22.71	24.38	25.89	–
		SD	–	17.94	23.69	16.62	20.47	–
Local (Murcia)	SOCs (GgC)	Sum	52,689.62	73,364.37	–	39,983.88	44,421.16	–
		Mean ± SD	0.05 ± 0.01	0.07 ± 0.01	–	0.04 ± 0.02	0.04 ± 0.01	–
	SOCc (g/kg)	Mean	–	13.78	11.6	18.88	19.51	14.07
		SD	–	5.87	7.15	7.85	6.06	14.88

* The available SOC products in the area refer to: Global soil organic carbon (GSOC), SoilGrids250m (SG), Organic Carbon Content in Topsoils (OTCOP), Topsoil Soil Organic Carbon Content (ocCont), Soil Organic Carbon Stock in Spain (SCSS), Soil Organic Carbon Map in Region of Murcia (OCMRM).

similar for all tested spatial resolutions. This is probably related to the complex challenge of estimating SOC across highly heterogeneous semiarid regions with complex terrain (Hoffmann et al., 2014; Hounkpatin et al., 2018; Jobbágy et al., 2000; Kulmatiski et al., 2004; Kunkel et al., 2011). Calvo de Anta et al. (2020) also revealed this heterogeneity in a national-level study where the semiarid Region of Murcia showed the highest coefficient of variation in both SOCc (76%) and SOCs (61%) at 0–30 cm depth.

Our results showed that the SOCc and SOCs maps at 100 m spatial resolution of the QRF model had the best balance among accuracy,

Table 4
External validation of the available SOC products (SOCs and SOCc) using synthetic profile values (255 soil samples) calculated from the LUCDEME local database (MurDB). The available SOC products in the area refer to Global soil organic carbon (GSOC), SoilGrids250m (SG), Organic Carbon Content in Topsoils (OTCOP), Organic Carbon Stock in Spain (SCSS), and Soil Organic Carbon Map in Region of Murcia (OCMRM).

AVAILABLE PRODUCTS	SOCs (tC/ha)		SOCc (g/kg)	
	R^2	RMSE (GgC)	R^2	RMSE (GgC)
MurDB vs. SG (Global)	0.055	25.73	0.211	11.316
MurDB vs. GSOC (Global)	$0.004 \cdot 10^{-1}$	20.88	–	–
MurDB vs. OCTOP (Europe)	–	–	0.067	12.74
MurDB vs. SCSS (Spain)	0.033	21.86	0.017	14.90
MurDB vs. OCMRM (Mur)	–	–	0.525	14.740

Table 5
Predicted values and associated uncertainty of Soil Organic Carbon (SOC) stock and the SOC concentration estimated by quantile regression forest (QRF = QuantregForest R package; QRF_G = Gstat R package) at different spatial resolutions for the Region of Murcia, Spain.

QRF (QuantregForest)					
			Spatial Resolution		
Variable			100 m	250 m	1000 m
SOC stock (GgC)	Predicted values	Sum	27,572	27,646	28,071
	Uncertainty	Mean	0.02	0.02	0.02
SOC concentration (g/kg)	Predicted values	Mean	12.18	12.89	10.28
	Uncertainty	Mean	10.54	9.99	8.87

QRF_G (GSIF)					
			Spatial Resolution		
Variable			100 m	250 m	1000 m
SOC stock (GgC)	Predicted values	Sum	28,911	28,435	28,937
	Uncertainty	Mean	0.03	0.03	0.03
SOC concentration (g/kg)	Predicted values	Mean	11.20	10.68	10.81
	Uncertainty	Mean	9.46	8.81	8.79

Table 6
Accuracy percentage (the interval corresponding to predicted values and their associated uncertainty, expressed as a decimal) of qrf models estimates (QRF = QuantregForest R package; QRF_G = GSIF R package). Local values of SOCc of LUCDEME topsoil database were compared with the interval corresponding to predicted values and their associated uncertainty.

MODEL	Spatial Resolution		
	100 m	250 m	1000 m
QRF	0.787	0.722	0.692
QRF_G	0.490	0.438	0.445

external validation, and interpretability of results. The largest spatial disagreement between SOCc and SOCs maps at 100 m spatial resolution was in areas with low soil sampling density, where SOCc maps depicted a spatial variability coherent with land use and landscape patterns (Albaladejo et al., 2009). These areas represent the highest altitude for forested regions (>1400 m) in the northwestern and central zones, and have an upper horizon rich in organic matter with natural grassland and tree forest over Lithosols that, although shallow (≤ 10 cm), are covered by abundant vegetation. The eastern area of the Region of Murcia also showed disagreements, where the main vegetation cover type was sclerophyllous vegetation and rainfed crops over regosols, with scarce incorporation of organic matter. The highest uncertainty in SOCc and SOCs was associated with low values corresponding to areas with an advanced process of soil degradation. Therefore, our results highlight that predicting SOC across degraded soils is challenging because processes associated with environmental degradation may alter the expected relationships between the training data and environmental covariates, resulting in higher bias and prediction errors (Brungard et al., 2015; Stell et al., 2021). To better capture the representativeness of SOC spatial variability, stratified sampling of homogeneous sub-areas (Zeraatpisheh et al., 2019) or optimization of future soil surveys (Smith et al., 2022) could reduce model uncertainty.

5. Conclusions

In this study, we tested six available SOC products derived from local, regional, and global approaches, and provided a locally derived map for a Mediterranean area in the southeastern Iberian Peninsula. The available SOC products with different spatial resolutions showed large differences among their values regardless of the spatial domain (CV = 0.35 and CV = 0.28 at the national and local domains, respectively). We observed a lower accuracy (R^2 = 0.06, RMSE = 25.73 GgC for external validation) and an overestimation (ranging from 44% to 164% over the estimates of the generated local maps) in SOCs predictions compared with SOCc predictions. These differences are likely due to missing information on key soil parameters in the databases (e.g., bulk density and coarse fragments). We observed that SOCc predictions are less sensitive to these missing key soil parameters; therefore, these model predictions may be more relevant to inform the environmental policies of soil carbon management.

Our high-resolution SOC map framework for the generated local prediction was based on local legacy soil data, environmental covariates (including single and/or multitemporal remote sensing indices), DMS modeling, and spatially explicit uncertainty quantification. This latter aspect and independent external validation are essential to interpret the soil carbon property distribution, especially in SOC complex quantification areas. Our results show the potential to improve the representation of national domain SOC estimates, especially in Mediterranean areas. This is important to respond to the challenges of land management and climate change adaptation/mitigation policies and strategies.

Funding

PD was supported by a pre-doctoral grant [DI-15-08093] awarded by the 'National Programme for the Promotion of Talent and Its Employability' of the Ministry of Economy, Industry, and Competitiveness, which are partially funded by the European Social Fund (ESF) from the European Commission. RV was supported by NASA Carbon Monitoring System grant 80NSSC21K0964.

CRediT authorship contribution statement

Pilar Durante: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **Mario Guevara:** Conceptualization, Methodology. **Rodrigo Vargas:** Conceptualization, Methodology, Resources, Supervision, Writing – review & editing. **Cecilio Oyonarte:** Conceptualization, Methodology, Resources, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Rodrigo Vargas reports financial support was provided by NASA. Pilar Duarte reports financial support was provided by National

Programme for the Promotion of Talent and Its Employability. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

We have included the link where the data can be downloaded

Acknowledgments

The authors are grateful to the Andalusian Scientific Computing Centre (CICA) for the access to computational facilities, the Spanish National Geographic Institute (IGN) as the provider of LiDAR data used in this study, and Arantzazu Blanco Bernardeau for providing the OCMRM map.

M.G. acknowledges support from grant UNESCO-IGCP-IUGS, 2022 (no. 765).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geodrs.2024.e00780.

References

Albaladejo, J., Martinez-Mena, M., Almagro, M., Ruiz-Navarro, A., Ortiz, R., Albaladejo Montoro, J., Martínez-Mena García, M., Almagro Costa, M., Ruiz-Navarro, A., Ortiz Silla, R., 2009. Factores de control en la dinámica del Carbono Orgánico de los suelos de la Región de Murcia, Advances in Studies on Desertification, Murcia (Spain). Murcia.

Alias, L., Ortiz, R., 1986. Memorias y mapas de suelos de las hojas del MTN a escala 1: 50.000. Proyecto LUCDEME. Ministerio de Medio Ambiente.

Arenas-Castro, S., Regos, A., Gonçalves, J.F., Alcaraz-Segura, D., Honrado, J., 2019. Remotely sensed variables of ecosystem functioning support robust predictions of abundance patterns for rare species. Remote Sens. 11, 2086. https://doi.org/10.3390/rs11182086.

Arrouays, D., McKenzie, N., Hempel, J., Richer de Forges, A.C., McBratney, A., 2014. Basis of the Global Spatial Soil Information System, in: GlobalSoilMap Basis of the Global Spatial Soil Information System. Taylor & Francis Group, London, p. 868.

Arrouays, D., McBratney, A., Bouma, J., Libohova, Z., Richer-de-Forges, A.C., Morgan, C. L.S., Roudier, P., Poggio, L., Mulder, V.L., 2020. Impressions of digital soil maps: the good, the not so good, and making them ever better. Geoderma Reg. 20, e00255 https://doi.org/10.1016/j.geodrs.2020.e00255.

Barahona, E., Santos, F., 1981. Estudios de correlación y regresión de diversos parámetros analíticos de 52 perfiles de suelos del sector Montiel-Alcaraz-Bienservida (Ciudad Real-Albacete). An Edafol Agrobiol. 40, 761–773.

Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modeling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91, 27–45. https://doi.org/10.1016/S0016-7061(99)00003-8.

Bivand, R.S., 2021. Progress in the R ecosystem for representing and handling spatial data. J. Geogr. Syst. 23, 515–546. https://doi.org/10.1007/s10109-020-00336-0.

Bivand, R., Rundel, C., 2020. rgeos: interface to geometry engine—open source ('GEOS'). R Package Version 0.5-3. https://CRAN.R-project.org/package=rgeos.

Blanco, A., 2015. Estudio de la Distribución Espacial y Cartografía Digital de Algunas Propiedades Físicas (Químicas e Hidrodinámicas de Suelos de la Cuenca del Segura).

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

de Brogniez, D., Ballabio, C., Stevens, A., Jones, R.J.A.A., Montanarella, L., van Wesemael, B., 2015. A map of the topsoil organic carbon content of Europe generated by a generalized additive model. Eur. J. Soil Sci. 66, 121–134. https://doi.org/10.1111/ejss.12193.

Brungard, C.W., Boettinger, J.L., Duniway, M.C., Wills, S.A., Edwards, T.C., 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239, 68–83. https://doi.org/10.1016/j.geoderma.2014.09.019.

Brunsdon, C., Chen, H., 2014. GISTools: some further GIS capabilities for R. R Packag Version 0.7-4. https://CRAN.R-project.org/package=GISTools.

Brus, D.J., Kempen, B., Heuvelink, G.B.M., 2011. Sampling for validation of digital soil maps. Eur. J. Soil Sci. 62, 394–407. https://doi.org/10.1111/j.1365-2389.2011.01364.x.

Calvo de Anta, R., Luís, E., Febrero-Bande, M., Galiñanes, J., Macías, F., Ortíz, R., Casás, F., 2020. Soil organic carbon in peninsular Spain: influence of environmental factors and spatial distribution. Geoderma 370. https://doi.org/10.1016/j.geoderma.2020.114365.

Carslaw, D.C., Ropkins, K., 2012. Openair - an R package for air quality data analysis. Environ. Model Softw. 27–28, 52–61. https://doi.org/10.1016/j.envsoft.2011.09.008.

Cash, D.W., Moser, S.C.S.C., 2000. Linking global and local scales: designing dynamic assessment and management processes. Glob. Environ. Chang. 10, 109–120. https://doi.org/10.1016/S0959-3780(00)00017-0.

Chambers, J., Hastie, T., Pregibon, D., 1990. Statistical models in S. In: Compstat. Physica-Verlag HD, Heidelberg, pp. 317–321. https://doi.org/10.1007/978-3-642-50096-1_48.

Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J., Meersmans, J., Richer-de-Forges, A. C., Walter, C., 2022. Digital mapping of GlobalSoilMap soil properties at a broad scale: a review. Geoderma 409, 115567. https://doi.org/10.1016/j.geoderma.2021.115567.

Conant, R.T., Ryan, M.G., Ågren, G.I., Birge, H.E., Davidson, E.A., Eliasson, P.E., Evans, S.E., Frey, S.D., Giardina, C.P., Hopkins, F.M., Hyvönen, R., Kirschbaum, M.U. F., Lavallee, J.M., Leifeld, J., Parton, W.J., Megan Steinweg, J., Wallenstein, M.D., Martin Wetterstedt, J.Å., Bradford, M.A., 2011. Temperature and soil organic matter decomposition rates - synthesis of current knowledge and a way forward. Glob. Chang. Biol. 17, 3392–3404. https://doi.org/10.1111/j.1365-2486.2011.02496.x.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (SAGA) v. 2.1.4. Geosci. Model Dev. 8, 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015.

Copenhaver, M.D., Holland, B., 1988. Computation of the distribution of the maximum studentized range statistic with application to multiple significance testing of simple effects. J. Stat. Comput. Simul. 30, 1–15. https://doi.org/10.1080/00949658808811082.

Dobos, E., Bialkó, T., Micheli, E., Kobza, J., 2010. Legacy soil data harmonization and database development. In: Digital Soil Mapping. Springer, Netherlands, Dordrecht, pp. 309–323. https://doi.org/10.1007/978-90-481-8863-5_25.

Durante, P., Martín-Alcón, S., Gil-Tena, A., Algeet, N., Tomé, J., Recuero, L., Palacios-Orueta, A., Oyonarte, C., 2019. Improving aboveground forest biomass maps: from high-resolution to national scale. Remote Sens. 11, 795. https://doi.org/10.3390/rs11070795.

Durante, P., Guevara, M., Vargas, R., Algeet, N., Oyonarte, C., 2020. Uncertainties in estimating the soil carbon sequestration service. In: EGU General Assembly Conference Abstracts, p. 18408.

Fathololoumi, S., Vaezi, A.R., Alavipanah, S.K., Ghorbani, A., Saurette, D., Biswas, A., 2020. Improved digital soil mapping with multitemporal remotely sensed satellite data fusion: a case study in Iran. Sci. Total Environ. 721, 137703 https://doi.org/10.1016/j.scitotenv.2020.137703.

Filippi, P., Cattle, S.R., Pringle, M.J., Bishop, T.F.A., 2021. Space-time monitoring of soil organic carbon content across a semi-arid region of Australia. Geoderma Reg. 24, e00367 https://doi.org/10.1016/j.geodrs.2021.e00367.

Guevara, M., Olmedo, G.F., Stell, E., Yigini, Y., Aguilar Duarte, Y., Arellano Hernández, C., Arévalo, G.E., Arroyo-Cruz, C.E., Bolivar, A., Bunning, S., Bustamante Cañas, N., Cruz-Gaistardo, C.O., Davila, F., Dell Acqua, M., Encina, A., Figueredo Tacona, H., Fontes, F., Hernández Herrera, J.A., Ibelles Navarro, A.R., Loayza, V., Manueles, A.M., Mendoza Jara, F., Olivera, C., Osorio Hermosilla, R., Pereira, G., Prieto, P., Ramos, I.A., Rey Brina, J.C., Rivera, R., Rodríguez-Rodríguez, J., Roopnarine, R., Rosales Ibarra, A., Rosales Riveiro, K.A., Schulz, G.A., Spence, A., Vasques, G.M., Vargas, R.R., Vargas, R., 2018. No silver bullet for digital soil mapping: country-specific soil organic carbon estimates across Latin America. Soil 4, 173–193. https://doi.org/10.5194/soil-4-173-2018.

Han, S.Y., Filippi, P., Singh, K., Whelan, B.M., Bishop, T.F.A., 2022. Assessment of global, national and regional-level digital soil mapping products at different spatial supports. Eur. J. Soil Sci. 73 https://doi.org/10.1111/ejss.13300.

Harden, Jennifer W., Hugelius, Gustaf, Ahlstr€om, Anders, Blankinship, Joseph C., Bond-Lamberty, Ben, Lawrence, Corey R., Loisel, Julie, Malhotra, Avni, Jackson, Robert B., Ogle, Stephen, Phillips, Claire, Ryals, Rebecca, Todd-Brown, Katherine, Vargas, Rodrigo, Vergara, Sintana E., Cotrufo, M. Francesca, Keiluweit, Marco, Heckman, Katherine A., Crow, Susan E., Silver, Whendee L., DeLonge, Marcia, Nave, Lucas E., 2017. Networking our science to characterize the state, vulnerabilities, and management opportunities of soil organic matter. Global Change Biology 24 (2), e705–e718.

Hartemink, A.E., 2006. The Future of Soil Science. International Union of Soil Sciences, Wageningen.

Heiberger, R., Holland, B., Azuaje, F., 2005. Statistical Analysis and Data Display: An Intermediate Course with Examples in S-PLUS, R, and SAS. BioMedical Engineering OnLine. https://doi.org/10.1186/1475-925X-4-18.

Hendriks, C.M.J., Stoorvogel, J.J., Lutz, F., Claessens, L., 2019. When can legacy soil data be used, and when should new data be collected instead? Geoderma 348, 181–188.

Hengl, T., MacMillan, R.A., 2019. Predictive Soil Mapping with R. Lulu.com. https://soilmapper.org/. GSIF: Global Soil Information Facilities. https://CRAN.R-project.org/package=GSIF.

Hengl, T., Mendes de Jesus, J., 2016. Understanding World Soils: Machine Learning as a Framework for Analyzing Global Soil-Landscape Relationships. ISRIC - World Soil Information, Wageningen.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: global gridded soil information based on machine learning. PLoS One 12, e0169748. https://doi.org/10.1371/journal.pone.0169748.

Heuvelink, G.B.M., 1998. Uncertainty analysis in environmental modelling under a change of spatial scale. In: Soil and Water Quality at Different Scales. Springer, Netherlands, Dordrecht, pp. 255–264. https://doi.org/10.1007/978-94-017-3021-1_24.

Hiederer, R., Köchy, M., 2012. Global soil organic carbon estimates and the harmonized world soil database. JRC Scientific and Technical Reports. EUR Scientific and Technical Research series. ISSN 1831-9424 (online), ISSN 1018-5593 (print), ISBN 978-92-79-23108-7. https://doi.org/10.2788/1326.

Hijmans, R.J., 2021. Raster: geographic data analysis and modelling. R package version 3.4-13. https://CRAN.R-project.org/package=raster.

Hoffmann, U., Hoffmann, T., Johnson, E.A.A., Kuhn, N.J., 2014. Assessment of variability and uncertainty of soil organic carbon in a mountainous boreal forest (Canadian Rocky Mountains, Alberta). Catena 113, 107–121. https://doi.org/10.1016/j.catena.2013.09.009.

Hounkpatin, O.K.L., Op de Hipt, F., Bossa, A.Y., Welp, G., Amelung, W., 2018. Soil organic carbon stocks and their determining factors in the Dano catchment (Southwest Burkina Faso). Catena 166, 298–309. https://doi.org/10.1016/j.catena.2018.04.013.

IGN, 2012. IGN-CORINE Land Cover (España) [WWW Document]. URL. https://www.idee.es:80/csw-inspire-idee/static/api/records/spaignCLC2012.

Jalabert, S.S.M., Martin, M.P., Renaud, J.P., Boulonne, L., Jolivet, C., Montanarella, L., Arrouays, D., 2010. Estimating forest soil bulk density using boosted regression modelling. Soil Use Manag. 26, 516–528. https://doi.org/10.1111/j.1475-2743.2010.00305.x.

Jobbágy, E.G., Jackson, R.B., Processes, B., Change, G., 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecol. Appl. 10, 423–436. https://doi.org/10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2.

Jones, R.J.A., Hiederer, R., Rusco, E., Loveland, P.J., Montanarella, L., 2004. The map of organic carbon in topsoils in Europe. Eur. J. Soil Sci. 56, 655–671.

Kibblewhite, M.G., Jones, R.J.A., Montanarella, L., Baritz, R., Huber, S., Arrouays, D., Micheli, E., Stephens, M., 2008. Environmental assessment of soil for monitoring volume VI: soil monitoring system for Europe. JRC Sci. Tech. Rep. https://doi.org/10.2788/95007.

Kravchenko, A., Bullock, D.G., 1999. A comparative study of interpolation methods for mapping soil properties. Agron. J. 91, 393–400. https://doi.org/10.2134/agronj1999.00021962009100030007x.

Kuhn, M., 2019. Caret: classification and Regression Training. R package version v6.0–88. https://CRAN.Rproject.org/web/packages/caret/index.html.

Kulmatiski, A., Vogt, D.J., Siccama, T.G., Tilley, J.P., Kolesinskas, K., Wickwire, T.W., Larson, B.C., 2004. Landscape determinants of soil carbon and nitrogen storage in southern New England. Soil Sci. Soc. Am. J. 68, 2014–2022. https://doi.org/10.2136/sssaj2004.2014.

Kunkel, M.L., Flores, A.N., Smith, T.J., McNamara, J.P., Benner, S.G., 2011. A simplified approach for estimating soil carbon and nitrogen stocks in semi-arid complex terrain. Geoderma 165, 1–11. https://doi.org/10.1016/j.geoderma.2011.06.011.

Lark, R.M., 2006. Chapter 23 Decomposing digital soil information by spatial scale. Dev. Soil Sci. https://doi.org/10.1016/S0166-2481(06)31023-9.

Lemercier, B., Lagacherie, P., Amelin, J., Sauter, J., Pichelin, P., Richer-de-Forges, A.C., Arrouays, D., 2022. Multiscale evaluations of global, national and regional digital soil mapping products in France. Geoderma 425, 116052. https://doi.org/10.1016/j.geoderma.2022.116052.

Liang, P., Qin, C.-Z., Zhu, A.-X., Hou, Z.-W., Fan, N.-Q., Wang, Y.-J., Peng, L., Cheng Zhi, Q., Xing, Z.A., Zhi Wei, H., Nai Qing, F., Yi Jie, W., 2020. A case-based method of selecting covariates for digital soil mapping. J. Integr. Agric. 19, 2127–2136. https://doi.org/10.1016/S2095-3119(19)62857-1.

Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2, 18–22.

Lombardo, L., Saia, S., Schillaci, C., Mai, P.M., Huser, R., 2018. Modeling soil organic carbon with quantile regression: dissecting predictors' effects on carbon stocks. Geoderma 318, 148–159. https://doi.org/10.1016/j.geoderma.2017.12.011.

Malone, B.P., McBratney, A.B., Minasny, B., Laslett, G.M., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. Geoderma 154, 138–152. https://doi.org/10.1016/j.geoderma.2009.10.007.

Manning, P., de Vries, F.T., Tallowin, J.R.B., Smith, R., Mortimer, S.R., Pilgrim, E.S., Harrison, K.A., Wright, D.G., Quirk, H., Benson, J., Shipley, B., Cornelissen, J.H.C., Kattge, J., Bönisch, G., Wirth, C., Bardgett, R.D., 2015. Simple measures of climate, soil properties and plant traits predict national-scale grassland soil carbon stocks. J. Appl. Ecol. 52, 1188–1196. https://doi.org/10.1111/1365-2664.12478.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4.

Meinshausen, N., 2006. Quantile regression forests. J. Mach. Learn. Res. 7, 983–999.

Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.-C., Vågen, T.-G., van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86. https://doi.org/10.1016/j.geoderma.2017.01.002.

Moni, C., Rumpel, C., Virto, I., Chabbi, A., Chenu, C., 2010. Relative importance of sorption versus aggregation for organic matter storage in subsoil horizons of two contrasting soils. Eur. J. Soil Sci. 61, 958–969. https://doi.org/10.1111/j.1365-2389.2010.01307.x.

Montanarella, L., 2015. Agricultural policy: govern our soils. Nature 528, 32–33. https://doi.org/10.1038/528032a.

Ninyerola, M., Pons, X., Roure, J., 2005. Atlas climático digital de la Península Ibérica: metodología y aplicaciones en bioclimatología y geobotánica. Barcelona.

Omran, E.-S.E., 2012. Improving the prediction accuracy of soil mapping through Geostatistics. Int. J. Geosci. 03, 574–590. https://doi.org/10.4236/ijg.2012.33058.

Pásztor, L., Laborczi, A., Szatmári, G., Koós, S., Bakacsi, Z., Makó, A., Tóth, B., 2019. Digital soil maps for the support of national mapping and assessment of ecosystem services. Geophys. Res. Abstr. 21, 5645.

Pecl, G.T., Araújo, M.B., Bell, J.D., Blanchard, J., Bonebrake, T.C., Chen, I.-C., Clark, T. D., Colwell, R.K., Danielsen, F., Evengård, B., Falconi, L., Ferrier, S., Frusher, S., Garcia, R.A., Griffis, R.B., Hobday, A.J., Janion-Scheepers, C., Jarzyna, M.A., Jennings, S., Lenoir, J., Linnetved, H.I., Martin, V.Y., McCormack, P.C., McDonald, J., Mitchell, N.J., Mustonen, T., Pandolfi, J.M., Pettorelli, N., Popova, E., Robinson, S.A., Scheffers, B.R., Shaw, J.D., Sorte, C.J.B., Strugnell, J.M., Sunday, J. M., Tuanmu, M.-N., Vergés, A., Villanueva, C., Wernberg, T., Wapstra, E., Williams, S.E., 2017. Biodiversity redistribution under climate change: impacts on ecosystems and human well-being. Science 355. https://doi.org/10.1126/science.aai9214 eaai9214.

Poeplau, C., Vos, C., Don, A., 2017. Soil organic carbon stocks are systematically overestimated by misuse of the parameters bulk density and rock fragment content. SOIL 3, 61–66. https://doi.org/10.5194/soil-3-61-2017.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Richer-de-Forges, A.C., Arrouays, D., Bardy, M., Bispo, A., Lagacherie, P., Laroche, B., Lemercier, B., Sauter, J., Voltz, M., 2019. Mapping of soils and land-related environmental attributes in France: analysis of end-users' needs. Sustainability 11, 2940. https://doi.org/10.3390/su11102940.

Robinson, T.P., Metternicht, G., 2006. Testing the performance of spatial interpolation techniques for mapping soil properties. Comput. Electron. Agric. 50, 97–108. https://doi.org/10.1016/j.compag.2005.07.003.

Rodríguez Martín, J.A., Álvaro-Fuentes, J., Gonzalo, J., Gil, C., Ramos-Miras, J.J., Grau Corbí, J.M., Boluda, R., 2016. Assessment of the soil organic carbon stock in Spain. Geoderma 264, 117–125. https://doi.org/10.1016/j.geoderma.2015.10.010.

Rossel, R.A.V., Behrens, T., 2010. Using data mining to model and interpret soil diffuse reflectance spectra. Geoderma 158, 46–54. https://doi.org/10.1016/j.geoderma.2009.12.025.

Savin, I.Y., Zhogolev, A.V., Prudnikova, E.Y., 2019. Modern trends and problems of soil mapping. Eurasian Soil Sci. 52, 471–480. https://doi.org/10.1134/S1064229319050107.

Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappiè, M., Märker, M., Saia, S., 2017. Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: the role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. Sci. Total Environ. 601–602, 821–832. https://doi.org/10.1016/j.scitotenv.2017.05.239.

Searle, R., McBratney, A., Grundy, M., Kidd, D., Malone, B., Arrouays, D., Stockman, U., Zund, P., Wilson, P., Wilford, J., Van Gool, D., Triantafilis, J., Thomas, M., Stower, L., Slater, B., Robinson, N., Ringrose-Voase, A., Padarian, J., Payne, J., Orton, T., Odgers, N., O'Brien, L., Minasny, B., Bennett, J.M., Liddicoat, C., Jones, E., Holmes, K., Harms, B., Gray, J., Bui, E., Andrews, K., 2021. Digital soil mapping and assessment for Australia and beyond: a propitious future. Geoderma Reg. 24, e00359 https://doi.org/10.1016/j.geodrs.2021.e00359.

Shangguan, W., Hengl, T., Mendes de Jesus, J., Yuan, H., Dai, Y., 2017. Mapping the global depth to bedrock for land surface modeling. J. Adv. Model. Earth Syst. 9, 65–88. https://doi.org/10.1002/2016MS000686.

Silatsa, F.B.T., Yemefack, M., Tabi, F.O., Heuvelink, G.B.M., Leenaars, J.G.B., 2020. Assessing countrywide soil organic carbon stock using hybrid machine learning modelling and legacy soil data in Cameroon. Geoderma 367. https://doi.org/10.1016/j.geoderma.2020.114260.

Smith, P., Soussana, J.F., Angers, D., Schipper, L., Chenu, C., Rasse, D.P., Batjes, N.H., Egmond, F., McNeill, S., Kuhnert, M., Arias-Navarro, C., Olesen, J.E., Chirinda, N., Fornara, D., Wollenberg, E., Álvaro-Fuentes, J., Sanz-Cobena, A., Klumpp, K., van Egmond, F., McNeill, S., Kuhnert, M., Arias-Navarro, C., Olesen, J.E., Chirinda, N., Fornara, D., Wollenberg, E., Álvaro-Fuentes, J., Sanz-Cobena, A., Klumpp, K., 2020. How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. Glob. Chang. Biol. 26, 219–241. https://doi.org/10.1111/gcb.14815.

Smith, E.M., Vargas, R., Guevara, M., Tarin, T., Pouyat, R.V., 2022. Spatial variability and uncertainty of soil nitrogen across the conterminous United States at different depths. Ecosphere 13. https://doi.org/10.1002/ecs2.4170.

Stell, E., Warner, D., Jian, J., Bond-Lamberty, B., Vargas, R., 2021. Spatial biases of information influence global estimates of soil respiration: how can we improve global predictions? Glob. Chang. Biol. 27, 3923–3938. https://doi.org/10.1111/gcb.15666.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmos. 106, 7183–7192. https://doi.org/10.1029/2000JD900719.

Thompson, J.A., Pena-Yewtukhiw, E.M., Grove, J.H., 2006. Soil-landscape modeling across a physiographic region: topographic patterns and model transportability. Geoderma 133, 57–70. https://doi.org/10.1016/j.geoderma.2006.03.037.

Tobler, W.R., 1988. Resolution, resampling, and all that. In: Mounsey, H., Tomlinson, R. F. (Eds.), Building Databases for Global Science: The Proceedings of the First Meeting of the International Geographical Union Global Database Planning Project. Taylor and Francis, Hampshire, U.K., pp. 129–137

Trnka, M., Olensen, J.E., Kersebaum, K.C., Skjelvag, A.O., Eitzinger, J., Seguin, B., Peltonen-Sainio, P., Rötter, R., Iglesias, A., Orlandini, S., Dubrovský, M., Hlavinka, P., Balek, J., Eckersten, H., Cloppet, E., Calanca, P., Gobin, A., Vucetc, V., Nejedlik, P., Kumar, S., Lalic, B., Mestre, A., Rossi, F., Kozyra, J., Alexandrov, V., Semerádová, D., Žalud, Z., 2011. Agroclimatic conditions in Europe under climate change. Glob. Chang. Biol. 17, 2298–2318. https://doi.org/10.1111/j.1365-2486.2011.02396.x.

Vargas, R., Alcaraz-Segura, D., Birdsey, R., Brunsell, N.A., Cruz-Gaistardo, C.O., de Jong, B., Etchevers, J., Guevara, M., Hayes, D.J., Johnson, K., Loescher, H.W., Paz, F., Ryu, Y., Sanchez-Mejia, Z., Toledo-Gutierrez, K.P., 2017. Enhancing interoperability to facilitate implementation of REDD+: case study of Mexico. Carbon Manag. 8, 57–65. https://doi.org/10.1080/17583004.2017.1285177.

Vargas-Rojas, R., Cuevas-Corona, R., Yigini, Y., Tong, Y., Bazza, Z., Wiese, L., 2019. Unlocking the potential of soil organic carbon: a feasible way forward. In: International Yearbook of Soil Law and Policy, pp. 373–395. https://doi.org/10.1007/978-3-030-00758-4_18.

Villarreal, S., Guevara, M., Alcaraz-Segura, D., Brunsell, N.A., Hayes, D., Loescher, H.W., Vargas, R., 2018. Ecosystem functional diversity and the representativeness of environmental networks across the conterminous United States. Agric. For. Meteorol. 262, 423–433. https://doi.org/10.1016/j.agrformet.2018.07.016.

Vitharana, U.W.A., Mishra, U., Mapa, R.B., 2019. National soil organic carbon estimates can improve global estimates. Geoderma 337, 55–64. https://doi.org/10.1016/j.geoderma.2018.09.005.

Wang, B., Waters, C., Orgill, S., Gray, J., Cowie, A., Clark, A., Liu, D.L., 2018. High resolution mapping of soil organic carbon stocks using remote sensing variables in the semi-arid rangelands of eastern Australia. Sci. Total Environ. 630, 367–378. https://doi.org/10.1016/j.scitotenv.2018.02.204.

Webster, R., Oliver, M.A., 2001. Geostatistics for Environmental Scientists, second. ed. John Wiley & Sons, Ltd.

Weston, J., Watkins, C., 1999. Support vector machines for multi-class pattern recognition. In: Proc. 7th Eur. Symp. Artif. Neural Networks, pp. 219–224.

Wiesmeier, M., Urbanski, L., Hobley, E., Lang, B., von Lützow, M., Marin-Spiotta, E., van Wesemael, B., Rabot, E., Ließ, M., Garcia-Franco, N., Wollschläger, U., Vogel, H.-J., Kögel-Knabner, I., 2019. Soil organic carbon storage as a key function of soils - a review of drivers and indicators at various scales. Geoderma 333, 149–162. https://doi.org/10.1016/j.geoderma.2018.07.026.

Willaarts, B.A., Oyonarte, C., Muñoz-Rojas, M., Ibáñez, J.J., Aguilera, P.A., 2016. Environmental factors controlling soil organic carbon stocks in two contrasting Mediterranean climatic areas of southern Spain. L Degrad. Dev. 27, 603–611. https://doi.org/10.1002/ldr.2417.

WRB-IUSS, 2014. World Reference Base for Soil Resources. 2014, International Soil Classification System for Naming Soils and Creating Legends for Soil Maps. FAO, Rome, Italy.

Xiong, X., Grunwald, S., Corstanje, R., Yu, C., Bliznyuk, N., 2016. Scale-dependent variability of soil organic carbon coupled to land use and land cover. Soil Tillage Res. 160, 101–109. https://doi.org/10.1016/j.still.2016.03.001.

Yigini, Y., Olmedo, G.F., Reiter, S., Baritz, R., Viatkin, K., Vargas, R., 2018. Soil Organic Carbon Mapping: Cookbook, 2nd edition. FAO, Rome.

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.

Zhou, T., Geng, Y., Lv, W., Xiao, S., Zhang, P., Xu, X., Chen, J., Wu, Z., Pan, J., Si, B., Lausch, A., 2023. Effects of optical and radar satellite observations within Google earth engine on soil organic carbon prediction models in Spain. J. Environ. Manag. 338, 117810 https://doi.org/10.1016/j.jenvman.2023.117810.