This is the preprint of the contribution published as:

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. Manage. **338** , art. 117810

The publisher’s version is available at:

http://dx.doi.org/10.1016/j.jenvman.2023.117810

Title page

Effects of optical and radar satellite observations within Google Earth Engine on soil organic carbon prediction models in Spain

Tao Zhou[a, b, c], Yajun Geng[a], Wenhao Lv[a],Shancai Xiao[d], Peiyu Zhang[e],Xiangrui Xu[f], Jie Chen[g], Zhen Wu[h], Jianjun Pan[h], Bingcheng Si[a, *] and Angela Lausch[b, c]

a Ludong University, School of Resources and Environmental Engineering, Middle Hongqi Road 186, 264025 Yantai, China

b Humboldt-Universität zu Berlin, Department of Geography, Unter den Linden 6, 10099 Berlin, Germany

c Helmholtz Centre for Environmental Research, Department of Computational Landscape Ecology, Permoserstraße 15, 04318 Leipzig, Germany

d Peking University, College of Urban and Environmental Sciences, Yiheyuan Road 5, 100871 Beijing, China

e Hunan Normal University, College of Geographical Sciences, Lushan Road 36, 410081 Changsha, China

f Zhejiang University City College, School of Spatial Planning and Design, Huzhou Street 51, 31000 Hangzhou, China

g Hunan Academy of Agricultural Sciences, Yuanda 2nd Road 560, 410125 Changsha, China

h Nanjing Agricultural University, College of Resources and Environmental Sciences, Weigang 1, 210095 Nanjing, China

* Corresponding author.

E-mail address: bing.si@usask.ca (B. Si).

Highlights

- The composite mode and selection of satellite imagery, as well as the radar data utilization strategies, influenced the prediction performances.
- The mixed information of different orbital directions and polarization modes effectively improved the mapping accuracy.
- The predictive performance of C-band Sentinel-1 versus multispectral Sentinel-2 was comparable, although the complementary use of the two provided higher accuracy.
- The accuracy rankings based on long-term optical and microwave sensor observations were as follows: Sentinel-3 > Sentinel-2 > Sentinel-1 > ALOS-2.
- The synergistic utilization of the Sentinel series gave better prediction models and more accurate prediction results.

Abstract

The modeling and mapping of soil organic carbon (SOC) has advanced through the rapid growth of Earth observation data (e.g., Sentinel) collection and the advent of appropriate tools such as the Google Earth Engine (GEE). However, the effects of differing optical and radar sensors on SOC prediction models remain uncertain. This research aims to investigate the effects of different optical and radar sensors (Sentinel-1/2/3 and ALOS-2) on SOC prediction models based on long-term satellite observations on the GEE platform. We also evaluate the relative impact of four synthetic aperture radar (SAR) acquisition configurations (polarization mode, band frequency, orbital direction and time window) on SOC mapping with multiband SAR data from Spain. Twelve experiments involving different satellite data configurations, combined with 4027 soil samples, were used for building SOC random forest regression models. The results show that the synthesis mode and choice of satellite images, as well as the SAR acquisition configurations, influenced the model accuracy to varying degrees. Models based on SAR data involving cross-polarization, multiple time periods and "ASCENDING" orbits outperformed those involving copolarization, a single time period and "DESCENDING" orbits. Moreover, combining information from different orbital directions and polarization modes improved the soil prediction models. Among the SOC models based on long-term satellite observations, the Sentinel-3-based models (R^2 = 0.40) performed the best, while the ALOS-2-based model performed the worst. In addition, the predictive performance of MSI/Sentinel-2 (R^2 = 0.34) was comparable with that of SAR/Sentinel-1 (R^2 = 0.33); however, the combination (R^2 > 0.39) of the two improved the

model performance. All the predicted maps involving Sentinel satellites had similar spatial patterns that were higher in northwest Spain and lower in the south, which is consistent with the land use. Overall, this study provides insights into the effects of different optical and radar sensors and radar system parameters on soil prediction models and improves our understanding of the potential of Sentinels in developing soil carbon mapping.

Keywords: Google Earth Engine; Multisensor; Sentinel; Soil organic carbon; Digital soil mapping; Synthetic aperture radar

1. Introduction

Soil organic carbon constitutes a large carbon pool in the biosphere and greatly influences the global carbon cycle (Caddeo et al., 2019; Rowley et al., 2018). Spatial information on SOC is essential for food production, environmental quality management, soil health monitoring and ecosystem health (Dharumarajan et al., 2021; Mallik et al., 2022; Zhang et al., 2022). There is a great demand for detailed, accurate and up-to-date spatial information on SOC that extends from the local to the national and global levels. Many countries have made significant efforts to compile their own national-level SOC maps (Calvo de Anta et al., 2020; Lamichhane et al., 2021; Liu et al., 2022a; Szatmári et al., 2021). However, conventional soil mapping with polygon-based methods is considered time-consuming, expensive and labor intensive, especially for large-scale soil mapping (Minasny and McBratney, 2016). Therefore, there is a need for an economical, reliable and suitable method for estimating the spatial distribution of SOC at large scales.

Digital soil mapping techniques, unlike traditional mapping approaches, are an effective method to characterize the spatial patterns of soil properties (He et al., 2021). Digital soil mapping mainly relies on establishing quantitative relationships between environmental variables (e.g., climate, topography, satellite imagery and parent material) and field soil observations, and then the relationship is used to make predictions (Baltensweiler et al., 2021; Wadoux et al., 2020). The accuracy of digital soil mapping techniques is strongly affected by the selected environmental variables; thus, developing influential environmental variables to generate accurate SOC maps is an important task (Yang et al., 2021a). At present, the amount and availability of environmental data is growing rapidly, especially for remote sensing products. The rapid growth in the number of Earth Observation (EO) satellites in orbit and the unprecedented improvements in computing power have created an enormous potential for improving soil mapping techniques.

Many sensors (i.e., multispectral, hyperspectral and radar) on board satellite platforms can acquire rich land surface information for the digital mapping of SOC (Guo et al., 2021; Sothe et al., 2022; Vaudour et al., 2021; Venter et al., 2021). These studies have tested the possibility of different satellite sensors and methods in predicting SOC and have validated their ability to monitor and map soil properties in different ecosystems. Currently, research on SOC mapping relies heavily on optical satellite sensors, and the commonly derived variables include surface reflectance and vegetation indices (Nguyen et al., 2022). However, optical satellite imagery is susceptible to cloud cover, which hinders its application for soil mapping. The synergy of optical imagery and synthetic aperture radar (SAR) data is considered an innovative solution. Compared with optical data, the application of SAR imagery in SOC mapping has not been fully explored and developed.

Several researchers have recently attempted to use SAR data to characterize the spatial pattern of SOC (Nguyen et al., 2022; Wang et al., 2020; Zhou et al., 2020b). These studies carried out SOC mappings at local, regional, or national scales, indicating the potential of SAR data in mapping SOC. However, the application of SAR data for SOC mapping is complicated by the various possible configurations of these data, such as with variations in polarization mode, band frequency, orbital direction and time window, especially since the configurations may affect the outcome of SOC predictions (Mahdianpari et al., 2017). The shorter wavelength C-band SAR system mainly interacts with the upper part of the canopy layer and has difficulty penetrating the vegetation canopy. In contrast, the longer-wavelength L-band SAR sensors can penetrate dense vegetation layers and better characterize vegetation structures (Li et al., 2019; Ottinger and Kuenzer, 2020). In addition to the band frequency, the interaction of the SAR signal with the surface also depends on other radar system parameters, such as polarization and incident angle (Prudente et al., 2022). In this context, it is important to study the impact of SAR data-acquisition configurations on their application in scientific fields (Hosseini et al., 2015; Mohammadimanesh et al., 2018; Purinton and Bookhagen, 2020; Rapinel et al., 2020). However, the relative impact of each acquisition configuration on SOC mapping remains to be addressed.

The advent of the era of big EO data is driving a significant shift in digital soil mapping (Tziolas et al., 2020). The recently launched Sentinel satellites (Sentinel-1/2/3) provide excellent observation capabilities for many different applications, including SOC mapping and environmental monitoring (Esch et al., 2018b; Song et al., 2020; Wang et al., 2022). Each Sentinel-1/2/3 mission consists of two satellites to ensure a high revisit frequency and rapid coverage of large areas. Sentinel-1 (6-day revisit) and Sentinel-2 (5-day revisit) provide C-band SAR data and multispectral imagery, respectively, while Sentinel-3 (with a revisit time of 1-2 days) uses multiple sensing instruments for land and ocean missions (Prikaziuk and van der Tol, 2019; Wang et al., 2019a). Sentinel-1/2/3 missions collect approximately 20 TB of data per day (Esch et al., 2018a), and such an unpredented revisit frequency and multiplication of sensing intensities will likely revolutionize soil mapping by advancing EO data into the era of big data. Moreover, the advent of cloud computing platforms such as Google Earth Engine (GEE) is leading to a shift in how EO data are processed, allowing users to process large-scale and intensive satellite time-series data (DeVries et al., 2020). Recently, the GEE platform has been shown to be a powerful tool for soil mapping, and it has been reported that soil prediction models based on multitemporal synthetic images provide more robust prediction results (Luo et al., 2022b; Luo et al., 2022c). However, most of the existing efforts are limited to specific optical sensors (e.g., Landsat-8 and Sentinel-2). Tamiminia et al. (2020) used a meta-analysis to review 349 peer-reviewed GEE articles, of which only 14 were on soil research-related topics. Based on this review, optical imagery (90% of studies), which is more accessible and familiar to users, was found to remain the most commonly used data source, especially Landsat (82%) and MODIS (6%) data, while only 9 studies utilized SAR data. The provision of analysis-ready SAR data within GEE represents an important step forward in applying SAR, as the complexity of SAR preprocessing has previously hindered its application; furthermore, the existence of other geospatial data sources, such as the global Sentinel archive, allows for relative ease in integrating the different EO data sources (DeVries et al., 2020). There is still a lack of comprehensive analysis supported by high-performance GEE computing platforms to evaluate and compare the potential and capabilities of the different optical and radar Sentinel sensors (Sentinel-1/2/3) in mapping SOC.

The focus of this study is to investigate the effects of different optical and radar Sentinel sensors on SOC prediction models based on long-term Sentinel-1/2/3 satellite observations. The specific aims are to: (1) evaluate the relative impact of four SAR acquisition configurations (i.e., polarization mode, band frequency, orbital direction and time window) on predicting SOC content and; (2) analyze and to compare the ability of optical and radar sensors (Sentinel-1/2/3 and ALOS-2), with different characteristics, in mapping SOC.

2. Materials and methods

2.1. Study area

The area for this study is located in mainland Spain (hereafter Spain), which covers nearly 500,000 km^2 (Fig. 1). The region is divided into three main climatic zones, oceanic, continental, and Mediterranean (Mulomba Mukadi and González-García, 2021; Rodríguez Martín et al., 2016). The Mediterranean climate dominates most of the study area, with an Atlantic climate in northern Spain and a semiarid climate in the southeastern part (Moreno-García et al., 2020). The mean annual precipitation varies widely from <200 mm (southeastern Spain) to >2200 mm (northeastern Spain); the annual mean temperature ranges from <2.5 °C in the high altitudes to >17 °C in the southern regions (Calvo de Anta et al., 2020; López-Senespleda et al., 2021). The terrain of Spain is highly heterogeneous, with elevations ranging from 0 to 3479 m; the average elevation is 600 m to 800 m (Moreno-García et al., 2020; Wang et al., 2020). Agricultural and forestland occupies 91% of the area, the latter is mainly on the northwest coast of Spain (Calvo de Anta et al., 2020; Wang et al., 2020). As one of the major agricultural producers in Europe, wheat and maize are the two most important crops in Spain (Allende-Montalbán et al., 2022; Morales-Polo et al., 2021). The main soil types in the different regions of Spain are described by Calvo de Anta et al. (2020).

2.2. Soil data

The soil dataset for this study came from the LUCAS project, a statistical survey initiated by Eurostat and conducted every three years in the European Union (EU) (Borrelli et al., 2022). The newly published LUCAS 2015 topsoil data covering Spain were used in this study. The LUCAS 2015 Survey was conducted in all EU-28 member states; a total of 27,069 locations were selected for sampling, of which 22,631 were collected in the EU-28 (Jones et al., 2020). In the LUCAS 2015 Survey, 90% of the soil sampling points in the LUCAS 2009/2012 Survey were repeated, with the remaining 10% displaced by new locations, including some points above 1,000 m in elevation (Fernández-Ugalde et al., 2020). These soil samples, representative of European landscape features, cover all land use types and were collected using a standard sampling scheme. Soil sample analysis was performed in an ISO-certified laboratory following standard procedures (Panagos et al., 2014; Zhong et al., 2021). The results of the analysis recorded the physicochemical properties (e.g., SOC) of the soil samples with details of the vegetation, soil type and altitude (Khan and Chiti, 2022). LUCAS topsoil is perhaps globally the largest, most comprehensive, and open-access topsoil dataset and has been used by many scholars for soil mapping at different scales (Ballabio et al., 2019; Castaldi et al., 2018; Zhou et al., 2021). We modeled the soil properties using soil samples (n = 4027) covering Spain based on the LUCAS database (Fig. 1).

2.3. Satellite data

2.3.1. Sentinel-1

Sentinel-1 is a C-band SAR system provided by the European Space Agency (ESA) since 2014; it makes acquisitions in four imaging modes with different spatial resolutions and swath widths (Zhao et al., 2022). The satellite is a two-satellite constellation of Earth-imaging satellites that collect data from both the descending and ascending orbits. Interferometric wide swath mode (IW) is the main operating mode over land, with a high spatial resolution (5 m × 20 m) and wide coverage (250 km) (Huang et al., 2018). We leveraged the GEE platform to access and process all the IW-mode Sentinel-1 ground range detected (GRD) images available in 2015 (Fig. 2). In this study, the Sentinel-1 data were filtered by orbital direction to obtain the images with "DESCENDING" and "ASCENDING" orbits. Sentinel-1 data were subjected to several preprocessing algorithms in the GEE platform to generate VV and VH backscatter coefficients, as described by Singha et al. (2020). The VV and VH polarizations of the Sentinel-1 data in different orbital directions were time-composited over 6 time periods at bimonthly intervals (e.g., period 1: January–February 2015); all Sentinel-1 images acquired during these time periods were used to calculate median and mean composite images for VV and VH polarizations in different composite modes (median and average). In total, we integrated 48 Sentinel-1 features—two bands (VV and VH) multiplied by six time periods, two orbital directions and two composite modes —to be considered in the next modeling step.

2.3.2. Sentinel-2

Sentinel-2, a two-satellite multispectral imaging mission, provides multispectral image (13 spectral bands) data with high spatial resolution (10–60 m) and wide area coverage (swath width of 290 km) at a 5-day interval (Murphy et al., 2016). High-level Sentinel-2 surface reflectance images have been available in GEE only since 2017. All available Sentinel-2 Top of Atmosphere (TOA) data with cloud cover less than 10% in 2015 were obtained using the GEE platform (Fig. 2). The QA60 band containing cloud information was used for cloud removal in the early stages of image preprocessing (Wang et al., 2019b; Yang et al., 2020). We synthesized all the Sentinel-2 images in two composite modes, resulting in a total of 20 feature synthesis results, 10 bands (i.e., bands 2-8a, 11, and 12) times two composite modes (median and average).

2.3.3. Sentinel-3

Sentinel-3 is the latest in the Sentinel family and consists of two satellites, Sentinel-3A, launched in February 2016, and Sentinel-3B, launched in April 2018 (Fernandez-Moran et al., 2021; Odebiri et al., 2022). The Sentinel-3 sensor (with a revisit time of 1–2 days) is equipped with scientific instruments, the Ocean and Land Color Imager (OLCI), Sea and Land Surface Temperature Radiometer (SLSTR), SAR Radar Altimeter (SRAL) and MicroWave Radiometer (MWR) (Prikaziuk and van der Tol, 2019; Wooster et al., 2012). The OLCI has 21 spectral bands with a spatial resolution of 300 m and a large swath width of 1270 km (Clevers and Gitelson, 2013; Kravitz et al., 2020). This study used the GEE platform to collect and process all the available Sentinel-3 OLCI TOA data for the study area in 2016 (Fig. 2). We filtered out the bright pixels, which mainly consist of clouds, snow and ice, using each image's quality flag band (Liu et al., 2022b). The median and mean of all the Sentinel-3 images were calculated to produce composite images consisting of 21 median and mean bands, respectively.

2.3.4. ALOS-2 PALSAR-2

ALOS-2 is a follow-up mission to ALOS that launched in May 2014 and is equipped with an enhanced L-band SAR sensor (PALSAR-2) operating in the 1215-1300 MHz frequency range (Rosenqvist et al., 2014). The Japan Aerospace Exploration Agency (JAXA) provides annual global 25 m ALOS-2 PALSAR-2 mosaic data by mosaicking SAR images of backscatter coefficients (Xu et al., 2020). PALSAR-2 data were preprocessed by JAXA; each mosaic contains HH and HV polarizations (Li et al., 2022). We acquired the 2015 ALOS-2 PALSAR-2 annual mosaic via GEE and converted the digital numbers (DNs) of the original HH and HV polarizations to backscatter gamma-naught (γ^0) values using Equation (1) (Yang et al., 2021b). There are two polarization models (HH and HV) in the composite PALSAR-2 imagery that are available for further analysis.

$$\gamma^0 = 10 \times log_{10}(DN^2) - 83 \qquad (1)$$

2.4. Random forest

Random forest (RF) is an ensemble tree-based algorithm for regression and classification tasks that combines the concepts of decision trees and bagging (Jia et al., 2021). RF relies on building decision trees on a training dataset, where each tree is planted using bootstrap samples from the training dataset (Nabiollahi et al., 2021). Bagging stands for bootstrapping and aggregation techniques and employs voting or averaging strategies to aggregate learners (Wadoux et al., 2019). The RF model is robust to overfitting and noise, has higher accuracy and works well with large datasets (Zhang et al., 2021). The algorithm has two important user-defined parameters that need to be optimized, the number of trees (ntree) and the number of randomly selected variables (mtry). We used the grid search method of the "caret" package in R software to optimize these parameters (Tiyasha et al., 2021).

2.5. Statistical analyses

A Pearson's correlation analysis between SOC and remote sensing-derived predictors was carried out, following the approaches of previous soil mapping studies (Gholizadeh et al., 2018; Nguyen et al., 2022; Yang and Guo, 2019). A descriptive statistical analysis of the soil properties was performed in the SPSS 21.0 software. The RF modeling process was performed in R software using the "randomForest" package.

2.6. Model evaluation

Twelve experiments (Table 1) were conducted to assess the suitability and potential of the different datasets and their combinations in mapping and predicting SOC at the national scale in Spain. We evaluated the performance of the soil prediction models based on each experiment using a 10-fold cross-validation procedure. Chen et al. (2022) reviewed studies on digital soil mapping at a broad scale and reported that the cross-validation techniques are the most common validation strategy for evaluating the accuracy of digital soil maps. Here, the data were randomly divided into ten folds; nine folds were used to fit the model, and the remaining fold was used for validation (Mello et al., 2022; Taghipour et al., 2022). This process was repeated ten times; then, the validation folds were aggregated together, and the quality of the predictions was evaluated by three accuracy metrics, the coefficient of determination (R^2), the root mean square error (RMSE) and the mean absolute error (MAE) (Equations (2)–(4)).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|P_i - O_i\right| \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^{n}(P_i - \overline{O}_i)^2}{\sum_{i=1}^{n}(O_i - \overline{O}_i)^2} \quad (4)$$

where n represents the number of samples and P_i and O_i represent the predicted and observed values at site i, respectively.

3. Results

3.1. Descriptive statistics and correlation analysis

A descriptive statistical analysis of the SOC content is shown in Table 2. The SOC content ranged from 0.10 to 406.10 g/kg, with a mean and median of 22.95 and 14.30 g/kg, respectively. The SOC data showed a strongly skewed distribution, yielding a skewness coefficient of 3.65. A natural logarithmic transformation reduced the skewness coefficient to 0.03. The SD values of the original and converted SOC were 25.42 g/kg and 0.90 g/kg, respectively.

We performed a Pearson correlation analysis and built correlograms between SOC and the quantitative predictors (Fig. 3). There were statistically significant correlations between all remote sensing-derived predictors and SOC ($P < 0.01$). Overall, higher correlations between the predictors and SOC were observed after log transformations. The highest correlations in the four remote sensing datasets were obtained from the spectral bands from B6 to B11 of Sentinel-3 (from -0.52 to -0.55); meanwhile, SOC had the weakest correlation with the VV polarization of Sentinel-1. Moreover, good correlations were also observed between SOC and the remaining bands of Sentinel-3, with r values above 0.30, with the exception of B20; in general, higher r-values were obtained from the Sentinel-3 median composite images in the two composite modes. Among the Sentinel-2 bands, B4 and B5 had the highest correlation with SOC (r > 0.50); the remaining Sentinel-2 bands also provided good correlations with SOC, with correlation coefficients above 0.25. For the two radar datasets used, the cross-polarizations (HV and VH) provided higher correlations with SOC than copolarizations (VV and HH); for example, for the Sentinel-1 derived predictors, SOC had a weak correlation with the VV polarizations (from 0.07 to 0.27) and a relatively strong correlation with the VH polarizations (from 0.32 to 0.44); the highest correlations for the VV and VH polarizations of Sentinel-1 came from VVa_2_median and VHa_2_mean, respectively (Fig. 3).

3.2. Effects of different inputs on SOC mapping

We compared the overall accuracy of the twelve input datasets under different synthesis modes to evaluate the effects of different input variables on the model outputs (Fig. 4). Models based only on ALOS-2 had the lowest accuracy in predicting SOC (HH: R^2 = 0.0522; HV: R^2 = 0.0854; HH+HV: R^2 = 0.1518); SOC was better predicted by cross-polarization (HV) than HH polarization; the co/cross-polarization combination yielded higher accuracy. The Sentinel series had higher accuracies than ALOS-2. Sentinel-3-based models had the highest accuracy (R^2: 0.3965-0.4019; RMSE: 0.6974-0.7011; MAE: 0.5352-0.5384) when using a single sensor, followed by Sentinel-2 (R^2: 0.3449-0.3486; RMSE: 0.7282-0.7304; MAE: 0.5594-0.5601) and Sentinel-1 (R^2: 0.3250-0.3322; RMSE: 0.7369-0.7413; MAE: 0.5629-0.5669); we also observed that the latter two sensors exhibited similar predictive performances.

The results showed that the prediction models built by Sentinel-1 had obvious differences in accuracy under different polarization modes and orbital directions. Consistent with the results from ALOS-2, the cross-polarization (i.e., VH) from Sentinel-1 yielded a better SOC prediction than the copolarization (i.e., VV). The combined use of copolarization and cross-polarization improved the prediction accuracy compared to the single-polarization mode, with a higher R^2 (0.3250-0.3322) and lower MAE (0.5629-0.5669) and RMSE (0.7369-0.7413). The SOC prediction models based on Sentinel-1 imagery with "ASCENDING" orbits (R^2: 0.2912-0.2984; RMSE: 0.7556-0.7596; MAE: 0.5781-0.5834) had higher R^2 and lower RMSE and MAE values than the models constructed from data with "DESCENDING" orbitals (R^2: 0.2759-0.2831; RMSE: 0.7635-0.7669; MAE: 0.5877-0.5899). Meanwhile, improved accuracy was found when using Sentinel-1 with two orbital orientations, with a 17% increase in R^2 compared to the model constructed with "DESCENDING" orbital data. The R^2 values of the SOC prediction models increased when six time periods were used relative to a single time period (Fig. 4, Fig. 5). The accuracy of SOC predictions using the composite images of six different time intervals varied. For different composite modes (median and mean), the highest R^2 values were obtained in the first and second time periods, respectively. Overall, the SOC prediction accuracy using the mean composite mode was slightly higher than that using the median composite mode.

Although the mapping accuracy of different composite modes and sensors differed, the synergistic utilization of the Sentinel series gave better prediction models and more accurate prediction results. For example, the performance of the SOC prediction models established by

Sentinel-1/2 showed that the R^2 values were > 0.39, the RMSE was between 0.7015 and 0.7026 and the MAE was between 0.5369 and 0.5376, which was better than the results based on Sentinel-1. As we expected, all available Sentinel sensors modeled together had the highest prediction accuracy (R^2 > 0.44), with a greater improvement in the SOC prediction accuracy than with a single Sentinel. The R^2 values of the predictive models built by Sentinel-1/2/3 indicated that these models could explain approximately 33%, 34% and 40% of the SOC variability, respectively, and together they explained 44% of the SOC variation.

3.3. Variable importance

The predictor importance in four experiments conducted in SOC modeling under mean composite mode is shown in Fig. 6. For models based on all the Sentinels, the backscatter bands (Sentinel-1) accounted for 64% of the relative importance of the SOC predictions in the models, followed by Sentinel-2 (18%) and Sentinel-3 (18%); S3_B20 and S3_B10 were the first and third most important predictors in the model, respectively, and the Sentinel-1 VH backscatter bands from the fourth time period and S2_B4 also ranked in the top five in our model. For the model built by Sentinel-2, the most important predictors were S2_B2 (relative importance of 29%) and S2_B4 (16%), which were also the most important Sentinel-2-derived predictors in the model based on dataset D12. S3_B20 and S3_B2 had the highest contribution to the SOC prediction model constructed by Sentinel-3, followed by S3_B1 and S3_B3. The contribution of the SAR product bands to the SOC prediction model based on Sentinel-1 ranged from 3% to 9%; of the top six most important predictors, only one was from VV polarization, and the remaining five were from VH polarization, which further suggests that VH polarization has a greater impact on the model than VV.

3.4. Spatial distribution of SOC using different inputs

Fig. 7 shows the modeled SOC maps for the study area with different inputs under two composite modes. The spatial pattern of the mapping results of the SOC prediction model with the same input under the two composite modes was consistent. The overall trend was similar among the digital maps produced by different SOC prediction models based on the Sentinel series, but the spatial details differed slightly; the ALOS-2-based model had the highest SOC prediction error, resulting in a significantly different spatial pattern of mapping results from these models. The mean and SD of the predicted maps from these prediction models based on different datasets were very close despite differences in spatial patterns and spatial details.

4. Discussion

4.1. The influence of input satellite sensors on prediction performance

Our results show that the composite mode and selection of satellite imagery, as well as the radar data utilization strategies, influenced the prediction performances (Fig. 4, Fig. 5). The L-band ALOS-2 images had the poorest performance, where HV polarization was more effective in predicting SOC than HH polarization. Combining HV and HH polarization performed better than a single polarization mode. This is consistent with Sentinel-1 predictions, where cross-polarization provided more accurate results than copolarization, and their combination yielded higher accuracy. These differences are attributed to the different ways in which microwaves interact with the surface or canopy, and the integration of co/cross-polarization provides richer information on radar wave scattering and effectively improves the accuracy (Chen et al., 2020; Hosseini et al., 2015).

Our results show that predictive models based on Sentinel-1 data with "ASCENDING" orbits perform better than models built from "DESCENDING" orbital data. We recommend building soil prediction models based on mixed information with different orbital directions, which can effectively improve the mapping accuracy (Fig. 4). Several studies have highlighted the differences in the backscatter and scattering mechanisms in the two orbital directions of SAR (Elfadaly et al., 2020; Mahdavi et al., 2019). Our results also show that the prediction model based on Sentinel-1 with multiple time periods is more stable and accurate than Sentinel-1 with a single time period (Fig. 4, Fig. 5). Similar findings were reported by Dou et al. (2019) and Silvero et al. (2021). The advantages of multitemporal remote sensing imagery can be explained by its ability to capture changes in land surface characteristics over time (Fathololoumi et al., 2020). Furthermore, our results show that the prediction performance of L-band ALOS-2 is lower than that of C-band Sentinel-1. Several scholars have emphasized the importance of studying the effects of band frequencies and polarization modes on the modeling of SAR data (Hosseini et al., 2015; Rapinel et al., 2020; Shu et al., 2020). However, to our knowledge, few scholars have comprehensively evaluated the effects of the polarization mode, band frequency, orbital direction and time window of SAR data on SOC modeling.

Our results demonstrate that Sentinel-1/2/3 with different spatiotemporal resolutions are well suited to map SOC content at the national scale in Spain. The accuracy ranking based on these sensors was as follows: Sentinel-3 > Sentinel-2 > Sentinel-1 (Fig. 4). This result agrees with previous studies by Zhou et al. (2021) and Silvero et al. (2021), who reported that satellite imagery with higher spatial resolutions did not necessarily provide better results. It is worth noting that the Sentinel-2 (optical) and Sentinel-1 (SAR) data had similar performance in SOC modeling. Therefore, the prediction of SOC based on the integration of open access SAR data obtained from ESA is very promising during periods when ideal optical data are not available. To the best of our knowledge, this is the first study that employed long-term composite Sentinel-1/2/3 images on the GEE platform to analyze and compare their potential for predicting SOC, but some studies have reported suitable results based on single-date or multitemporal Sentinel data. For example, several recent studies have demonstrated the usefulness of Copernicus Sentinel families in SOC mapping, including Sentinel-1 (Tripathi and Tiwari, 2022), Sentinel-2 (Gholizadeh et al., 2018) and Sentinel-3 (Odebiri et al., 2022).

Our study reveals that the GEE cloud platform has promising prospects in soil mapping due to its powerful data processing capability and sufficient satellite image data. Amani et al. (2020) reviewed various applications of the GEE platform and found that the Landsat and Sentinel datasets are widely used by GEE users, but only 3% are soil-related studies. Luo et al. (2022a) reported the advantage of using the GEE platform for soil attribute prediction, avoiding the impact of insufficient data and environmental differences on regional soil mapping. Combining Sentinel series data yielded the best results in all experiments (Fig. 4). Similar findings have been reported by other studies using Sentinel-1 and Sentinel-2 (as well as other optical and SAR data) in soil property prediction (Nguyen et al., 2022; Wang et al., 2020). This can be attributed to the differences in image characteristics, information content, and imaging techniques of the two systems (Forkuor et al., 2020), which help improve SOC predictions when they are combined. This knowledge notwithstanding, this study achieved efficient mapping of large-scale SOC using long-term composite images on the GEE platform, taking into account the influence of sensor types, composite modes and radar data utilization strategies in the modeling, which provides new insights into the complementarity and strengths of the Sentinel series for future rapid, large-scale, high-resolution modeling of soil properties.

4.2. The importance of using satellite-derived predictors in the GEE platform for SOC modeling

Previous SOC mapping studies have relied heavily on optical sensors, such as Landsat or MODIS. However, our results showed that optical and radar sensors and their combinations from the Sentinel series are essential for effective SOC modeling. Among the prediction models constructed by all the Sentinel series, Sentinel-1 showed a stronger impact than Sentinel-2 and Sentinel-3 in SOC prediction, with a sum of importance of 64%, 18% and 18%, respectively (Fig. 6). This is similar to previous studies, which highlighted the importance of the Sentinel series, that included radar (e.g., (Zhou et al., 2020a)) and optical (e.g., (Castaldi et al., 2019)) sensors, in explaining SOC variability. For the model built with Sentinel-1, five of the top six most important predictors were derived from VH polarization, revealing that VH polarization had a greater impact than VV polarization in our predictions. Among the Sentinel-1-derived predictors, we also observed a weaker correlation of SOC with VV polarization and a relatively strong correlation with VH polarization. Long-term composite backscatter bands in our prediction model were identified as effective predictors of SOC. Although radar data have rarely been used to map SOC, previous studies using these images to monitor vegetation have demonstrated the ability of radar data to capture short-term changes in vegetation characteristics, which can be further applied to predict SOC due to the relationship in the soil-vegetation system (Yang and Guo, 2019; Yang et al., 2019).

Among the Sentinel-2-built models, S2_B2 (relative importance of 29%) and S2_B4 (16%) were identified as the most influential predictors for SOC mapping and were also the most significant Sentinel-2-derived predictors when the models were built with all the Sentinels (Fig. 6). This finding was reported by other studies using Sentinel-2 to predict SOC. For example, Zhou et al. (2021) found that S2_B2 was the most important predictor among all the spectral bands of Sentinel-2 for SOC prediction. On the other hand, S2_B4 was the most correlated spectral band in the correlation analysis between the SOC and Sentinel-2 bands. With respect to Sentinel-3, S3_B20 had the highest importance in experiments involving Sentinel-3. The spectral reflectance of optical satellites is the most commonly used practical indicator for SOC prediction. The spectral bands based on remote sensing can reflect the biophysical properties related to vegetation cover and soil conditions (Lamichhane et al., 2019; Xu et al., 2017). The close relationship of soil-vegetation systems observed by satellite images helps explain the spatial variability of SOC (Yang and Guo, 2019). Long-term vegetation cover conditions may be more influential than short-term snapshots. The GEE platform is very efficient in acquiring long-term time series images, and its potential in soil mapping cannot be underestimated.

4.3. Digital SOC maps and spatial pattern analysis using different inputs

The SOC maps predicted by the models involving Sentinel data (i.e., Sentinel-1/2/3) in this study were compared with those produced by Wang et al. (2020) and Calvo de Anta et al. (2020). These predicted SOC maps had roughly the same spatial distribution trends as our mapping results involving Sentinel satellites (Fig. 7). The spatial pattern of the mapping results of the models with the same Sentinel input was consistent in the two composite modes. The spatial details of the mapping results from satellite sensors with different image characteristics, information content and imaging techniques varied slightly. The predicted SOC maps exhibited strong spatial variation due to the complex interaction of environmental factors such as land use, climate, topography and vegetation (Calvo de Anta et al., 2020). The spatial pattern of SOC was in accordance with the land use distribution, with higher SOC concentrations in the northwest dominated by closed broadleaved deciduous forest. We also found high values in some mountainous areas of the Central System and the Iberian System, which is consistent with the results of López-Senespleda et al. (2021). Another area of noteworthy high values was the high-altitude Pyrenees, a mountain system that originates from the Alps and is unconnected to the

Central Plateau. The agricultural systems (especially woody crops) had lower SOC concentrations than the above areas located under forests and scrubs. The lowest SOC levels were observed in the southern regions, which depends mainly on the climatic conditions in Spain (Rodríguez Martín et al., 2016). Land use and climate are the key variables in determining SOC concentrations in Spain, as confirmed by other studies (Hontoria et al., 1999; Rodríguez Martín et al., 2016). Rodríguez Martín et al. (2016) modeled the spatial variability of SOC in Spain and found that the average SOC content of forestlands and grasslands was more than 3 times higher than that of croplands.

5. Conclusions

This study combined machine learning and multiple types of long-term optical and microwave satellite observations acquired from the GEE platform to estimate the spatial distribution of SOC concentrations in Spain. The main conclusions can be summarized as follows:

- The results show that the composite mode and choice of satellite imagery, as well as the SAR acquisition configurations, affect the model results to varying degrees. The models constructed from SAR data involving cross-polarization, multiple time periods and "ASCENDING" orbits performed better than models involving copolarization, a single time period and "DESCENDING" orbits. The soil prediction models based on mixed information of different orbital directions and polarization modes effectively improved the mapping accuracy.

- The accuracy rankings based on long-term optical and microwave sensor observations were as follows: Sentinel-3 > Sentinel-2 > Sentinel-1 > ALOS-2. The predictive performance of C-band Sentinel-1 versus multispectral Sentinel-2 was comparable, although the complementary use of the two provided higher accuracy.
- Our mapping results involving Sentinel satellites had similar spatial patterns with slightly different spatial details, exhibiting extensive spatial variability. The predicted SOC was higher in northwestern Spain and lower in southern Spain.
- This study emphasizes the benefits of the GEE platform allowing rapid, dynamic analysis of EO data in near real-time to support soil mapping and health monitoring.
- Our results confirm the good predictive power of long-term Sentinel-1/2/3 observations, and the massive amounts of freely available high-quality Sentinel data is expected to accelerate and advance digital soil mapping research.

Acknowledgments

We thank the European Soil Data Centre, ESA and JAXA for their data support. We are also very grateful to the anonymous reviewers. This work was performed with the support of the Ludong University.

References

Allende-Montalbán, R., Martín-Lammerding, D., del Mar Delgado, M., Porcel, M.A., Gabriel, J.L., 2022. Nitrate Leaching in Maize (Zea mays L.) and Wheat (Triticum aestivum L.) Irrigated Cropping Systems under Nitrification Inhibitor and/or Intercropping Effects. Agriculture 12(4).

Amani, M., Ghorbanian, A., Ahmadi, S.A., Kakooei, M., Moghimi, A., Mirmazloumi, S.M., Moghaddam, S.H.A., Mahdavi, S., Ghahremanloo, M., Parsian, S., Wu, Q., Brisco, B., 2020. Google Earth Engine Cloud Computing Platform for Remote Sensing Big Data Applications: A Comprehensive Review. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 13, 5326-5350.

Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., Montanarella, L., Panagos, P., 2019. Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. Geoderma 355, 113912.

Baltensweiler, A., Walthert, L., Hanewinkel, M., Zimmermann, S., Nussbaum, M., 2021. Machine learning based soil maps for a wide range of soil properties for the forested area of Switzerland. Geoderma Regional 27, e00437.

Borrelli, P., Poesen, J., Vanmaercke, M., Ballabio, C., Hervás, J., Maerker, M., Scarpa, S., Panagos, P., 2022. Monitoring gully erosion in the European Union: A novel approach based on the Land Use/Cover Area frame survey (LUCAS). International Soil and Water Conservation Research 10(1), 17-28.

Caddeo, A., Marras, S., Sallustio, L., Spano, D., Sirca, C., 2019. Soil organic carbon in Italian forests and agroecosystems: Estimating current stock and future changes with a spatial modelling approach. Agricultural and Forest Meteorology 278, 107654.

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, 114365.

Castaldi, F., Chabrillat, S., Jones, A., Vreys, K., Bomans, B., Van Wesemael, B., 2018. Soil Organic Carbon Estimation in Croplands by Hyperspectral Remote APEX Data Using the LUCAS Topsoil Database. Remote Sensing 10(2).

Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B., 2019. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. ISPRS Journal of Photogrammetry and Remote Sensing 147, 267-282.

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.

Chen, S., Useya, J., Mugiyo, H., 2020. Decision-level fusion of Sentinel-1 SAR and Landsat 8 OLI texture features for crop discrimination and classification: case of Masvingo, Zimbabwe. Heliyon 6(11), e05358.

Clevers, J.G.P.W., Gitelson, A.A., 2013. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. IJAEO 23, 344-351.

DeVries, B., Huang, C., Armston, J., Huang, W., Jones, J.W., Lang, M.W., 2020. Rapid and robust monitoring of flood events using Sentinel-1 and Landsat data on the Google Earth Engine. RSEnv 240, 111664.

Dharumarajan, S., Kalaiselvi, B., Suputhra, A., Lalitha, M., Vasundhara, R., Kumar, K.S.A., Nair, K.M., Hegde, R., Singh, S.K., Lagacherie, P., 2021. Digital soil mapping of soil organic carbon stocks in

Western Ghats, South India. Geoderma Regional 25, e00387.

Dou, X., Wang, X., Liu, H., Zhang, X., Meng, L., Pan, Y., Yu, Z., Cui, Y., 2019. Prediction of soil organic matter using multi-temporal satellite images in the Songnen Plain, China. Geoderma 356, 113896.

Elfadaly, A., Abate, N., Masini, N., Lasaponara, R., 2020. SAR Sentinel 1 Imaging and Detection of Palaeo-Landscape Features in the Mediterranean Area. Remote Sensing 12(16).

Esch, T., Adam, F., Bachofer, F., Boudinaud, L., Hirner, A., Gessner, U., Knauer, K., Metz-Marconcini, A., Marconcini, M., Üreyen, S., Winkler, K., Zeidler, J., 2018a. New Opportunities for Urban Land Governance by Exploiting Big Data from Earth Observation, Annual World Bank Conference on Land and Poverty 2018, Washington D.C., USA, pp. 1-20.

Esch, T., Üreyen, S., Zeidler, J., Metz–Marconcini, A., Hirner, A., Asamer, H., Tum, M., Böttcher, M., Kuchar, S., Svaton, V., Marconcini, M., 2018b. Exploiting big earth data from space – first experiences with the timescan processing chain. Big Earth Data 2(1), 36-55.

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. ScTEn 721, 137703.

Fernandez-Moran, R., Gómez-Chova, L., Alonso, L., Mateo-García, G., López-Puigdollers, D., 2021. Towards a novel approach for Sentinel-3 synergistic OLCI/SLSTR cloud and cloud shadow detection based on stereo cloud-top height estimation. ISPRS Journal of Photogrammetry and Remote Sensing 181, 238-253.

Fernández-Ugalde, O., Ballabio, C., Lugato, E., Scarpa, S., Jones, A., 2020. Assessment of changes in topsoil properties in LUCAS samples between 2009/2012 and 2015 surveys. Publications Office of the European Union, 11-12.

Forkuor, G., Benewinde Zoungrana, J.-B., Dimobe, K., Ouattara, B., Vadrevu, K.P., Tondoh, J.E., 2020.

Above-ground biomass mapping in West African dryland forest using Sentinel-1 and 2 datasets - A case study. RSEnv 236, 111496.

Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L., 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. RSEnv 218, 89-103.

Guo, L., Sun, X., Fu, P., Shi, T., Dang, L., Chen, Y., Linderman, M., Zhang, G., Zhang, Y., Jiang, Q., Zhang, H., Zeng, C., 2021. Mapping soil organic carbon stock by hyperspectral and time-series multispectral remote sensing images in low-relief agricultural areas. Geoderma 398, 115118.

He, X., Yang, L., Li, A., Zhang, L., Shen, F., Cai, Y., Zhou, C., 2021. Soil organic carbon prediction using phenological parameters and remote sensing variables generated from Sentinel-2 images. Catena 205, 105442.

Hontoria, C., Saa, A., Rodríguez-Murillo, J.C., 1999. Relationships Between Soil Organic Carbon and Site Characteristics in Peninsular Spain. SSSAJ 63(3), 614-621.

Hosseini, M., McNairn, H., Merzouki, A., Pacheco, A., 2015. Estimation of Leaf Area Index (LAI) in corn and soybeans using multi-polarization C- and L-band radar data. RSEnv 170, 77-89.

Huang, W., DeVries, B., Huang, C., Lang, M.W., Jones, J.W., Creed, I.F., Carroll, M.L., 2018. Automated Extraction of Surface Water Extent from Sentinel-1 Data. Remote Sensing 10(5).

Jia, X., O'Connor, D., Shi, Z., Hou, D., 2021. VIRS based detection in combination with machine learning for mapping soil pollution. Environ. Pollut. 268, 115845.

Jones, A., Fernandez-Ugalde, O., Scarpa, S., 2020. LUCAS 2015 Topsoil Survey. Presentation of dataset and.

Khan, M.Z., Chiti, T., 2022. Soil carbon stocks and dynamics of different land uses in Italy using the LUCAS soil database. J. Environ. Manage. 306, 114452.

Kravitz, J., Matthews, M., Bernard, S., Griffith, D., 2020. Application of Sentinel 3 OLCI for chl-a retrieval over small inland water targets: Successes and challenges. RSEnv 237, 111562.

Lamichhane, S., Kumar, L., Adhikari, K., 2021. Updating the national soil map of Nepal through digital soil mapping. Geoderma 394, 115041.

Lamichhane, S., Kumar, L., Wilson, B., 2019. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. Geoderma 352, 395-413.

Li, H., Kato, T., Hayashi, M., Wu, L., 2022. Estimation of Forest Aboveground Biomass of Two Major Conifers in Ibaraki Prefecture, Japan, from PALSAR-2 and Sentinel-2 Data. Remote Sensing 14(3).

Li, H., Zhang, C., Zhang, S., Atkinson, P.M., 2019. Full year crop monitoring and separability assessment with fully-polarimetric L-band UAVSAR: A case study in the Sacramento Valley, California. IJAEO 74, 45-56.

Liu, F., Wu, H., Zhao, Y., Li, D., Yang, J.-L., Song, X., Shi, Z., Zhu, A.X., Zhang, G.-L., 2022a. Mapping high resolution National Soil Information Grids of China. Science Bulletin 67(3), 328-340.

Liu, Y., Wu, H., Wang, S., Chen, X., Kimball, J.S., Zhang, C., Gao, H., Guo, P., 2022b. Evaluation of trophic state for inland waters through combining Forel-Ule Index and inherent optical properties. ScTEn 820, 153316.

López-Senespleda, E., Calama, R., Ruiz-Peinado, R., 2021. Estimating forest floor carbon stocks in woodland formations in Spain. ScTEn 788, 147734.

Luo, C., Wang, Y., Zhang, X., Zhang, W., Liu, H., 2022a. Spatial prediction of soil organic matter content using multiyear synthetic images and partitioning algorithms. Catena 211, 106023.

Luo, C., Zhang, X., Meng, X., Zhu, H., Ni, C., Chen, M., Liu, H., 2022b. Regional mapping of soil organic matter content using multitemporal synthetic Landsat 8 images in Google Earth Engine. Catena 209, 105842.

Luo, C., Zhang, X., Wang, Y., Men, Z., Liu, H., 2022c. Regional soil organic matter mapping models based on the optimal time window, feature selection algorithm and Google Earth Engine. Soil and Tillage Research 219, 105325.

Mahdavi, S., Amani, M., Maghsoudi, Y., 2019. The effects of orbit type on synthetic aperture RADAR (SAR) backscatter. Remote Sensing Letters 10(2), 120-128.

Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Motagh, M., 2017. Random forest wetland classification using ALOS-2 L-band, RADARSAT-2 C-band, and TerraSAR-X imagery. ISPRS Journal of Photogrammetry and Remote Sensing 130, 13-31.

Mallik, S., Bhowmik, T., Mishra, U., Paul, N., 2022. Mapping and prediction of soil organic carbon by an advanced geostatistical technique using remote sensing and terrain data. GeoIn 37(8), 2198-2214.

Mello, F.A.O., Demattê, J.A.M., Rizzo, R., Mello, D.C.d., Poppiel, R.R., Silvero, N.E.Q., Safanelli, J.L., Bellinaso, H., Bonfatti, B.R., Gomez, A.M.R., Sousa, G.P.B., 2022. Complex hydrological knowledge to support digital soil mapping. Geoderma 409, 115638.

Minasny, B., McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264, 301-311.

Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Brisco, B., Motagh, M., 2018. Multi-temporal, multi-frequency, and multi-polarization coherence and SAR backscatter analysis of wetlands. ISPRS Journal of Photogrammetry and Remote Sensing 142, 78-93.

Morales-Polo, C., Cledera-Castro, M.D., Revuelta-Aramburu, M., Hueso-Kortekaas, K., 2021. Bioconversion Process of Barley Crop Residues into Biogas—Energetic-Environmental Potential in Spain. Agronomy 11(4).

Moreno-García, M., Repullo-Ruibérriz de Torres, M.Á., González-Sánchez, E.J., Ordóñez-Fernández, R., Veroz-González, Ó., Carbonell-Bojollo, R.M., 2020. Methodology for estimating the impact of no tillage on the 4perMille initiative: The case of annual crops in Spain. Geoderma 371, 114381.

Mulomba Mukadi, P., González-García, C., 2021. Time Series Analysis of Climatic Variables in Peninsular Spain. Trends and Forecasting Models for Data between 20th and 21st Centuries. Climate 9(7).

Murphy, S.W., de Souza Filho, C.R., Wright, R., Sabatino, G., Correa Pabon, R., 2016. HOTMAP: Global hot target detection at moderate spatial resolution. RSEnv 177, 78-88.

Nabiollahi, K., Taghizadeh-Mehrjardi, R., Shahabi, A., Heung, B., Amirian-Chakan, A., Davari, M., Scholten, T., 2021. Assessing agricultural salt-affected land using digital soil mapping and hybridized random forests. Geoderma 385, 114858.

Nguyen, T.T., Pham, T.D., Nguyen, C.T., Delfos, J., Archibald, R., Dang, K.B., Hoang, N.B., Guo, W., Ngo, H.H., 2022. A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion. ScTEn 804, 150187.

Odebiri, O., Mutanga, O., Odindi, J., 2022. Deep learning-based national scale soil organic carbon mapping with Sentinel-3 data. Geoderma 411, 115695.

Ottinger, M., Kuenzer, C., 2020. Spaceborne L-Band Synthetic Aperture Radar Data for Geoscientific Analyses in Coastal Land Applications: A Review. Remote Sensing 12(14), 2228.

Panagos, P., Meusburger, K., Ballabio, C., Borrelli, P., Alewell, C., 2014. Soil erodibility in Europe: A high-resolution dataset based on LUCAS. ScTEn 479-480, 189-200.

Prikaziuk, E., van der Tol, C., 2019. Global Sensitivity Analysis of the SCOPE Model in Sentinel-3 Bands: Thermal Domain Focus. Remote Sensing 11(20).

Prudente, V.H.R., Skakun, S., Oldoni, L.V., A. M. Xaud, H., Xaud, M.R., Adami, M., Sanches, I.D.A., 2022. Multisensor approach to land use and land cover mapping in Brazilian Amazon. ISPRS

Journal of Photogrammetry and Remote Sensing 189, 95-109.

Purinton, B., Bookhagen, B., 2020. Multiband (X, C, L) radar amplitude analysis for a mixed sand- and gravel-bed river in the eastern Central Andes. RSEnv 246, 111799.

Rapinel, S., Betbeder, J., Denize, J., Fabre, E., Pottier, É., Hubert-Moy, L., 2020. SAR analysis of wetland ecosystems: Effects of band frequency, polarization mode and acquisition dates. ISPRS Journal of Photogrammetry and Remote Sensing 170, 103-113.

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.

Rosenqvist, A., Shimada, M., Suzuki, S., Ohgushi, F., Tadono, T., Watanabe, M., Tsuzuku, K., Watanabe, T., Kamijo, S., Aoki, E., 2014. Operational performance of the ALOS global systematic acquisition strategy and observation plans for ALOS-2 PALSAR-2. RSEnv 155, 3-12.

Rowley, M.C., Grand, S., Verrecchia, É.P., 2018. Calcium-mediated stabilisation of soil organic carbon. Biogeochemistry 137(1), 27-49.

Shu, M., Zhou, L., Gu, X., Ma, Y., Sun, Q., Yang, G., Zhou, C., 2020. Monitoring of maize lodging using multi-temporal Sentinel-1 SAR data. AdSpR 65(1), 470-480.

Silvero, N.E.Q., Demattê, J.A.M., Amorim, M.T.A., Santos, N.V.d., Rizzo, R., Safanelli, J.L., Poppiel, R.R., Mendes, W.d.S., Bonfatti, B.R., 2021. Soil variability and quantification based on Sentinel-2 and Landsat-8 bare soil images: A comparison. RSEnv 252, 112117.

Singha, M., Dong, J., Sarmah, S., You, N., Zhou, Y., Zhang, G., Doughty, R., Xiao, X., 2020. Identifying floods and flood-affected paddy rice fields in Bangladesh based on Sentinel-1 imagery and Google Earth Engine. ISPRS Journal of Photogrammetry and Remote Sensing 166, 278-293.

Song, Y., Chen, B., Kwan, M.-P., 2020. How does urban expansion impact people's exposure to green environments? A comparative study of 290 Chinese cities. Journal of Cleaner Production 246, 119018.

Sothe, C., Gonsamo, A., Arabian, J., Snider, J., 2022. Large scale mapping of soil organic carbon concentration with 3D machine learning and satellite observations. Geoderma 405, 115402.

Szatmári, G., Pásztor, L., Heuvelink, G.B.M., 2021. Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics. Geoderma 403, 115356.

Taghipour, K., Heydari, M., Kooch, Y., Fathizad, H., Heung, B., Taghizadeh-Mehrjardi, R., 2022. Assessing changes in soil quality between protected and degraded forests using digital soil mapping for semiarid oak forests, Iran. Catena 213, 106204.

Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., Brisco, B., 2020. Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. ISPRS Journal of Photogrammetry and Remote Sensing 164, 152-170.

Tiyasha, T., Tung, T.M., Bhagat, S.K., Tan, M.L., Jawad, A.H., Mohtar, W.H.M.W., Yaseen, Z.M., 2021. Functionalization of remote sensing and on-site data for simulating surface water dissolved oxygen: Development of hybrid tree-based artificial intelligence models. Mar. Pollut. Bull. 170, 112639.

Tripathi, A., Tiwari, R.K., 2022. Utilisation of spaceborne C-band dual pol Sentinel-1 SAR data for simplified regression-based soil organic carbon estimation in Rupnagar, Punjab, India. AdSpR 69(4), 1786-1798.

Tziolas, N., Tsakiridis, N., Ogen, Y., Kalopesa, E., Ben-Dor, E., Theocharis, J., Zalidis, G., 2020. An integrated methodology using open soil spectral libraries and Earth Observation data for soil organic carbon estimations in support of soil-related SDGs. RSEnv 244, 111793.

Vaudour, E., Gomez, C., Lagacherie, P., Loiseau, T., Baghdadi, N., Urbina-Salazar, D., Loubet, B., Arrouays, D., 2021. Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content mapping in croplands. IJAEO 96, 102277.

Venter, Z.S., Hawkins, H.-J., Cramer, M.D., Mills, A.J., 2021. Mapping soil organic carbon stocks and trends with satellite-driven high resolution maps over South Africa. ScTEn 771, 145384.

Wadoux, A.M.J.C., Brus, D.J., Heuvelink, G.B.M., 2019. Sampling design optimization for soil mapping with random forest. Geoderma 355, 113913.

Wadoux, A.M.J.C., Minasny, B., McBratney, A.B., 2020. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. Earth-Sci. Rev. 210, 103359.

Wang, J., Xiao, X., Bajgain, R., Starks, P., Steiner, J., Doughty, R.B., Chang, Q., 2019a. Estimating leaf area index and aboveground biomass of grazing pastures using Sentinel-1, Sentinel-2 and Landsat images. ISPRS Journal of Photogrammetry and Remote Sensing 154, 189-201.

Wang, L., Zhou, Y., Liu, J., Liu, Y., Zuo, Q., Li, Q., 2022. Exploring the potential of multispectral satellite images for estimating the contents of cadmium and lead in cropland: The effect of the dimidiate pixel model and random forest. Journal of Cleaner Production 367, 132922.

Wang, M., Liu, Z., Ali Baig, M.H., Wang, Y., Li, Y., Chen, Y., 2019b. Mapping sugarcane in complex landscapes by integrating multi-temporal Sentinel-2 images and machine learning algorithms. Land Use Policy 88, 104190.

Wang, X., Zhang, Y., Atkinson, P.M., Yao, H., 2020. Predicting soil organic carbon content in Spain by combining Landsat TM and ALOS PALSAR images. IJAEO 92, 102182.

Wooster, M.J., Xu, W., Nightingale, T., 2012. Sentinel-3 SLSTR active fire detection and FRP product: Pre-launch algorithm development and performance evaluation using MODIS and ASTER datasets. RSEnv 120, 236-254.

Xu, Y., Smith, S.E., Grunwald, S., Abd-Elrahman, A., Wani, S.P., 2017. Incorporation of satellite remote sensing pan-sharpened imagery into digital soil prediction and mapping models to characterize soil property variability in small agricultural fields. ISPRS Journal of Photogrammetry and Remote Sensing 123, 1-19.

Xu, Y., Yu, L., Li, W., Ciais, P., Cheng, Y., Gong, P., 2020. Annual oil palm plantation maps in Malaysia and Indonesia from 2001 to 2016. Earth Syst. Sci. Data 12(2), 847-867.

Yang, L., Cai, Y., Zhang, L., Guo, M., Li, A., Zhou, C., 2021a. A deep learning method to predict soil organic carbon content at a regional scale using satellite-based phenology variables. IJAEO 102, 102428.

Yang, R.-M., Guo, W.-W., 2019. Using time-series Sentinel-1 data for soil prediction on invaded coastal wetlands. Environ. Monit. Assess. 191(7), 462.

Yang, R.-M., Guo, W.-W., Zheng, J.-B., 2019. Soil prediction for coastal wetlands following Spartina alterniflora invasion using Sentinel-1 imagery and structural equation modeling. Catena 173, 465-470.

Yang, X., Qin, Q., Yésou, H., Ledauphin, T., Koehl, M., Grussenmeyer, P., Zhu, Z., 2020. Monthly estimation of the surface water extent in France at a 10-m resolution using Sentinel-2 data. RSEnv 244, 111803.

Yang, X., Xiao, X., Qin, Y., Wang, J., Neal, K., 2021b. Mapping forest in the southern Great Plains with ALOS-2 PALSAR-2 and Landsat 7/8 data. IJAEO 104, 102578.

Zhang, S., Liu, G., Chen, S., Rasmussen, C., Liu, B., 2021. Assessing soil thickness in a black soil watershed in northeast China using random forest and field observations. International Soil and Water Conservation Research 9(1), 49-57.

Zhang, Z., Zhang, H., Xu, E., 2022. Enhancing the digital mapping accuracy of farmland soil organic carbon in arid areas using agricultural land use history. Journal of Cleaner Production 334, 130232.

Zhao, F., Sun, R., Zhong, L., Meng, R., Huang, C., Zeng, X., Wang, M., Li, Y., Wang, Z., 2022. Monthly mapping of forest harvesting using dense time series Sentinel-1 SAR imagery and deep learning. RSEnv 269, 112822.

Zhong, L., Guo, X., Xu, Z., Ding, M., 2021. Soil properties: Their prediction and feature extraction from the LUCAS spectral library using deep convolutional neural networks. Geoderma 402, 115366.

Zhou, T., Geng, Y., Chen, J., Liu, M., Haase, D., Lausch, A., 2020a. Mapping soil organic carbon content using multi-source remote sensing variables in the Heihe River Basin in China. Ecol. Indicators 114, 106288.

Zhou, T., Geng, Y., Chen, J., Pan, J., Haase, D., Lausch, A., 2020b. High-resolution digital mapping of soil organic carbon and soil total nitrogen using DEM derivatives, Sentinel-1 and Sentinel-2 data based on machine learning algorithms. ScTEn 729, 138244.

Zhou, T., Geng, Y., Ji, C., Xu, X., Wang, H., Pan, J., Bumberger, J., Haase, D., Lausch, A., 2021. Prediction of soil organic carbon and the C:N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images. ScTEn 755, 142661.

Figure Legends

Fig. 1. Overview of the study area and soil observations.

Fig. 2. Spatial distribution of the total number of observations per pixel for the different Sentinels over Spain.

Fig. 3. Correlation matrix using Pearson's correlation coefficients between SOC and the satellite-derived predictors. The S2_B2_mean to the S2_B12_mean correspond to bands 2 to 12 of the Sentinel-2 data under the mean composite mode, respectively; The S3_B1_mean to the S3_B21_mean correspond to bands 1 to 21 of the Sentinel-3 data under the mean composite mode, respectively; The VHa_1_mean to the VHa_6_mean correspond to the Sentinel-1 VH polarizations (mean composite mode) with "ASCENDING" orbits at 1 to 6 different time periods, respectively; The VHd_1_mean to the VHd_6_mean correspond to the Sentinel-1 VH polarizations (mean composite mode) with "DESCENDING" orbits at 1 to 6 different time periods, respectively.

Fig. 4. Prediction accuracy of twelve experiments conducted with SOC modeling under different synthesis modes (For the definition of the datasets used in the experiments, see Table 1).

Fig. 5. Prediction accuracy of the Sentinel-1 composite images for six time periods (e.g., period 1: January–February 2015) under different composite modes.

Fig. 6. Variable importance of the experiments conducted with SOC modeling under the mean composite mode (for the specific meaning of the abbreviations, see Fig. 3). (a) - (d) correspond to the results of datasets D8, D9, D10 and D12, respectively (see Table 1 for a description of the different datasets).

Fig. 7. Modeled SOC maps of the experiments conducted under different synthesis modes (the subareas of the predicted map are displayed to the right). The first map represents the results from the D3 dataset; the predicted maps under the mean composite mode correspond to datasets D8 (b), D9 (d), D10 (f) and D12 (h); and the predicted maps under the median composite mode correspond to datasets D8 (c), D9 (e), D10 (g) and D12 (i) (see Table 1 for a description of the different datasets).

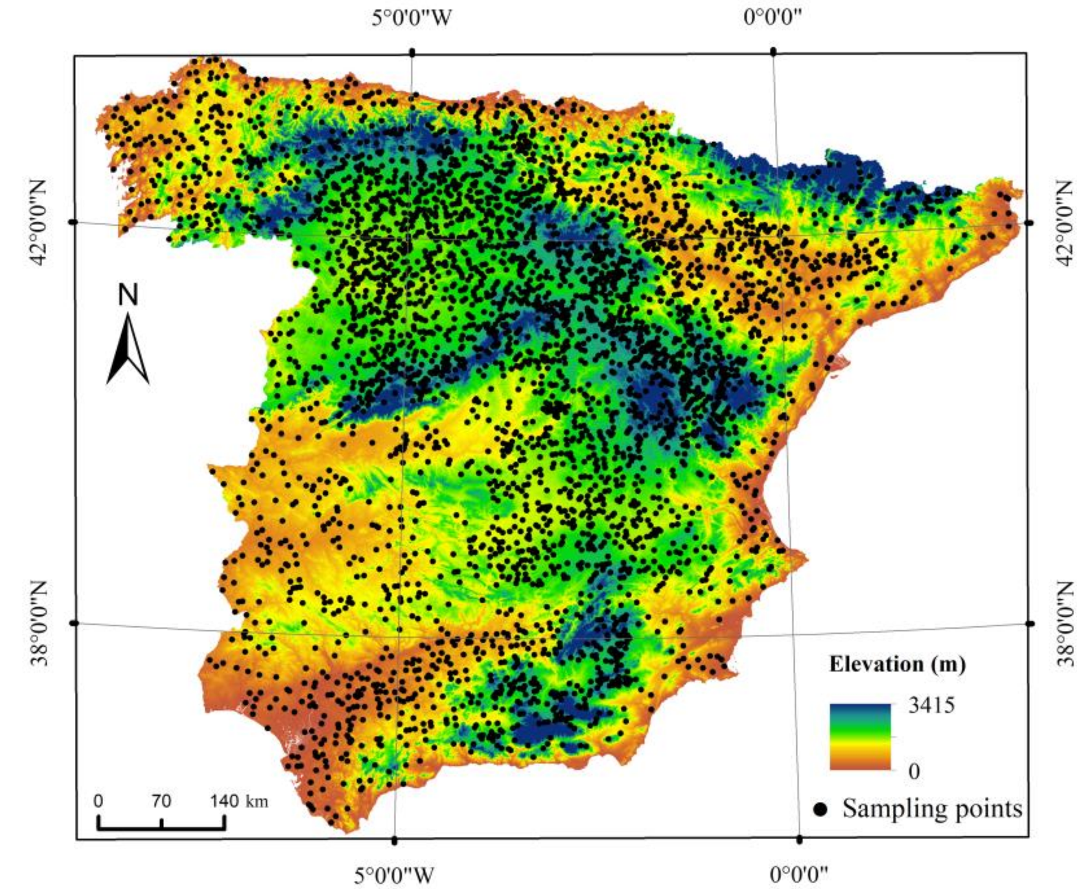
5°0'0"W
0°0'0"
42°0'0"N
38°0'0"N
N
Elevation (m)
3415
0
Sampling points
0 70 140 km

Fig. 1.

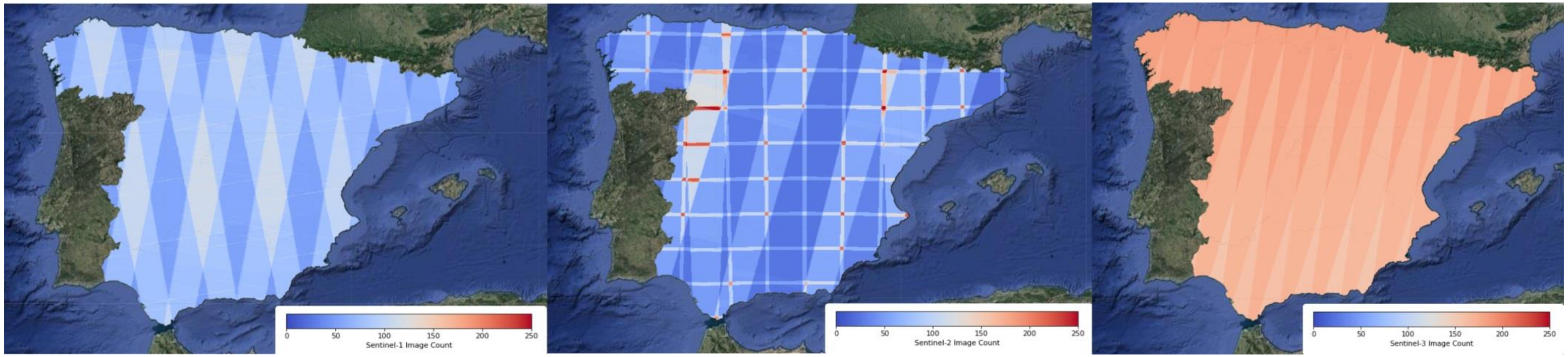
0 50 100 150 200 250
Sentinel-1 Image Count
0 50 100 150 200 250
Sentinel-2 Image Count
0 50 100 150 200 250
Sentinel-3 Image Count

Fig. 2.

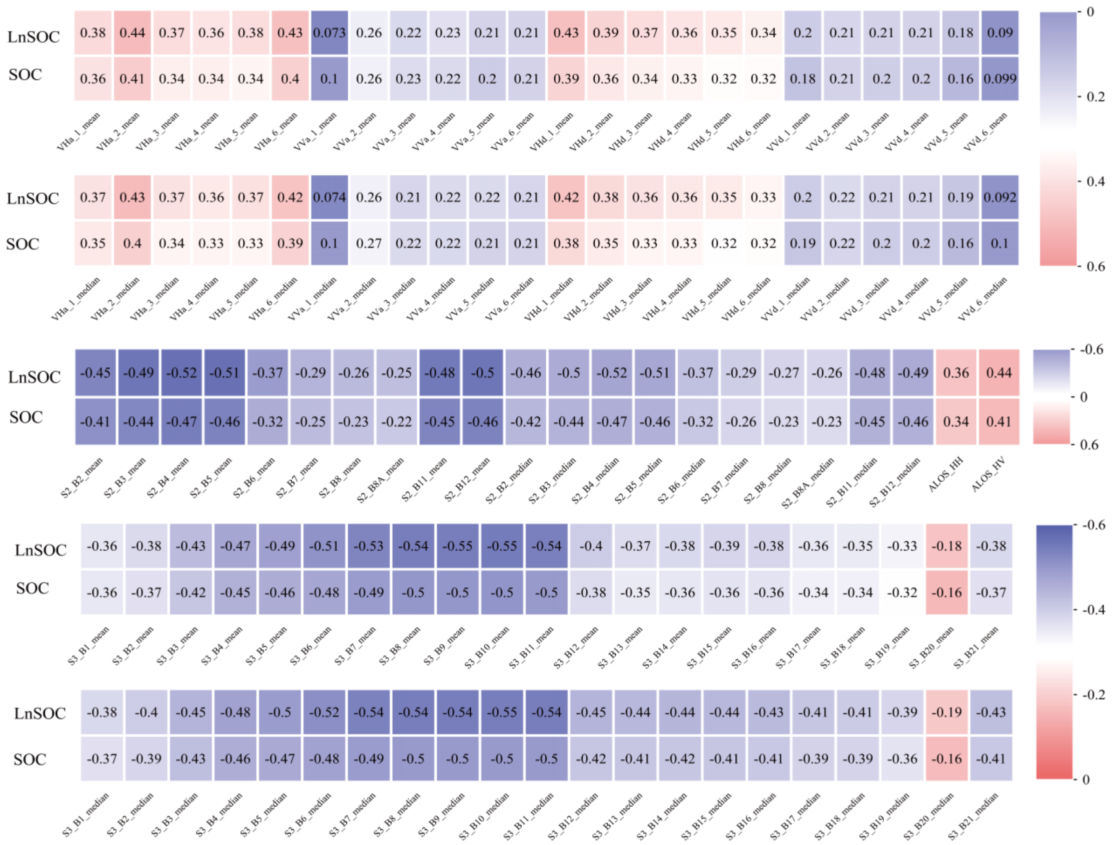

Fig. 3.

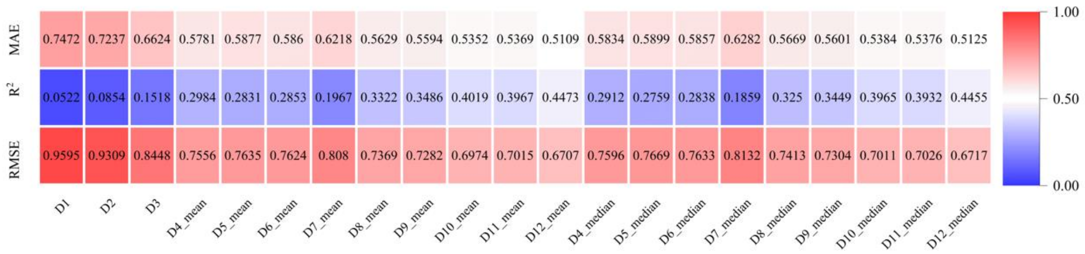
MAE
R2
RMSE
0.7472 0.7237 0.6624 0.5781 0.5877 0.586 0.6218 0.5629 0.5594 0.5352 0.5369 0.5109 0.5834 0.5899 0.5857 0.6282 0.5669 0.5601 0.5384 0.5376 0.5125
0.0522 0.0854 0.1518 0.2984 0.2831 0.2853 0.1967 0.3322 0.3486 0.4019 0.3967 0.4473 0.2912 0.2759 0.2838 0.1859 0.325 0.3449 0.3965 0.3932 0.4455
0.9595 0.9309 0.8448 0.7556 0.7635 0.7624 0.808 0.7369 0.7282 0.6974 0.7015 0.6707 0.7596 0.7669 0.7633 0.8132 0.7413 0.7304 0.7011 0.7026 0.6717
D1 D2 D3 D4_mean D5_mean D6_mean D7_mean D8_mean D9_mean D10_mean D11_mean D12_mean D4_median D5_median D6_median D7_median D8_median D9_median D10_median D11_median D12_median
1.00
0.50
0.00

Fig. 4.

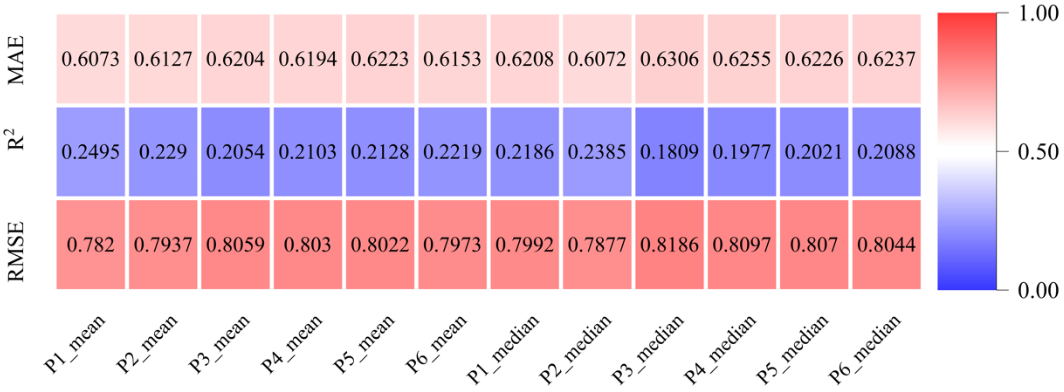
MAE
R2
RMSE
0.6073
0.6127
0.6204
0.6194
0.6223
0.6153
0.6208
0.6072
0.6306
0.6255
0.6226
0.6237
0.2495
0.229
0.2054
0.2103
0.2128
0.2219
0.2186
0.2385
0.1809
0.1977
0.2021
0.2088
0.782
0.7937
0.8059
0.803
0.8022
0.7973
0.7992
0.7877
0.8186
0.8097
0.807
0.8044
P1_mean
P2_mean
P3_mean
P4_mean
P5_mean
P6_mean
P1_median
P2_median
P3_median
P4_median
P5_median
P6_median
1.00
0.50
0.00

Fig. 5.

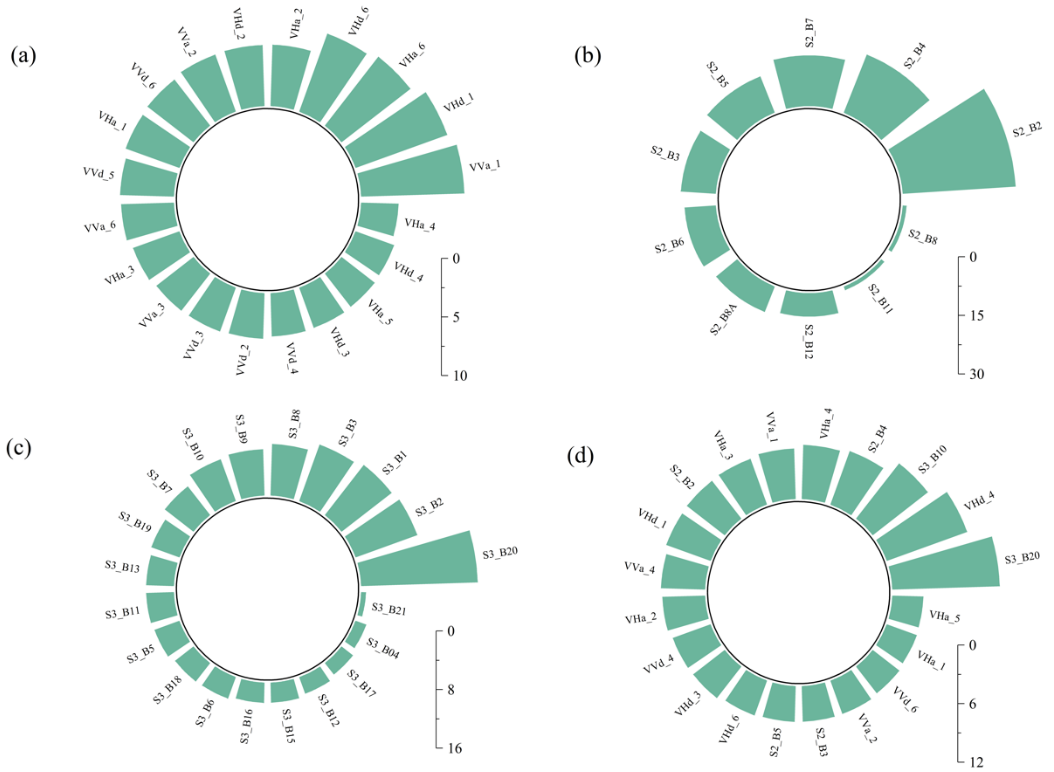
(a)
VVa_1
VHd_1
VHa_6
VHd_6
VHa_2
VHd_2
VVa_2
VVd_6
VHa_1
VVd_5
VVa_6
VHa_3
VVa_3
VVd_3
VVd_2
VVd_4
VHd_3
VHa_5
VHd_4
VHa_4
0
5
10
(b)
S2_B2
S2_B4
S2_B7
S2_B5
S2_B3
S2_B6
S2_B8A
S2_B12
S2_B11
S2_B8
0
15
30
(c)
S3_B20
S3_B2
S3_B1
S3_B3
S3_B8
S3_B9
S3_B10
S3_B7
S3_B19
S3_B13
S3_B11
S3_B5
S3_B18
S3_B6
S3_B16
S3_B15
S3_B12
S3_B17
S3_B04
S3_B21
0
8
16
(d)
S3_B20
VHd_4
S3_B10
S2_B4
VHa_4
VVa_1
VHa_3
S2_B2
VHd_1
VVa_4
VHa_2
VVd_4
VHd_3
VHd_6
S2_B5
S2_B3
VVa_2
VVd_6
VHa_1
VHa_5
0
6
12

Fig. 6.

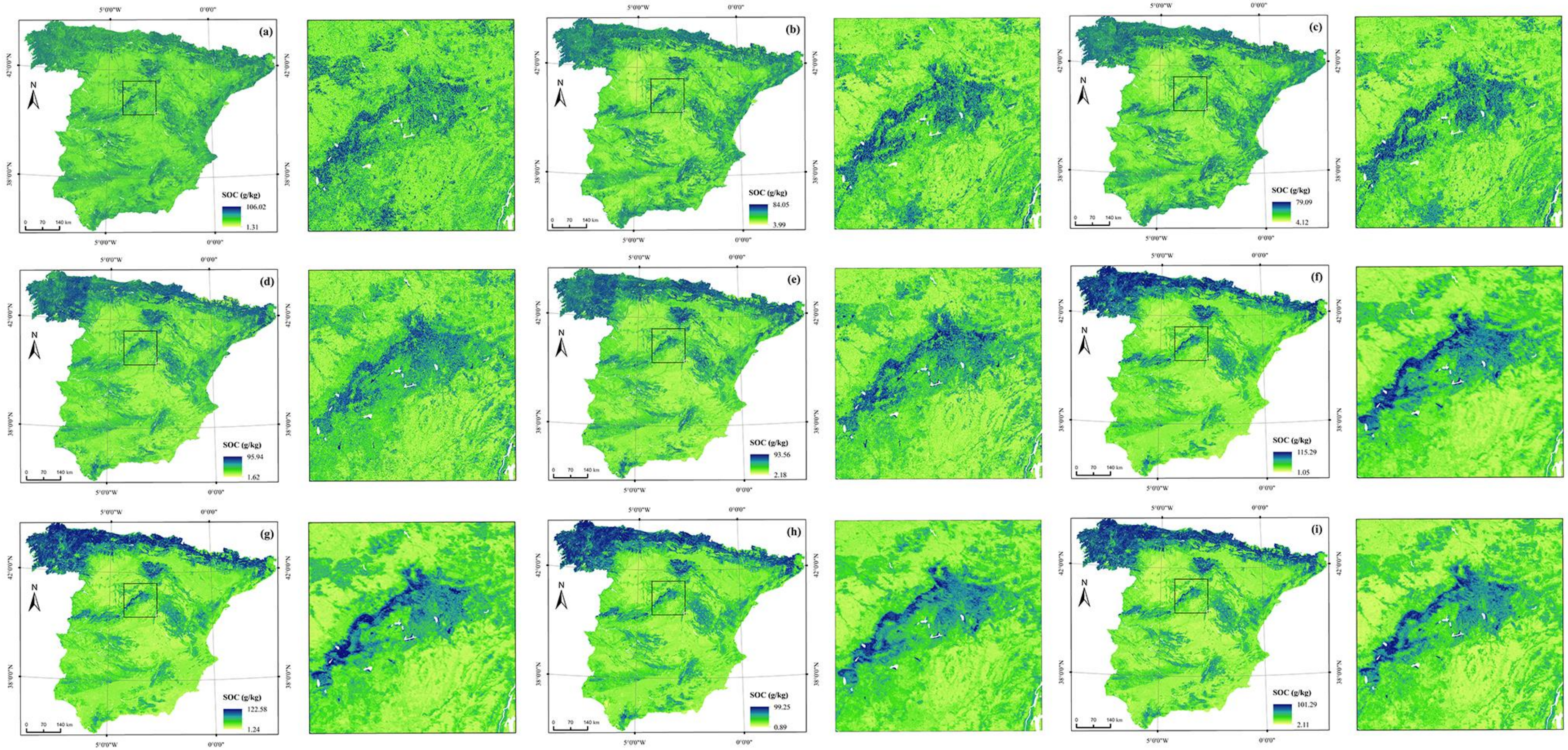
(a)
SOC (g/kg)
106.02
1.31
(b)
SOC (g/kg)
84.05
3.99
(c)
SOC (g/kg)
79.09
4.12
(d)
SOC (g/kg)
95.94
1.62
(e)
SOC (g/kg)
93.56
2.18
(f)
SOC (g/kg)
115.29
1.05
(g)
SOC (g/kg)
122.58
1.24
(h)
SOC (g/kg)
99.25
0.89
(i)
SOC (g/kg)
101.29
2.11
5°0'0"W
0°0'0"
42°0'0"N
38°0'0"N
N
0 70 140 km

Fig. 7.

Table 1. Details of the experimental setups formed from the different satellite observation datasets.

Sensors	Datasets	Description
ALOS-2	D1	Backscatter coefficient in HH polarization
ALOS-2	D2	Backscatter coefficient in HV polarization
ALOS-2	D3	Backscatter coefficient in HH and HV polarization
Sentinel-1	D4	Backscatter coefficient in "ASCENDING" orbit
Sentinel-1	D5	Backscatter coefficient in "DESCENDING" orbit
Sentinel-1	D6	Backscatter coefficient in VH polarization
Sentinel-1	D7	Backscatter coefficient in VV polarization
Sentinel-1	D8	All available Sentinel-1 polarization metrics
Sentinel-2	D9	All available Sentinel-2 derived predictors
Sentinel-3	D10	All available Sentinel-3 derived predictors
Sentinel-1/2	D11	All available Sentinel-1/2 derived predictors
Sentinel-1/2/3	D12	All available Sentinel-1/2/3 derived predictors

Table 2. Descriptive statistics of raw and log-transformed SOC (g/kg).

	Minimum	Maximum	Mean	Median	Standard deviation (SD)	Skewness
SOC	0.10	406.10	22.95	14.30	25.42	3.65
LnSOC	-2.30	6.00	2.72	2.66	0.90	0.03

Notes: LnSOC, log-transformed SOC.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

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