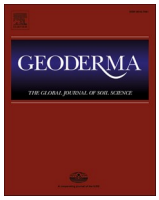
Geoderma 452 (2024) 117113

Contents lists available at ScienceDirect

Geoderma

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



A European soil organic carbon monitoring system leveraging Sentinel 2 imagery and the LUCAS soil data base

Bas van Wesemael [a,*], Asmaa Abdelbaki [b,c], Eyal Ben-Dor [d], Sabine Chabrillat [b,e], Pablo d'Angelo [f], José A.M. Demattê [g], Giulio Genova [h], Asa Gholizadeh [i], Uta Heiden [f], Paul Karlshoefer [f], Robert Milewski [b], Laura Poggio [h], Marmar Sabetizade [a], Adrián Sanz [j], Peter Schwind [f], Nikolaos Tsakiridis [k], Nikolaos Tziolas [l], Julia Yagüe [j], Daniel Žížala [i]

[a] *Earth and Life Institute, Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium*
[b] *Section Remote Sensing and Geoinformatics, Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany*
[c] *Soils and Water Science Department, Faculty of Agriculture, Fayoum University, Fayoum 63514, Egypt*
[d] *Porter School of Environment and Earth Science, Faculty of Exact Sciences, Tel Aviv University, 69978 Tel Aviv, Israel*
[e] *Leibniz University of Hannover, Institute of Earth System Science, Department of Soil Science, Herrenhäuser Str. 2, 30419 Hannover, Germany*
[f] *German Aerospace Center (DLR), Remote Sensing Technology Institute (IMF), Oberpfaffenhofen, 82234 Wessling, Germany*
[g] *Luiz de Queiroz College of Agriculture, Department of Soil Science, University of São Paulo, Piracicaba, Brazil*
[h] *ISRIC - World Soil Information, Droevendaalsesteeg 3, 6708 PB Wageningen, the Netherland*
[i] *Department of Soil Science and Soil Protection, Faculty of Agrobiology, Food and Natural Resources, Czech University of Life Sciences Prague, Kamycka 129, Prague 16500, Czech Republic*
[j] *GMV - Remote Sensing and Geospatial Analytics Division, Isaac Newton 11, P.T.M. Tres Cantos, E-28760 Madrid, Spain*
[k] *Laboratory of Remote Sensing, Spectroscopy, and GIS, Department of Agriculture, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece*
[l] *Southwest Florida Research and Education Center, Department of Soil, Water and Ecosystem Sciences, Institute of Food and Agricultural Sciences, University of Florida, 2685 State Rd 29N, Immokalee, FL 34142, USA*

ARTICLE INFO

Handling Editor: B. Minasny

Keywords:
SOC content
Sentinel-2
Temporal composite
Soil parameter modelling
LUCAS soil
Validation
European soil monitoring

ABSTRACT

The Worldsoils project has developed a pre-operational Soil Organic Carbon (SOC) monitoring system in a cloud environment. The system predicts topsoil organic carbon content at regional and continental scales from Earth Observation (EO) satellite data with a continuous cover over Europe. The system utilizes spectral models for croplands and a digital soil mapping approach for permanently vegetated areas such as grasslands and forests. Models strongly rely on soil reflectance composites from the Sentinel 2 multispectral instrument providing the median reflectance for all valid pixels over a period of three years. The bare soil frequency, a proxy for the degree of crop cover, is clearly lower in a Mediterranean pilot region compared to croplands in temperate regions. This is due to the extensive crop cover in the Mediterranean with winter cereals and fodder crops. The graphical user interface provides SOC content and the prediction interval ratio (i.e. 90 % uncertainty interval divided by the median) for 50 m pixels in three pilot regions and 100 m pixels for the rest of Europe. The SOC prediction algorithms are reasonable compared to others at the continental scale (R^2: 0.41 for croplands and 0.28 for permanently vegetated areas). Apart from tree crops in Macedonia (Greece) the soil reflectance composite attributes the correct model to validation sets of cropland and grassland in the pilot regions. The SOC prediction is satisfactory in Wallonia (Belgium; R^2 0.51) but is less accurate in Greece and the Czech Republic. In particular in Greece, the poor performance is linked to the low bare soil frequency due to the abundance of tree crops, cereals and fodder crops. The monitoring system can reproduce spatial patterns in SOC content similar to the ones obtained from a detailed regional algorithm using the new generation of hyperspectral satellites. However, the very high values in kettle holes in a morainic landscape of Northern Germany are underestimated.

* Corresponding author.
E-mail addresses: bas.vanwesemael@uclouvain.be (B. van Wesemael), asmaa@gfz-potsdam.de (A. Abdelbaki), bendor@tauex.tau.ac.il (E. Ben-Dor), chabri@gfz-potsdam.de (S. Chabrillat), Pablo.Angelo@dlr.de (P. d'Angelo), jamdemat@usp.br (J.A.M. Demattê), giulio.genova@isric.org (G. Genova), gholizadeh@af.czu.cz (A. Gholizadeh), uta.heiden@dlr.de (U. Heiden), paul.karlshoefer@dlr.de (P. Karlshoefer), milewski@gfz-potsdam.de (R. Milewski), Laura.poggio@isric.org (L. Poggio), Marmar.sabetizadeh@uclouvain.be (M. Sabetizade), adsanz@gmv.com (A. Sanz), peter.schwind@dlr.de (P. Schwind), tsakirin@auth.gr (N. Tsakiridis), nziolas@ufl.edu (N. Tziolas), mjyague@gmv.com (J. Yagüe), zizala.daniel@vumop.cz (D. Žížala).

https://doi.org/10.1016/j.geoderma.2024.117113
Received 19 July 2024; Received in revised form 18 October 2024; Accepted 16 November 2024
Available online 26 November 2024

Nomenclature

BSF	Bare soil frequency
CAI	Cellulose absorption index
CHIME	The Copernicus Hyperspectral Imaging Mission
CNN	Convolutional neural networks
COG	Cloud Optimized GeoTIFF
DEM	Digital Elevation Model
DSM	Digital Soil mapping
ESA	European Space Agency
ESRIN	European Space Research Institute, also known as ESA Centre for Earth Observation
GEOCRADLE	Data Hub to access and share geospatial data and information collected from satellites and ground-based networks.
GIS	Geographic Information System
GUI	Graphic User Interface
ISO	International Standards Organisation
LPIS	Land Parcel Identification System
LUCAS	European Union Land Use/Cover Area Frame Statistical Survey
MEC	Model Efficiency Coefficient
MREF	Mean reflectance composite
NASA	The National Aeronautics and Space Administration of the United States federal government
NBR	Normalized Burn Ratio
NDVI	Normalised difference vegetation index
NIR	Near-Infrared
NPV	Non-photosynthetically active vegetation
NRC	National Reporting Centre on Soils
NRMSE	Normalized root mean squared error
OLI	Operational Land Imager of the Landsat 8 payload
PICP	Prediction Interval Coverage Probability
PIR	Prediction interval ratio
PLSR	Partial Least Squares regression, or Projection to Latent Structures,
PRISMA	Hyperspectral Precursor of the Application Mission
QRF	Quantile Random Forests
RE	Relative Error
RMSE	Root Mean Square Error
RPD	Ratio of performance to deviation (i.e. RMSE/Std)
RPIQ	Ratio of performance to interquartile range (i.e. RMSE/(Q3-Q1)
SBG	Surface Biology and Geology Mission
SCMaP	Soil composite mapping processor
SCORPAN	Soil, climate, organisms, relief, parent material, age, and site factors
SOC	Soil Organic Carbon
SRC	Spectral reflectance composites
SSL	Soil Spectral Library
SVM	Support Vector Machine
SWIR	Short-wavelength Infrared
TM	Thematic Mapper of the Landsat Mission
UKZUZ	Central Institute for Supervising and Testing in Agriculture in the Czech Republic
UTM	Universal Transverse Mercator
VIS	Visible spectral range
WORLDSOILS	ESA contract No. 400131273/20/I-NB to derive topsoil SOC predictions from satellite observations
WoSIS	World Soil Information Service (selection of standardised and ultimately harmonised soil profile data)

1. Introduction

The soil is the most important terrestrial carbon pool and therefore small losses or gains in SOC represent significant sources or sinks of atmospheric CO_2. The spatial variability of SOC is notable because of variations in driving factors: climate, soils, land use or historical management practices, amongst other factors such as microbial activity. Moreover, SOC content is required to calculate the SOC/clay ratio which has recently been recognized as one of the quantified descriptors in the EU soil monitoring law (Council of the European Union, 2024) as a parameter to evaluate soil structure. Earth observation plays an important role in monitoring SOC content (e.g. Angelopoulou et al., 2019; Vaudour et al., 2022). SOC is a heterogeneous material and does not show features in narrow spectral ranges. Instead, broad features occur in the visible range around 450, 590 and 664 nm and in the near infrared (NIR) and shortwave infrared (SWIR) regions, mainly related to lignin and cellulose (Chabrillat et al., 2019; Castaldi, 2021).

One of the main challenges for remote sensing of soils is the fact that the time windows during which the soils are bare (e.g. just before seeding) are relatively short and depend on the crop calendar for each field. Moreover, the soil surface conditions can disturb the signal reflecting the soil properties and hinder measurement standardization. Therefore, imagery acquired on a single date does not always provide a continuous and reliable map of soil properties (e.g. Vaudour et al., 2019).

In the last decade the soil composite mapping processor (SCMaP; Rogge et al., 2018) and similar techniques have been developed (Diek et al. 2017, Demattê et al., 2018). These techniques use the large multispectral satellite archives and can distinguish croplands and permanently vegetated areas (grasslands and croplands), retrieve bare soil spectral reflectance composites (SRC) and calculate statistics related to the length of the indicated period. This temporal compositing technique integrates the spectra for each pixel of exposed soils for a large number of satellite acquisitions during a defined time period. SRCs thus provide a spatial coverage of most croplands that could not be obtained in a single acquisition. After all, the windows when the soil is bare are short and cannot be exploited for all fields during a single satellite acquisition as the crops are in different stages of development. This results in various degrees of coverage of the soil. SCMaP provides bare soil composite spectra that together with an up-to-date geo referenced data base of soil characteristics can be used to construct SOC prediction models for croplands (Demattê et al., 2018; Safanelli et al., 2020, Safanelli et al., 2021). Moreover, Demattê et al. (2020) identified and nominated the bare soil frequency (BSF) while mapping at global scale. On the one hand, the authors used this BSF as an indicator of how many times a single pixel was bare and thus, could have an impact on soil degradation. On the other hand, the same BSF can be interpreted as a proxy for the crop cover dynamics during the period over which the satellite imagery was acquired and thus reflects the production of biomass and the input of carbon to the soil.

Safanelli et al. (2020 & 2021) used temporal soil composites of Landsat data, creating a median reflectance of a series of satellite acquisitions for each pixel. Although they were able to cover nearly all croplands for large areas in Europe and Brazil, these authors used long time series of Landsat imagery (36 years) and a somewhat coarse spatial and spectral resolution (30 m pixels and 6 Bands). This long integration period is neither optimal for detecting up-to-date SOC base line values, nor for monitoring, as required by soil health and climate mitigation policies. New sensors such as the Sentinel 2 constellation and the Landsat 8 multispectral instruments have recently become available with better spatial and spectral resolutions. The frequent overpass of these satellites ensures a nearly complete coverage of bare croplands for

temporal composites collected over a relatively short period. Castaldi (2021) has shown that three-year composites produced good quality SOC prediction algorithms covering a large part of the European territory.

For permanently vegetated areas such as grasslands and forests, a so-called digital soil mapping (DSM) approach can be used i.e. establishing a multi variate or machine learning prediction model for geo-referenced SOC contents using a variety of spatially continuous co-variates (Minasny & McBratney, 2016). DSM is a well-established approach to model and map soil properties at unknown locations, see for example Poggio et al., (2021). DSM techniques use legacy in situ soil data and relate them to spatially explicit environmental covariates describing the so-called SCORPAN (soil, climate, organisms, relief, parent material, age, and site) factors (Minasny & McBratney, 2016). So far, the use of high-resolution data products from optical sensors such as Sentinel-2 in digital soil mapping approaches are rare, although they showed potential in case studies or small-scale applications (Loiseau et al., 2019). The usefulness of innovative products from the Sentinel-2 archive, such as mean reflectance composites, soil reflectance composites and BSF information seems to be promising for digital soil mapping approaches. Indeed, BSF has had its importance corroborated by Sousa et al. (2024) on carbon impact monitoring.

As an example, many croplands have a negative carbon balance and have lost a large part of their carbon pool since the start of agriculture (Sanderman et al., 2017). Conservation agriculture aims at reversing this carbon balance and thus contributing to negative CO_2 emissions while at the same time increasing soil organic matter and restoring soil health (4 per 1000 initiative). For example, Castaldi et al. (2024) demonstrated that for a large number of fields in the Italian Po valley there was a clear spatial correlation between SOC and the farming systems, while the inclusion of fodder crops in the rotation and no till were the most effective practices. Many of the fodder crops such as alfalfa are multiannual and therefore cover the soil for (most of the) growing season. Our hypothesis is that increases in SOC content accompanied by decreases in exposure of bare soils indicate that on the one hand tillage intensity decreases and grades to no-till. This leads to concentration of SOC in a thinner topsoil horizon. On the other hand, C input from crops increases as the bare soil exposure decreases. The bare soil frequency product obtained from the temporal compositing approach indicates the percentage of bare soil acquisitions compared to the total number of acquisitions that exclude clouds, haze and snow. Due to the normalization with the total number of acquisitions, it is comparable across Europe even if the cloud cover conditions vary (Heiden et al., 2022). The bare soil frequency can be used as an indicator of activity i.e. cropland or permanently vegetated area or conservation agriculture in cropland i.e. length of the crop cover during the spring-autumn season. Moreover, the uncertainty of the SOC prediction in each pixel increases when the number of bare soil spectra included in the composite drops below 10 (Dvorakova et al., 2023).

Digital soil mapping has successfully been applied for SOC monitoring over regions (e.g. Cheng et al., 2024; Wang et al., 2024) to large countries (e.g. Australia, Wadoux et al., 2023). However, constraints have appeared. The co-variates, such as climate, topography and vegetation indices, generally reflect the driving factors for variability in SOC at rather large resolution (e.g. 300 m; Wang et al., 2024). For high resolution maps a geo-referenced large data set of SOC contents is required often to be collected over a rather long period (e.g. 1970–2020 in Australia; Wadoux et al., 2023). Moreover, some regions such as alluvial plains show little variability in co-variates (Wang et al., 2024) or are characterised by a combination of low sampling density and some extreme values of SOC (e.g. Scandinavia, de Brogniez et al., 2015). Spectroscopic approaches have demonstrated the potential for mapping SOC using satellite imagery of bare fields reducing the requirement for recent SOC data and exploiting the multispectral satellite signal to enhance the differences between cropland fields (e.g. Angelopoulou et al., 2019; Vaudour et al., 2022).

We present a pre-operational system that leverages the Sentinel 2 image archive, environmental covariates and the European land use/cover area frame statistical survey LUCAS soil data set (Tóth et al., 2013; Orgiazzi et al., 2018) to derive topsoil SOC content for the European continent. The system uses a novel concept of combining bare soil SOC prediction models with models for vegetated soils. The operational system covers the entire processing chain from the pre-treatment of the Sentinel 2 images, calibrating the algorithms for soil property prediction, integrating uncertainty estimations and producing combined maps for vegetated and bare soils. Results are provided for three consecutive years in a graphical user interface that supports soil monitoring. In the near future, such a system has the potential to evolve from an up-to-date baseline SOC assessment towards monitoring of SOC content and other soil properties. The system is standardized over Europe as it is calibrated on the LUCAS database with uniform sampling and analytical protocols. The system is scalable up to regions and even farm scale. The latter would require aggregation of the pixel predictions to the field level based on the field lay out using the European land parcel information system (LPIS; e.g. Samarinas et al., 2023).

The objectives of this paper are to describe the tests and validation of this operational system using a novel dual approach: (1) direct spectral soil mapping for soils that are visible in bare conditions and (2) digital soil mapping for those that are permanently vegetated. Specific technical objectives are:

- To provide up to date annual SOC content maps for Europe covering croplands, grasslands and forests
- To develop transparent SOC prediction models
- To validate results in three large pilot regions with external data from National Reporting Centres on Soils
- To enable the system to produce uncertainty maps for evaluating the relevance of spatial or temporal trends
- To provide an indicator for the length of the bare soil exposure during the growing season

The SOC monitoring system results are validated using external datasets in three European regions (Wallonia (Belgium), Central Macedonia (Greece) and the Czech Republic). Additionally, an example of the spatial trends for a region in northern Germany with a large SOC content variation is presented to demonstrate the validity and accuracy of the SOC predictions achieved exemplarily in a local area with large SOC variability.

2. Material and methods

The SOC monitoring system comprises a collection of modules that covers the whole processing pipeline from the Sentinel-2 archive exploitation to the final SOC content and uncertainty map (Fig. 1). It starts by accessing several external data sources which include all Sentinel-2 L2A reflectance data from 2018 to 2022 and the ESA WorldCover data set (Table S1). These data are used as input for the SCMaP to create temporal composite products (section 2.2). The SOC content values of the LUCAS soil data set and the Soil Reflectance Composite (SRC) are the input data for the subsequent modelling for bare soil pixels described in section 2.3.1. For soils that are always covered by vegetation, a digital soil mapping approach is followed integrating environmental covariates, novel products from SCMaP and the SOC content values from the LUCAS soil data set (section 2.3.2). Both models generate separate SOC content maps for bare soil pixels and for vegetated pixels at 20 m spatial resolution. The results are then mosaicked to get a combined SOC content and uncertainly map at 50 m spatial resolution for three pilot regions and 100 m pixels for Europe (section 2.4).

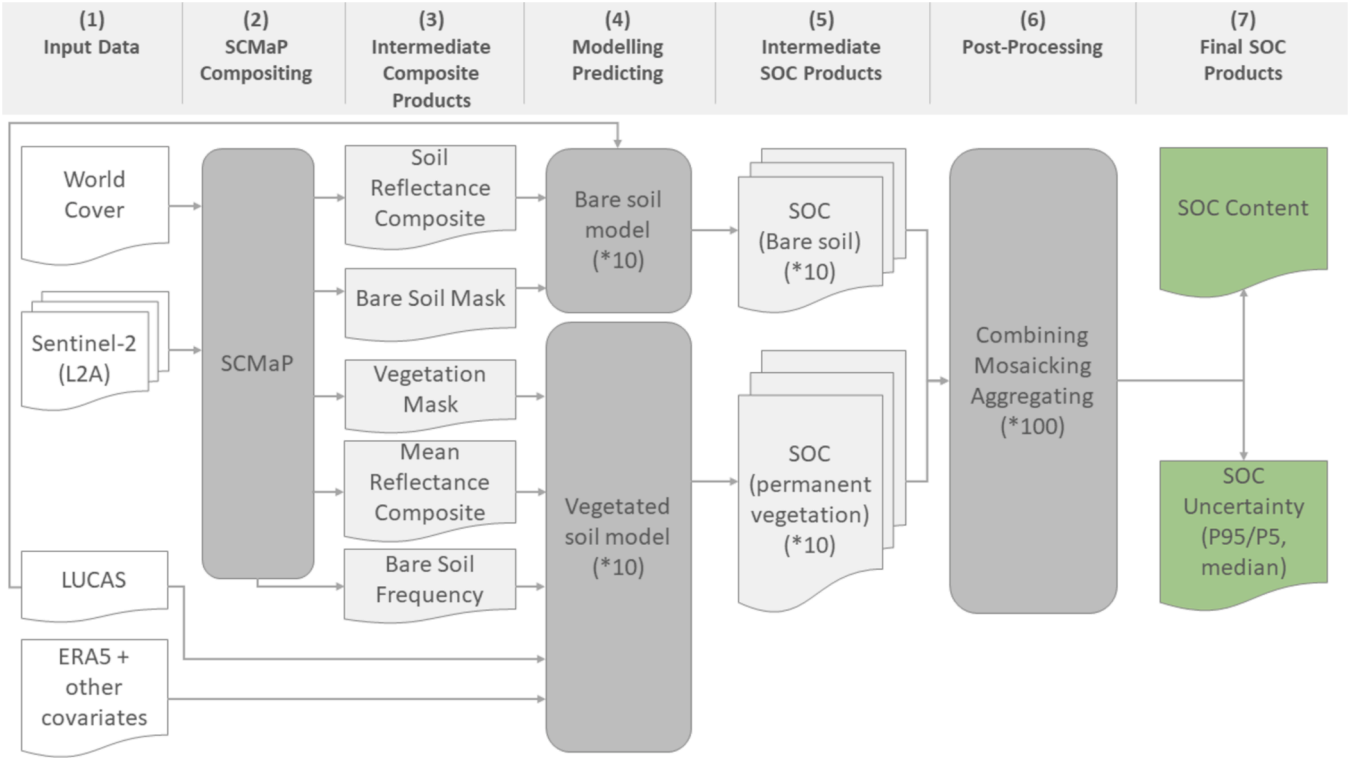


Fig. 1. Flowchart of the SOC monitoring system to derive topsoil SOC content at continental scale. The number in brackets indicates the number of iterations and resulting products with P5 and P95 being the 5th and 95th percentile.

2.1. *Architecture of the soil monitoring system*

The architecture for the EO based SOC monitoring system relies on a visual platform for operators to monitor and control the processing of SOC maps, as well as a user-friendly graphical interface for end users to view and download the maps produced by the system.

The design of the system adheres to a 3-tier architecture, a well-established software application architecture that organizes applications into three logical and physical computing tiers. In this design, application processing, and data management functions are independently developed and maintained in layers within each tier. This design allows developers to add or modify functional layers within a tier, eliminating the need to overhaul the entire application. Thus, the system is divided into three physically separate tiers, each further segmented

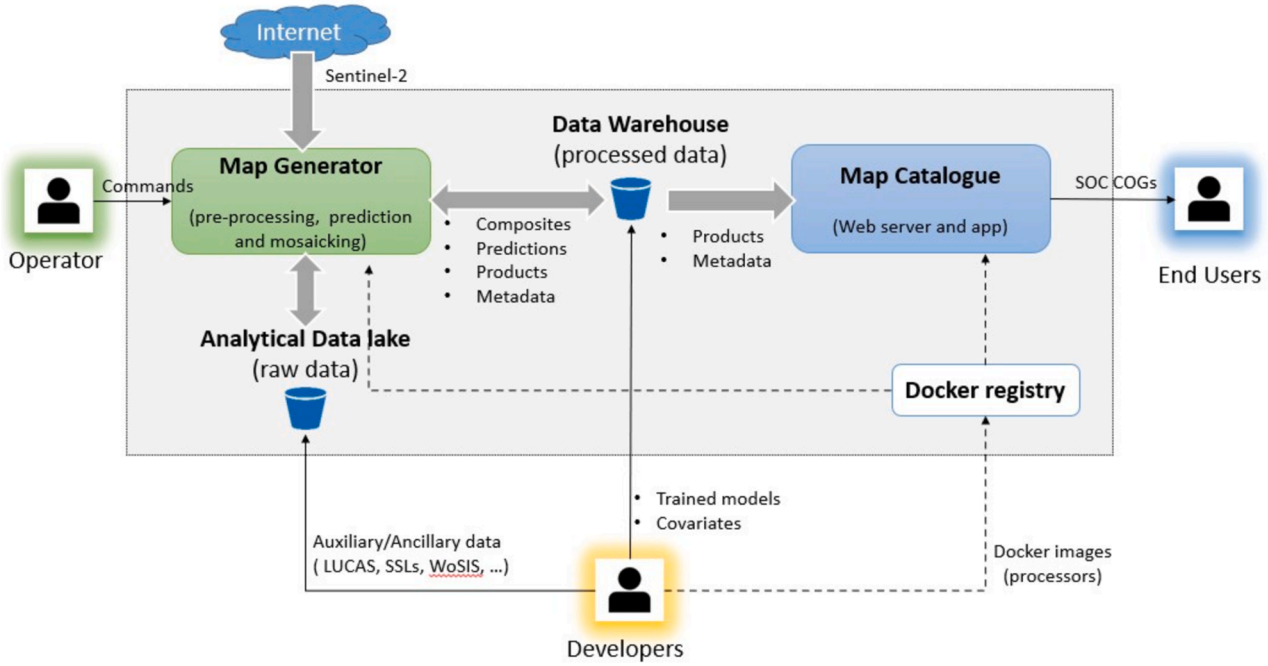


Fig. 2. Detailed modular architecture of the SOC prediction system and actors *(See abbreviations table after Conclusions)*.

into logical layers: Tier-1 or presentation tier which houses the *Map Catalogue* front end; Tier-2 or application tier which holds the *Map Catalogue* back-end and the *Map Generator*; and Tier-3 or data tier which operates the *Analytical Data Lake* and the *Data Warehouse*. This architectural choice was made to facilitate component reusability in future iterations of the system and to enable concurrent development of tiers by different actors: developers responsible for developing prediction models and processors, operators responsible for system configuration and triggering prediction processing, and end users.

There are four primary system components (Fig. 2): **(i)** The *Map Generator* which handles the back-end processing necessary to generate SOC predictions. It is supervised by a human operator who configures the execution of the SOC Prediction pipeline (i.e., the EO data pre-processing, soil prediction, mosaicking, and aggregation tasks). **(ii)** The *Map Catalogue* that serves as the front-end graphical user interface. This component provides the end users with 1) a map service to visually inspect the SOC prediction layers and product metadata (generated by the *Map Generator* component) and 2) a download interface via web browser to download SOC map layers in Cloud Optimized GeoTIFF (COG) format. **(iii)** The *Analytical Data Lake*, a storage system leveraging a cloud storage service, holding raw and/or granular data for analytical purposes. It holds big data from various sources with high accessibility, e.g., LUCAS, digital elevation models (DEM), and climate data from ERA5 (Tables S1-S3). **(iv)** The *Data Warehouse*, another cloud storage system storing structured processed data used for the end-user graphical interface and intermediate processing. It houses trained models, covariates, image composites, intermediate predictions, SOC output products, and metadata.

The back-end processing (Map Generator) utilizes a suite of processors developed by the WorldSoils project team, implemented as Docker images.

2.2. Temporal compositing approach using SCMAP

The advantage of temporal compositing is to (1) generate results that reflect not just one single point in time but a temporal average that balances out strong seasonal dynamics and preserves permanent differences between regions; (2) in case of bare SRC, it enlarges the area for which direct soil modelling can be applied. The larger this area, the better the prediction results since direct spectral models show mostly better cross-validation results. The challenge is to develop a rigorous pixel selection process to collect undisturbed bare soil pixels by covering an area as large as possible at the same time.

A summary of the temporal compositing based on SCMaP (step 2 in Fig. 1) follows, aligned with Rogge et al., (2018), Heiden et al., (2022).

In principle, four main products are generated (see Table 1) using the archive of Sentinel-2 based on the methodology shown in Fig. 3. The bare surface and the vegetation mask are used to assign each pixel of the processed area to the correct SOC model. The mean reflectance composite and the bare surface frequency are inputs for the digital SOC mapping for the vegetated surfaces. The bare SRC will be utilized in the spectral SOC mapping for the bare soils. For the bare soil and surface selection, a combined spectral index (PV+IR2) is used to separate photosynthetically active vegetation (PV) from bare surfaces (Normalized Difference Vegetation Index – NDVI, Rouse et al., 1974) and to reduce the influence of non-photosynthetically active vegetation (NPV) using the Normalized Burn Ratio (NBR, (García and Caselles, 1991; Key & Benson, 2003; eq. (1).

$$PV+IR2 = ((B8 - B4)/(B8 + B4)) + ((B8 - B12)/(B8 + B12)) \quad (1)$$

where Bx is the reflectance of the used Sentinel-2 band. Initially, thresholds need to be defined, which are calculated per UTM-Tile and analysing the time period 2019–2021. The regionalisation of the thresholds is necessary to reflect the differences in the environmental conditions of the region that emerged from a different climate, terrain properties and thus, affects the characteristics of the land cover. For each UTM tile, the minimum index composite is built by collecting for each pixel the minimum PV+IR2 value in the time period. The frequency distribution of spectral index values for the agricultural fields are then compared with the frequency of grasslands by overlaying the land cover classes grassland and cropland from the WorldCover data set for 2020 and 2021. These two landcover classes are used to optimize the separation between PV and NPV. In general, the clear separation of both land covers is not possible due to the limited spectral information of Sentinel-2. The histogram separation threshold method (Heiden et al., 2022) is then used to derive a tile-based threshold that is further the baseline to select bare surfaces.

Table 1
Specification of products generated using the Soil Composite Mapping processor (SCMaP).

Product	Description
Bare surface and the vegetation mask (MASK)	The 3-band mask contain information about: Band 1 – pixels with alternating cover from soil exposure to vegetation Band 2 – pixels with permanent vegetation cover Band 3 – pixels with permanent non-vegetation cover. The spatial resolution is 20 m.
Mean reflectance composite (MREF)	The mean reflectance composite in [reflectance values*10.000] is the mean of all valid pixels (vegetation, no vegetation, bare soils and it excludes clouds, snow, etc.). All original 10 m and 20 m bands of Sentinel-2 are resampled to 20 m.
Bare surface frequency (BSF)	This image contains information about the bare surface frequency and is scaled between 0 and 1. It quantifies the fraction of bare surface pixels compared to the total number of valid pixels. It is a measure for the use intensity of the soils.
Bare surface reflectance composite (SRC)	The bare surface reflectance composite in [reflectance values*10.000] is the mean of all bare surface reflectance pixels (except of clouds, haze, snow, etc.). All original 10 m and 20 m bands of Sentinel-2 are resampled to 20 m.

For the WorldSoils system, level 2A reflectance data including their mask product (scene classification layer) are used for the processing. For the exploitation of the Sentinel-2 archive, a rigorous cloud handling is necessary that includes several filter steps. Initially, only level 2A scenes are selected for processing that have a cloud coverage < 80 %. For scenes that are still contaminated with clouds, the scene classification layer of the Sentinel-2 L2A product generated during the atmospheric correction stage is used to filter out clouds, haze and snow by keeping all vegetation (class 4), no-vegetation (class 5) and water (class 6) pixels only. In a later stage of the processing, selected bare surface reflectance pixels are further masked out by two soil-specific filters. First, due to the distinct difference in the NIR and SWIR between clouds and almost all soils, pixels with the following conditions are masked out as clouds:

$$(B11 - B8A)/(B11 + B8A) > 0.02 \quad (2)$$

where Bx is the reflectance of the used Sentinel-2 band. Second, remaining haze and thin cloud contamination is detected based on higher blue reflectance. For this purpose, a local statistic outlier filter is implemented:

$$b \leqslant median(B) + 3\sigma; \sigma = 1.48 median(|B - median(B)|) \quad (3)$$

where B is the vector of all blue reflectance values in the time series and b the blue reflectance to be tested. Further, the Normalised Difference Snow Index (NDSI) is applied as defined by ESA (Sentiwiki, 2024) With this rigorous cloud and snow filtering, it is possible to average the pixels based on the mean. This has been done for every valid pixel in the time-period (mean reflectance composite) and for the bare surface/soil reflectance composites. Additionally, statistical products such as the bare surface frequency are derived. The latter is the ratio between the

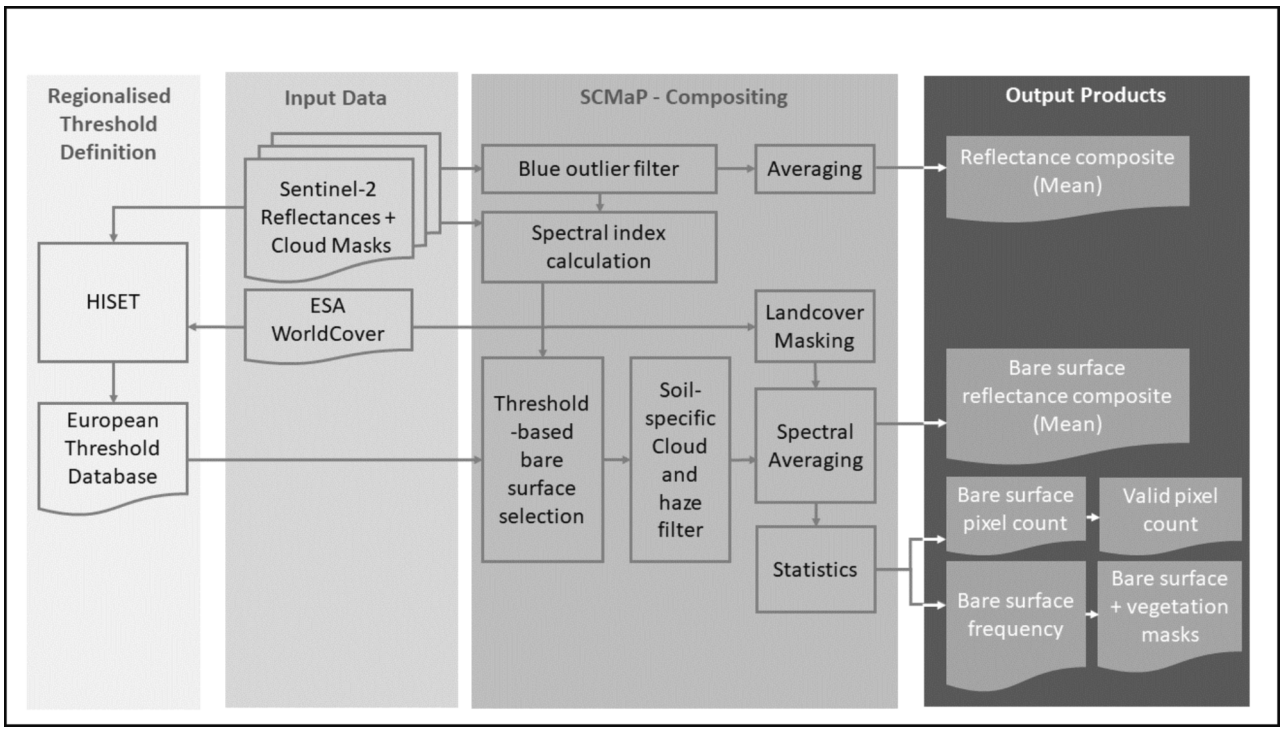


Fig. 3. Soil Composite Mapping processor (SCMaP) – temporal compositing flowchart.

bare surface pixel count product and the valid pixel count product.

For the WorldSoils system, the above-described composite products are generated for three different time periods with the same processing parameters (Table 2). For each of the time ranges, we used the spring (03, 04, 05) and autumn month (08, 09, 10) only. As described in Heiden et al. (2022), the winter month have been excluded due to the high solar zenith angle. For angle higher than 70°, the estimation of the atmospheric parameter (water vapour, aerosols) result in high uncertainties due to the low energy level that is received by the Sentinel-2 sensor (Vermote et al., 2016). The summer months June and July do not contain a lot of bare soil fields due to the ripening phase and thus, they are also excluded.

2.3. Data sets

2.3.1. Data sets for training the models

The LUCAS topsoil dataset is a large European soil spectral library (about 20,000 soil samples). The campaign was launched in 2009 and the sampling was repeated in 2015 and 2018 in the framework of the European land use/cover area frame statistical survey (Tóth et al., 2013; Orgiazzi et al., 2018). The topsoil sampling (0–20 cm) locations were selected using a Latin hypercube-base stratified sampling design from the LUCAS master sample grid of 2 km by 2 km. The LUCAS dataset consists of laboratory spectra and 12 chemical and physical variables for each soil sample that were acquired in the same laboratory. To train the bare soil model, first the bare soil reflectance composite for each of the LUCAS 2015 and LUCAS 2018 cropland points was extracted after examining a three-year period from 2018 to 2021 and calculating the mean over the bare soil signatures. It should be noted that if a LUCAS point was revisited in 2018, we kept only its most recent SOC analysis. In addition to the actual points themselves, to augment the data, a 3x3 grid (i.e., pixels each corresponding to 20 m of spatial resolution) around each point was constructed and its eight adjacent neighbours were also collected. This resulted in about 71,235 unique spectral signatures, each associated with either a neighbour or the central pixel corresponding to a LUCAS topsoil measurement. The assumption behind our effort to collect the neighbours originates from the fundamental principle that adjacent pixels ought to have similar SOC value (or very close to it).

Thus, to account for the inherent imbalanced distribution of SOC that has high positive skewness, it is possible to collect extra training data for the large SOC concentrations. For this reason, we kept only the central pixels from soil samples corresponding for which SOC is less than 50 g C kg^{-1} and kept both central and neighbouring pixels for which SOC content is equal to or exceeds 50 g C kg^{-1}. In total, 13,949 records formed the final set, the distribution of which is depicted in Fig. 4. Following this step, a spectral-based outlier detection methodology was employed to remove a few samples which had outlying nature, even if they passed from the bare soil filter, to ensure that the highest-quality data remained. This is necessary to account for unforeseen anomalies which may be present in the spectral signatures, particularly at points that are detected infrequently as bare, which could not be detected by the process used to select only the bare fields. To this end, the one-class Support Vector Machine (SVM) algorithm was employed in the first two principal components (Seliya et al. 2021). In one-class SVMs, outlier detection is achieved by learning a high-dimensional hypersphere that tightly encloses the training data, representing the expected normal

Table 2
Processing parameters for the Soil Composite Mapping processor (SCMaP) approach.

Processing parameter	Value
Sentinel-2 Level 2A input	Sen2Cor processor
Sentinel-2 Level 2A cloud masking	Scene classification layer: classes 4 (vegetation),5 (no vegetation) and 6 (water) are kept
Time range (years)	2018–2020, 2019–2021, 2020–2022
Time range (month used in each time period)	03, 04, 05, 08, 09, 10
Cloud cover	< 80 %
Spectral index	PV + IR2 (see equation (1)
Threshold	Variable per Sentinel-2 tile
Normalized Difference Snow Index	0.00 (ESA, Sentiwiki, 2024)
Minimum bare soil count	3

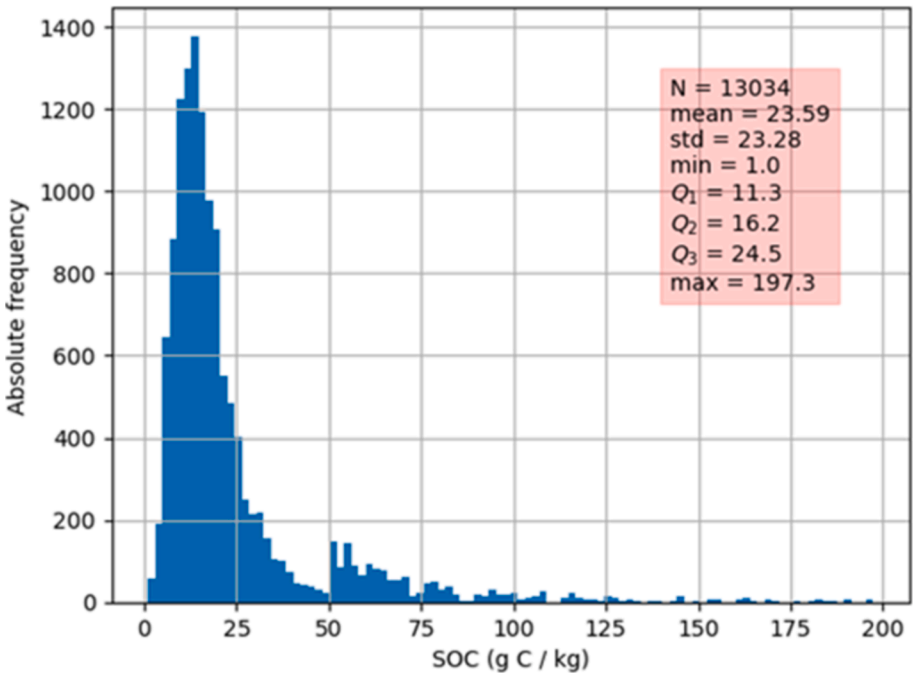


Fig. 4. SOC content distribution of the data collected after using the 3x3 neighbouring grid for high (i.e., >=50 g C kg^{-1}) SOC content values after removing the outliers. Statistical parameters are in g C kg^{-1}.

behaviour. This approach maximizes the margin between the data and the boundary, effectively isolating outlying points that fall outside the learned decision surface. The hyperparameter of SVM is nu, which is an upper bound on the fraction of training errors and a lower bound of the fraction of support vectors (Chang and Lin, 2001). We used a nu value 0.02 and employed a linear kernel, after visually checking the samples that were detected as outliers. After using the one-class SVM algorithm, 13,034 samples were left as inliers, and we proceeded to use them for modelling (Fig. 4).

With respect to the permanently vegetated model, additionally, a set of about 40 covariates was prepared as candidate predictors based on their likely influence on soil formation for the permanently vegetated areas. The mean reflectance composites (Table 1) for all pixels were included in the covariates set. An overview of the covariates is given in the supplementary materials. The World Soil Information Service (WoSIS; Batjes et al., 2024) provides a compilation of quality-assessed and standardised soil profile data that can be used to support digital soil mapping and environmental applications at broad scale levels. The data come from 174 countries and represent more than 900 k soil layers (or horizons) and over 6 million records. The number of measurements for each soil property vary (greatly) between profiles and with depth, this generally depending on the objectives of the initial soil sampling programmes. For this study the data for the region of interest were selected and used in addition to the LUCAS dataset.

2.3.2. *Data sets for evaluation of the SOC monitoring system in three pilot regions*

The Earth Observation-Soil Monitoring System has been validated in three contrasting regions: Wallonia, Belgium; Macedonia, Greece and the Czech Republic. These three focus studies include the use of local/regional reference data, independent form the training data used in the development of the SOC prediction algorithms and the derived relationship which are compared to the use of the default reference relationship based on the LUCAS topsoil database (Supplementary material). Different analytical methods are used to measure SOC in the pilot regions. Most frequently used methods are wet oxidation also called Walkley-Black method (Walkley & Black, 1934) or its modifications, and dry combustion, e.g., with a CN-analyser (direct analyses), as well as near infrared reflectance spectroscopy (NIR calibrated to SOC contents analysed with a CN analyser). The use of different methods and laboratories in the pilot regions can result in a bias compared to the LUCAS survey that used a standardized method in a single reference laboratory (ISO 10694:1995).

2.4. *SOC prediction algorithms*

2.4.1. *Spectroscopic models for bare (cropland) soils*

To develop the model from the SRC generated from the multi-temporal Sentinel-2 data which predicts topsoil SOC content in bare soils (corresponding to croplands), various AI modelling pipelines were tested. First, we considered standard machine learning models such as the Random Forest (Breiman, 2001) and Extreme Gradient Boost (XGBoost) learners (Chen & Guestrin, 2016) which are ensemble methodologies that work well on tabular data. Their hyperparameters were optimized using 5-fold cross-validation in the calibration set (i.e., data available for training only) and grid search to obtain the most robust models. These models exhibited relatively low performance given the highly imbalanced nature of the dataset, due to SOC's inherent high positive skewness where most of the patterns have low SOC content. To overcome this limitation, imbalanced learning techniques were also considered such as under-sampling, over-sampling, and custom weighted loss functions (e.g., weighted Random Forest). At this stage, we also implemented custom artificial neural networks and convolutional neural networks (CNNs) that were optimized using Bayesian optimization (Schmidinger & Heuvelink, 2023) while additionally considering custom loss functions to account for the imbalanced nature of the dataset. A 10-fold cross-validation strategy in the calibration set was employed to find the optimal hyperparameters by checking the metrics of the internal validation set of each fold. The folds were spatially stratified in the geodetic domain. Ultimately, the best model was the CNN model with a custom loss function and oversampling of the minority class (as detailed in the sections below), that outperformed all the other approaches and was used in the end.

The custom CNN model is a multi-input one-dimensional model which uses as input simultaneously the SRC (which corresponds to a mean bare soil reflectance) and the Standard Normal Variate

transformation thereof, each of length 10 (i.e., the 10 Sentinel-2 bands after removing B1 used for coastal aerosol, B9 that targets water vapor, and B10 that detects cirrus clouds). It then utilizes three convolutional layers to successively extract higher level features, with 18, 60, and 98 filters and a kernel size of 4. After flattening, 7 fully connected layers follow (with 200, 240, 168, 200, 160, 256, and 224) and then final layer produces the predicted value. The activations functions in each layer are a mix of linear and leaky ReLU (rectified linear unit), while the final layer uses the tanh activation function ensuring that the output is always constrained with the bounds seen from the training dataset. The model was trained using the Adam learner and a learning rate of 0.00156 with a batch size of 128 (Kingma & Ba, 2014).

2.4.2. Digital soil mapping

SOC in permanently vegetated areas was predicted using a digital soil mapping (DSM) approach. DSM is a well-established approach to model and map soil properties at unknown locations, see for example Minasny & McBratney (2016). DSM techniques use legacy in situ soil data and relate them to spatially explicit environmental information describing the so-called SCORPAN (S: soil, C: climate, O: organisms, R: relief, P: parent material, A: age and N: site, location) factors. Among these SCORPAN factors, organisms (within soil and above) play a dominant role in explaining the variability in SOC at the regional to landscape scale. A statistical relationship between measured soil properties and soil forming factors (e.g. terrain, vegetation, climate) as measured by environmental covariates will be established (see the covariates in Table S3). The layers were selected from a variety of open source datasets available for the test area. Remote sensing products are particularly effective in providing covariates characterising the role of organisms i.e. cropping systems such as land use (change) maps and indicators for C input from the vegetation (e.g., NDVI, Poggio et al.2013). SOC observations at various locations in Europe were related to the previously described environmental covariates. The soil observations were split in 10 equally sized folds. Model tuning was performed with a 10-fold cross-validation procedure applied to multiple combinations of hyper-parameters. The main method explored was Random Forest (Breiman, 2001). Models were obtained with the ranger package in R (Wright & Ziegler, 2017), with the option quantreg to build Quantile Random Forests (QRF; Meinshausen, 2006), for spatial uncertainty assessment. With this option, the prediction is not a single value, e.g., the average of predictions from the group of decision trees in the random forest, but rather a cumulative probability distribution of the soil property at each location. The importance of the variables was calculated using the Gini impurity Index (Ishwaran, 2015). A set of about 40 covariates was prepared as candidate predictors based on their likely influence on soil formation. The soil composite averages for all pixels were included in the covariates set.

The number of decision trees (ntree) was set as 500 and the number of covariates used in the tree splits (mtry) as the (rounded) square root of the number of covariates. Ten-fold cross-validation was used to assess the internal performances of the model. The observations in each fold were spatially stratified in the geodetic domain. The final model was fitted with all available data. The final prediction models were generated using the ranger package as computationally optimized implementation of Random Forest. The option quantreg was used to build quantile regression forests (Meinshausen, 2006). This yields a cumulative probability distribution of topsoil SOC content at each location, thus also quantifying prediction uncertainty.

2.5. Mosaicking

The inputs for the mosaicking are a raster created from the soil and permanently vegetated masks produced by SCMaP and the SOC predicted in croplands and permanently vegetated areas. The model output was a weight surface that can be used to calculate a SOC content in each pixel to be used as a contiguous map. The weight surface was obtained using a transition function around the border of the vegetated and bare soil predictions. We used a linear transition, with the weights defined in equation (4).

$$weight = (\max\_dist + 2) - buffer/(\max\_dist + 1) \quad (4)$$

where max_dist is the maximum buffer distance in number of cells.

This linear weight surface is transitioning from 1 inside the region to 0 in the cells farther from the region border. The same procedure was used for the uncertainty parameters (expressed as the ratio of the 90 % percentile and the median ((P95-P5)/median). The mosaicking step was carried out at 20m resolution for each of the 100 combinations. The combinations were derived from the 10-fold bootstrapping iterations for both bare and vegetated soil predictions. The 100 combinations were aggregated (resampled) at 50 and 100 m resolution, respectively for the test areas and the European scale.

2.6. Model evaluation and validation

2.6.1. Model attribution testing

The SOC data provided by the National Reporting Centres (NRC; see section 2.3.2; Figures S2, S4 and S6) was sampled in agricultural soils during the last decade, i.e. 2019–2021 for Wallonia (Figure S1), 1995–2022 for Macedonia (Figure S3) and from 2018 onwards for the Czech Republic (Figure S5). The samples were geo-referenced and the field technicians from the NRC's distinguished the land cover types at the time of sampling i.e. cropland (mainly annual crops), tree crops and grassland. We used these observed land use data to test the accuracy of the attribution to the correct SOC prediction algorithm using a confusion matrix. As can be seen in steps 3 and 4 of Fig. 1, annual crops should be attributed to the bare soil model and grasslands and tree crops to the vegetated soil model. We calculated the user's, producer's and overall accuracy according to the following theoretical example where A-F are number of observations in the classes (Table 3).

2.6.2. Cross validation of algorithms

This approach was used to perform an evaluation of the models' performance using the reference data (i.e., LUCAS and WoSIS). A tenfold cross validation was considered crucial for robustly assessing the model's generalization ability and minimizing the risk of overfitting. We divided the dataset into ten subsets, with nine of them used for training and one for testing in each iteration. The accuracy metrics were calculated in the test set of each fold and their mean values served as our primary performance indicators such as model efficiency coefficient (MEC), which is the equivalent of the R^2 of predicted and observed values computed against the 1:1 line, root mean square error (RMSE), normalized root mean square error (NRMSE), bias, relative error (RE), Ratio of performance to deviation (RPD) and Ratio of performance to interquartile range (RPIQ); eqs. (5–11).

Table 3
Theoretical Confusion matrix for algorithm attribution.

	Reference				
Classified		cropland	Tree crops+grassland	Others	User's accuracy
	Bare soil	A[1]	B	0	A/(A + B)
	Vegetated soil	C	D	0	D*/(C + D)
	others	E	F	0	–
	Producer's accuracy	A/(A + C + E)	D/(B + D + F)	Overall Accuracy	(A + D)/(A + B + C + D + E + F)

[1] A-F are number of observations in the classes.

$$MEC = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (5)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\widehat{y}_i - y_i)^2}{n}} \quad (6)$$

$$NRMSE = \frac{RMSE}{\bar{y}} \quad (7)$$

$$\text{Bias} = \sum_{i=1}^{n}\frac{y_i - \widehat{y}_i}{n} \quad (8)$$

$$RE = \frac{\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right|}{n} * 100 \quad (9)$$

$$RPD = \frac{SD}{RMSE} \quad (10)$$

$$RPIQ = \frac{IQ}{RMSE} \quad (11)$$

where $\widehat{y}$ = predicted value, $\bar{y}$ = mean observed value, y = observed values, n = number of samples with i = 1, 2, …, n, and SD the standard deviation and IQ the interquartile range (quartiles 0.25 and 0.75) of the observed values.

This method allowed us to obtain ten values per each accuracy metric, from which we calculated the mean. By averaging the values across the ten folds, we obtained a more reliable estimate of the model's overall performance, providing a scientific and precise evaluation of its predictive capabilities. This approach enhances the validity of our findings and strengthens the confidence in the model's efficacy.

2.6.3. Monitoring and uncertainty assessment

Considering that CNNs are considered as black-box models in the context of explainable AI, we used a post-hoc interpretability mechanism, namely the Shapley values, to interpret the CNN model (Schwalbe & Finzel, 2024). We attributed importance scores to each input feature, highlighting which features most influence the model's predictions. These values are calculated via the SHapley Additive exPlanations (SHAP) algorithm which computes the contribution of each feature based on coalitional game theory (Lundberg et al., 2017).

To assess the uncertainty of prediction of the two models, the PICP (prediction interval coverage probability) levels for various PI (prediction interval) levels are calculated, to ascertain if the uncertainty estimations are reliable (Safanelli et al., 2020; Schmidinger & Heuvelink, 2023). The underlying idea is to evaluate what percentage of soil samples from the test set lies in its respective prediction interval. In the ideal scenario, the PICP is equal to the PI-level, e.g., for a 90% PI we desire a PICP of 90%. If the PICP is over the PI-level, then the model is overly pessimistic in its uncertainty estimations, whereas in the opposite case the model is optimistic.

The two different learning algorithms (namely the CNN-based model for bare soil predictions and the QRF for permanently vegetated areas) used two different approaches to produce estimates of the uncertainty. On the one hand, QRFs by design provide not only the mean of the response variable, as conventional RF does, but also its entire distribution; thus, the task of extracting the quantiles of predictions is simple. On the other hand, for the CNN model a different approach was followed. Ten different models were trained using ten different bootstrap iterations of the training dataset; each model provided its own prediction, which were then stored separately.

Furthermore, to visualize the uncertainty as a map in the Worldsoils system, we calculated the prediction interval ratio (PIR; eq. (12).

$$PIR = (P95 - P05)/P50 \quad (12)$$

where P05, P50, and P95 are the 5th, 50th, and 95th percentile of prediction; P50 is the median.

2.6.4. Evaluation of the monitoring system in the pilot areas

The three pilot regions i.e. Wallonia (Belgium), Macedonia (Greece) and the Czech Republic are described in detail in the supplementary materials. For the validation we extracted the predicted SOC value of the pixel corresponding to the location of the sample in the data sets of the National Reporting Centres on Soils. The location of the samples is given as the centre of the field. The spectrum of the pixel corresponding to the recorded location of the soil sample is then linked to the measured SOC content. Unfortunately, the protocols for taking composite samples and recording their location varied between the pilot regions as explained in the next paragraphs. Therefore, we did not weight the spectra to represent an average of the field. An additional limitation is that we did not have access to the land parcel identification system (LPIS) for all pilot regions.

For each of the three years of our study the validation samples in Wallonia (Belgium) were collected by the network of agricultural laboratories (REQUASUD) and reached more than 10,000 with a mean SOC content of 17.4–18.0 g C kg^{-1} and a standard deviation of 10.6–11.6 g C kg^{-1} (Table S4 and Figure S2). Each soil sample was collected at the field-scale as a composite of ten sub-samples. As the sample points are poorly distributed across the region, we applied a stratified random selection. Samples were selected in such a way that they would be distributed according to the cropland or grassland surface throughout the entire region. Moreover, we removed the samples in a specific land cover type that were either attributed to the wrong model or for which the SCMaP could not attribute a model type (i.e. others; see section 2.6.1). This allowed to focus on the performance of the SOC prediction algorithms. The measured SOC contents were determined from samples collected in the middle of the period for which the soil reflectance composites were made (e.g. samples collected in 2019 for the 2018–2020 composite).

The reference soil data from the croplands of the Macedonian region are shown in Figure S4. It consists of 2208 samples from agricultural land (e.g., arable land and tree crops). The data originated from two different sources: i) National Paying Agency in conjunction with the Aristotle University of Thessaloniki – Laboratory of Applied Science; and ii) GEOCRADLE EU Initiative (Tziolas et al., 2019). The data were collected between 1995 and 2022, following the same principles (5 m radius, 0–30 topsoil). The SOC content is in general low with a mean of 12.60 g C kg^{-1} and a standard deviation of 5.79 g C kg^{-1}.

The dataset used for validation in the Czech Republic (Figure S6) comes from routine agrochemical soil testing (AZZP) carried out by the Central Institute for Supervising and Testing in Agriculture (UKZUZ). The testing is repeated in the period of six years on the whole area of agricultural soils (last cycle 2014–2019). Every year ca. 60,000 samples are collected (SOC only ca. 5000 per year based on a stratified selection according to soil type; Figure S6). SOC monitoring started in 2018, and since then about 20,000 samples for SOC have been collected. One composite sample consists of min. 30 individual samples from an area of 7–10 ha (coordinates known only for centroid of area). SOC is obtained by spectroscopy in the NIR spectral region (1000–2500 nm) using reflectance values.

3. Results

SOC contents for the 0–20 cm topsoil were predicted based on reflectance composites of a moving interval of three years: 2018–2020, 2019–2021, 2020–2022. The Graphical User Interface (GUI; https://world-soils.com) displays the SOC content maps for these three time periods covering Europe at 100 m resolution and the three pilot

areas at 50 m resolution are available on website **GUI** together with maps of the 90% uncertainty intervals.

3.1. The soil reflectance composites

3.1.1. Model attribution

For Wallonia, the overall accuracy of the bare soil vs vegetated soil distinction of the Worldsoils model was very high (93.6%; Table 4). For Macedonia, the confusion matrix displays that most of the tree crops were indeed identified as permanently vegetated (830 / 995 and a producer's accuracy of 83.4 %) while annual crops had mixed results (719 / 1213 and a producer's accuracy of 59.3 %). The crop type "olive groves", one of the main tree crops, will always be attributed to the permanently vegetated SOC prediction model due to (1) the evergreen character of the vegetation and (2) the varying spectral mixtures with photosynthetically active and non-active vegetation. Therefore, spectrally pure soils will not be visible in the Sentinel-2 data. Other tree crops such as vineyards and orchards are also represented by spectral mixtures and thus, not attributed directly to the bare soil prediction model. Most of the annual crops which were identified as permanently vegetated are cereals, which may exhibit high percentage of plant residues in the months in which the soil may typically be bare. For the Czech Republic, tree crops consisted mainly of vineyards, hop farms and orchards. The overall accuracy of distinguishing bare soil from vegetated soil of the Worldsoils model was very high (96.7%). Therefore, the influence of the inaccuracy of the classification and possible inadequate use of the bare soil model (97.8 % producer's accuracy) and the permanently vegetated model (87.3 % producer's accuracy) is limited.

3.1.2. Bare soil count and frequency

SCMaP also calculates the bare soil count in croplands (i.e. the number of bare soil acquisitions for each pixel) and the bare soil frequency (i.e. the number of bare soil counts divided by the number of valid pixels; Table 5). It is considered that many bare soil acquisitions in the composite for a pixel result in a more stable signal where disturbances related to moisture and crop are minimal. Dvorakova et al., (2023) have demonstrated for a region covering parts of the Netherlands and Belgium that the uncertainty of the SOC content predictions strongly increased when the number of bare soil acquisitions in the composite of a pixel dropped below 7. It can already be seen in the confusion matrix that about 40 % of the validation samples in Macedonian croplands (mainly cereals and fodder crops) are classified by the SCMaP as permanently vegetated areas (Table 4). Moreover, the median bare soil count in Macedonia is around the critical value for model uncertainty identified by Dvorakova et al. (2023), while only 25% of the validation points reach this critical value in the Czech Republic and Wallonia (Table 5).

The bare soil frequency can be compared across regions as it is normalized by the number of valid pixels (Heiden et al., 2022). It is a proxy of the crop cover during the period of compositing i.e. from March until October. The lowest bare soil frequency occurs in Macedonia followed by the Czech Republic and Wallonia (Fig. 5).

3.2. SOC prediction algorithms

3.2.1. Bare soil

The ten-fold cross-validated model evaluation metrics are shown in Table 6, while the scatter plot of the out-of-fold predictions is shown in Fig. 6, where we collated the predictions across the held-out sets in cross-validation to form one comprehensive set. They both demonstrate

Table 5
Bare soil count statistics of the pixels for cropland validation sample points. Q1: first quartile; Q3: third quartile.

Region	Year	Min	Q1	Median	Q3	Max	Std
Wallonia	2019	0	13	23	31	58	11.86
	2020	0	11	19	28	61	12.09
	2021	0	15	25	34	60	13.48
Czech Republic	2019	0	11	20	29	59	13.23
	2020	0	0	10	20	63	12.64
	2021	0	7	16	25	51	11.76
Macedonia	2019	0	0	7	18	111	13.27
	2020	0	0	9	20	104	13.33
	2021	0	0	10	20	100	13.10

Table 4
Confusion matrix for the attribution of pixels to the bare soil and vegetated pixels based on the 2019–2021 composite. The number of pixels corresponding to the validation points is given together with the Producer's, user's and overall accuracy in percent (see Table 3).

A		**Reference**				
Classified	**Wallonia**	**Croplands**	**Grasslands**	**Others**	**Total**	**User's accuracy (%)**
	Bare soil [1]	7,886	108	0	7,994	98.7
	Permanently vegetated[2]	495	1,666	0	2,161	77.1
	Others[3]	26	21	0	47	–
	Total	8407	1795	0	10,202	
	Producer's accuracy (%)	93.8 %	92.8 %	–	**Overall accuracy (%)**	93.63

B		**Reference**				
Classified	**Macedonia**	**Croplands**	**Tree crops**	**Others**	**Total**	**User's accuracy (%)**
	Bare soil	719	141	0	860	83.6
	Permanently vegetated	463	830	0	1,293	64.2
	Others	31	24	0	55	–
	Total	1,213	995	0	2,208	
	Producer's accuracy (%)	59.3	83.4	–	**Overall accuracy (%)**	70.2

C		**Reference**				
Classified	**Czech Republic**	**Croplands**	**Grasslands + Tree crops**	**Others**	**Total**	**User's accuracy (%)**
	Bare soil	4,496	63	0	4,559	98.6
	Permanently vegetated	102	440	0	542	81.2
	Others	1	1	0	2	–
	Total	4,599	504	–	5,103	
	Producer's accuracy (%)	97.8	87.3		**Overall accuracy (%)**	96.7

[1] See Fig. 1 for workflow and section 2.4.1 for algorithms for bare soils.
[2] See Fig. 1 for workflow and section 2.4.2 for the algorithms for permanently vegetated areas
[3] These are mixed pixels or built-up areas.

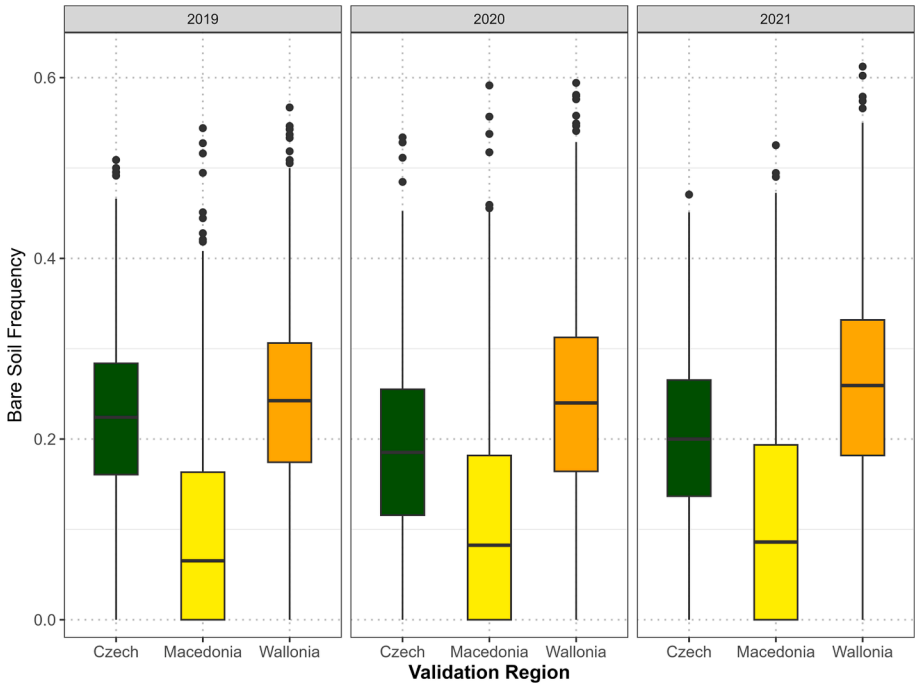


Fig. 5. Box plots of bare soil frequency for the validation points for annual crops in Wallonia, Macedonia and the Czech Republic validation sites.

that the model attains a fair prediction with an RMSE of 18 g C kg^{-1} and a low bias of 0.08 g C kg^{-1}. It is however evident that still the high SOC content values are very difficult to predict and the model underpredicts the high SOC content values.

To shed light into the underlying reasoning process, we calculated the feature importance of the model using the Shapley algorithm. The most important information as identified by the Shapley values appears to be in the visible range (B3) which indicates that the model tends to use the soil colour as one of the most important features (Fig. 7). Significant importance is also placed on B8 in the near infrared (at about 850 nm), while SWIR (B12) is also somewhat significant. This is consistent with the results of Castaldi (2021) who compared SOC prediction models derived from the laboratory spectral data base of LUCAS, and the Sentinel 2 and Landsat8 spectra for the pixels corresponding to the LUCAS sampling points. They found that the visible region was the most important for the laboratory models on the full spectra with some effect of the NIR and SWIR region.

As these two regions are only represented by two rather broad bands in the Sentinel 2 spectra, they are much less important in the variable importance graphs (Fig. 7). For Brazil, the red, NIR and SWIR2 bands were the most important contributors to the SOC prediction models (Safanelli et al., 2021). The shift in importance towards longer wavelengths compared to the European case is probably due to the high iron contents in the tropical soils in Brazil which masks the effect of colouring by organic matter (Chabrillat et al., 2019).

3.2.2. Permanently vegetated areas

The results of the cross-validation show that the model for permanently vegetated areas did not have great performances (MEC = 0.28, RMSE = 65.35 g C kg^{-1}; Table 6), whilst in line with other models across continental areas. In contrast to other continental models that have a resolution of 250 m (Soil Grids, (Poggio et al., 2021) or 500 m (European soil property maps based on LUCAS; (de Brogniez et al., 2015) (Ballabio et al., 2019) the Worldsoils algorithms are calculated at Sentinel 2 (i.e. 20 m) resolution and displayed in the GUI at 50m resolution for the pilot areas and 100 m for the remainder of Europe. This higher resolution enables integrating management practices that can vary within or more often between fields as co-variates in contrast to other DSM algorithms.

The model for permanently vegetated areas used SOC values across a wide range encompassing observations with very little SOC to observations with a maximum of more than 500 g C kg^{-1} (Fig. 8). This wide range of values with few observations for the extremes is challenging to model. The most important covariates in the model for permanently vegetated areas were the bands from the Sentinel 2 composites (mean reflectance composite − MREF), together with climatic and geomorphological features (Fig. 9).

3.3. The uncertainty

The uncertainty of the SOC predictions is expressed as the 90th percentile range (i.e., P95-P5) or in other words the interval that covers 90% of the predictions. A zoom on an area in the North of Wallonia clearly shows that the uncertainty of SOC predictions for the croplands with an average SOC content of 10 g C kg^{-1} is below 2.5 g C kg^{-1} (Fig. 10). The pixels with higher uncertainty often partly cover a road or are on field boundaries. The same uncertainty range of c. 25 % was also observed for croplands in Macedonia and Czech Republic.

Table 6
Cross validation of the SOC prediction algorithms.

	MEC[1]	RMSE (g kg^{-1})	NRMSE	Bias	RE (%)	RPD	RPIQ	PICP (90 %)
Bare soil	0.41	18.07	0.77	0.08	58.27	1.30	0.73	0.97
Permanently vegetated	0.28	67.35	1.58	−1.97	120.78	1.18	0.34	0.94

[1] MEC: model efficiency coefficient; RMSE: root mean squared error; NRMSE: normalized root mean squared error; RE: relative error; RPD ratio of performance to deviation (i.e. RMSE/Std); RPIQ: ratio of performance to interquartile range (i.e. RMSE/(Q3-Q1)); PICP: prediction interval coverage probability.

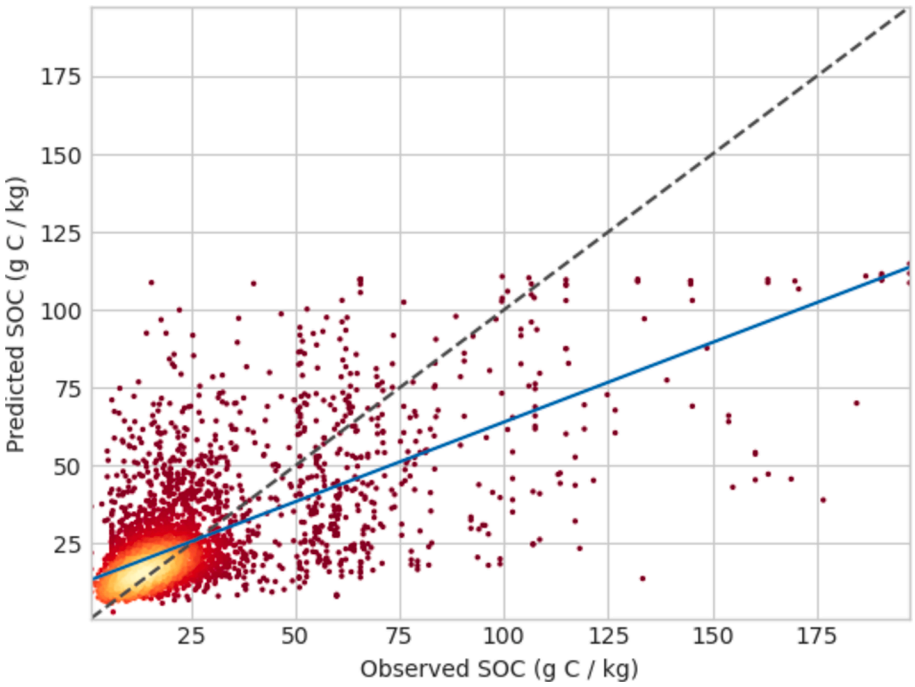


Fig. 6. Scatter plot between observed and predicted SOC (in g C kg^{-1}) for the bare soil model with the 1:1 (dashed) and the regression (solid blue) lines.

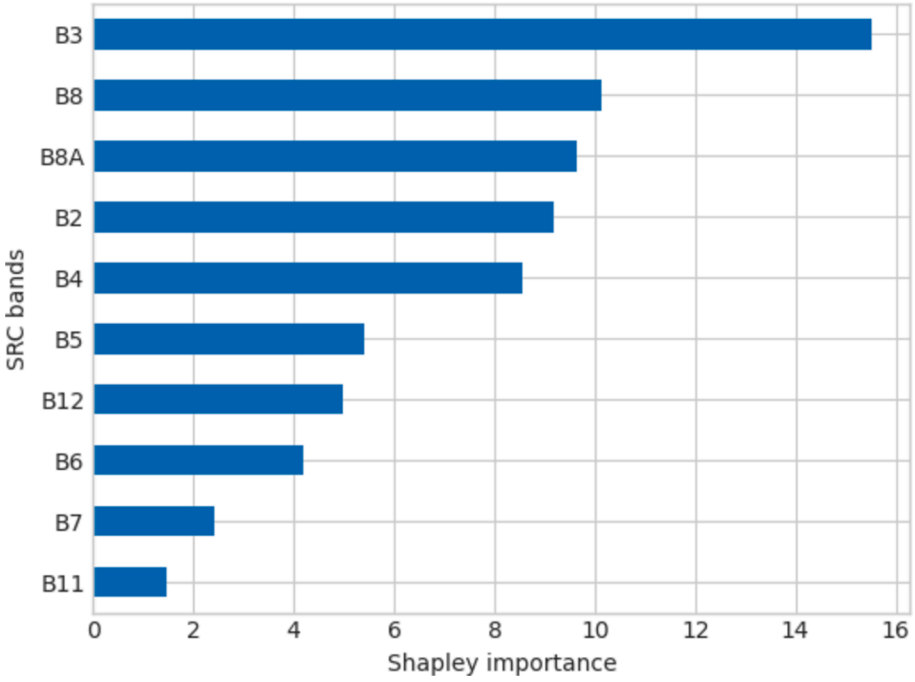


Fig. 7. Mean feature importance of Sentinel 2 soil reflectance composite (SRC) bands calculated over all data.

3.4. Evaluation of the SOC monitoring system

So far, we discussed the performance of each of the steps in the processing pipeline (Fig. 1). As there were different steps from attributing the pixels to one of the two SOC algorithms, mosaicking, aggregation from the Sentinel 2 resolution to the 50 m (pilot regions) and 100 m resolution (rest of Europe), we used recent georeferenced data sets of topsoil SOC contents in the pilot regions as the best approximation to an independent validation.

Obviously, the time period is still relatively short because of the Sentinel 2 twin satellite constellation only became operational in 2018. The SOC content predictions for agricultural soils were compared to independent data sets. The evaluation only concerns agricultural soils, as the sampling protocols for forest soils are different from the ones in agricultural soils and the date at which the former were collected does not correspond to the dates of the SOC prediction maps. Details on the three regions and their content datasets are given in the supplementary material.

Overall, the performance of the monitoring system remains stable for the three periods. This is to be expected as the periods are partly

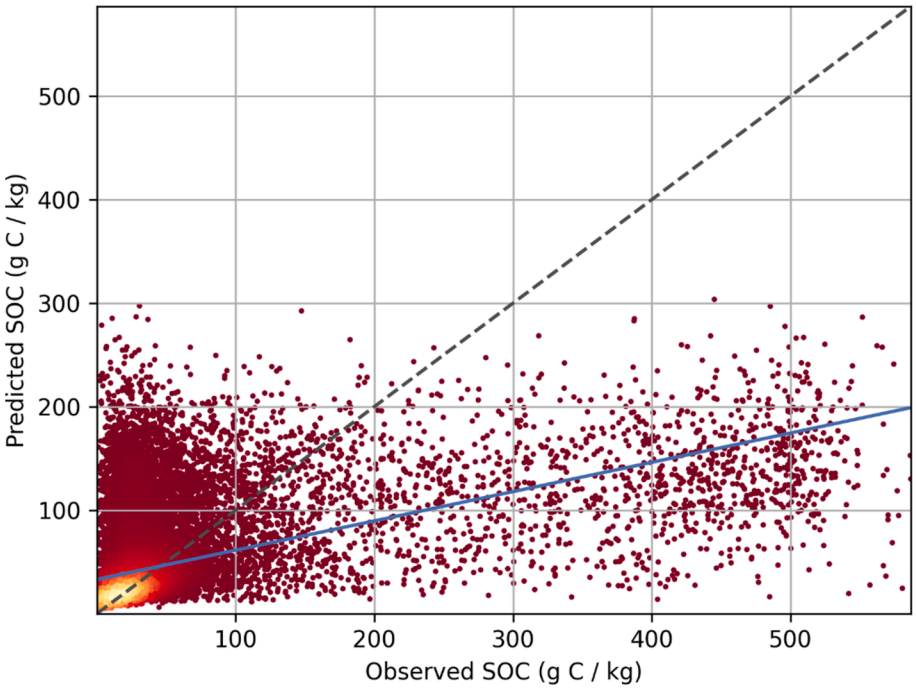


Fig. 8. Scatter plot between observed and predicted SOC (in g C kg^{-1}) for the permanently vegetated model with the 1:1 (dashed) and the regression (solid blue) lines.

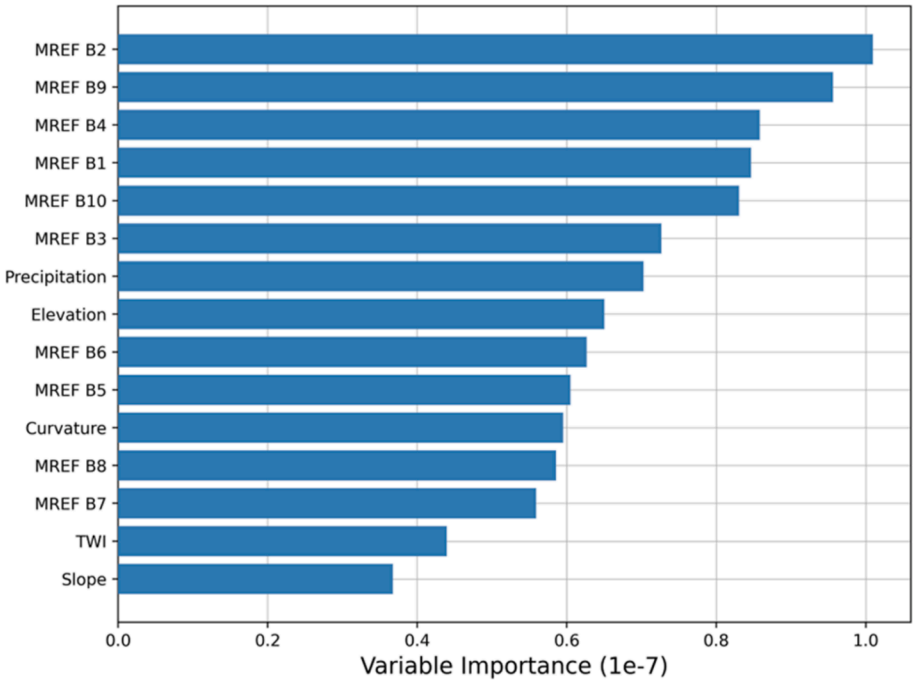


Fig. 9. Random forest variable importance computed as reduction of Gini impurity for the SOC prediction model of permanently vegetated areas. MREF B1-B10 refers to the mean reflectance spectrum of the Sentinel 2 bands; TWI: topographical wetness index. Detailed explanation of the predictors used can be found in Table S3.

overlapping and cover a short period for which SOC is generally considered to be a constant soil property (Fig. 11).

For Wallonia, the points are distributed around the 1:1 line on a graph of predicted against observed values, although there is a tendency to over predict the values above 80 g C kg^{-1} (Fig. 11). As these high SOC values are extremely rare, it was decided to consider both observed and predicted values above 80 g C kg^{-1} as outliers and the performance parameters were assessed removing these outliers. Overall, the performance of the model without these outliers is reasonable with R^2 around 0.5 and RPIQ between 1.46 and 1.69. The RMSE (or accuracy) is quite large at 8.4–9.0 g C kg^{-1} mainly because of poorer prediction of pixels with SOC contents higher than 25 g C kg^{-1}. The bias of the SOC predictions is small at –0.33 to –1 g C kg^{-1}.

For Macedonia, the SOC content in croplands with tree crops is

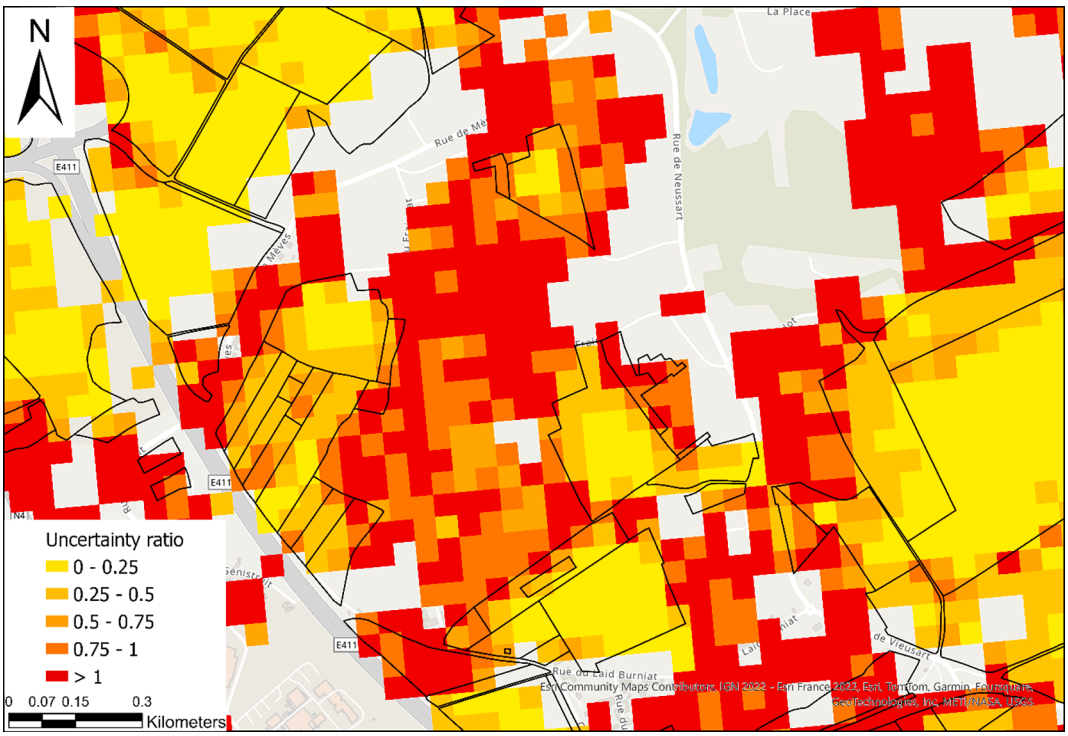


Fig. 10. Zoom of the uncertainty intervals of the predicted SOC content expressed as prediction interval ratio (PIR; eq. (12) in an area in the North of Wallonia. The field borders of the croplands are indicated.

largely over predicted (Fig. 11). This is not surprising as the tree crops are mostly predicted by the permanently vegetated model (Table 4). This model is trained with SOC data for entire Europe and relies heavily on the reflectance in the Vis NIR region, reflecting that vegetation cover increases SOC content as is the case in grasslands and forests in temperate regions (e.g. Poeplau et al. 2011). For Mediterranean tree crops, this is not necessarily the case as the soil is frequently tilled to reduce the water use by weeds, and therefore the return of biomass to these soils is quite low. For the annual crops the model predicts a constant SOC content of ca 15 g C kg^{-1}. However, the uncertainty of these predictions is large as already 40 % of the annual crops are attributed to the permanently vegetated model (Table 4) and for at least half of the validation points the bare soil count is in the critical range of 7–10 (Table 5).

For the Czech Republic, the model metrics validated on the national dataset show relatively low values. This shows, as expected, that the global model is not fully accurate in modelling values at the local level. However, in case of the validation dataset, the nature of the data and the way it is sampled, especially the composite characteristic of the data, should be considered (Fig. 11). Overall, the model performance is poor, with R^2 around 0.1 and RPIQ between 0.8. The RMSE is high at 6.93–7.32 g C kg^{-1}, mainly due to the poorer prediction of pixels with SOC content higher than 25 g C kg^{-1}. The bias in the SOC predictions is small.

4. Discussion

4.1. Bare soil frequency as a proxy for conservation practices

Winter cereals and fodder crops such as alfalfa have an extensive cover during the growing season, while spring seeded crops such as sugar beets, potatoes and maize still show bare soils during the start of the growing season (from March until May). As Castaldi et al. (2024) demonstrated these cereals and particularly fodder crops are a driver for maintaining and increasing SOC content that can therefore have a relationship with the bare soil frequency proxy. Crop residues decrease the bare soil frequency until it reaches critical levels for the SOC prediction as demonstrated by the samples in Macedonia (Table 5). As we used satellite imagery from March to October (Table 2) for our prediction models, there is a risk that the SOC prediction in Mediterranean environments is less accurate than the one in temperate environments. Summer crops, seeded in April and May are more common in temperate environments and therefore the bare soil after seeding is more frequent in these environments (Fig. 5).

Conservation agriculture will no doubt even lead to increased cover of residues as these are no longer ploughed into the soil by inversion tillage. Moreover, there will be a vertical SOC gradient in the topsoil as a result of the reduced depth and reduced intensity of tillage (Priori et al. 2024). Currently, the SWIR bands of the Sentinel 2 are too wide to capture the signal of crop residues. Hively et al., (2021) suggested to add a few narrow width spectral bands in the SWIR region for the detection of the cellulose absorption index (CAI) for the next generation of Landsat multispectral instruments. This would also be beneficial for the next generation of the Sentinel-2 instruments and could greatly support the frequent monitoring of agricultural transition and other applications such as soil erosion monitoring.

4.2. Performance of the algorithms

Although most studies are relatively recent, Vaudour et al. (2022) found a significant correlation between RMSE of the prediction and standard deviation and range of the observed SOC. The RMSE of the algorithm for the Worldsoils system (RMSE= 18.07 g C kg^{-1}, Table 6) is slightly higher than the one predicted by Vaudour et al. (2022) based on the same range: 15.24 g kg^{-1}. Castaldi (2021) used part of the LUCAS 2015 dataset and Sentinel 2 composite for their SOC content algorithm. Not surprisingly, they obtained a similar RMSE at 16.31 g C kg^{-1}. The SOC content prediction algorithm using Landsat composites over 40-year period in Brazil covered a much smaller range of SOC contents (max. 48 g C kg^{-1}; Safanelli et al., 2021) compared to the LUCAS dataset for Europe (max. 197 g C kg^{-1}; Fig. 4). Considering the smaller range, the RMSE in Brazil was also quite low at 3.76 g C kg^{-1}. Meng et al.

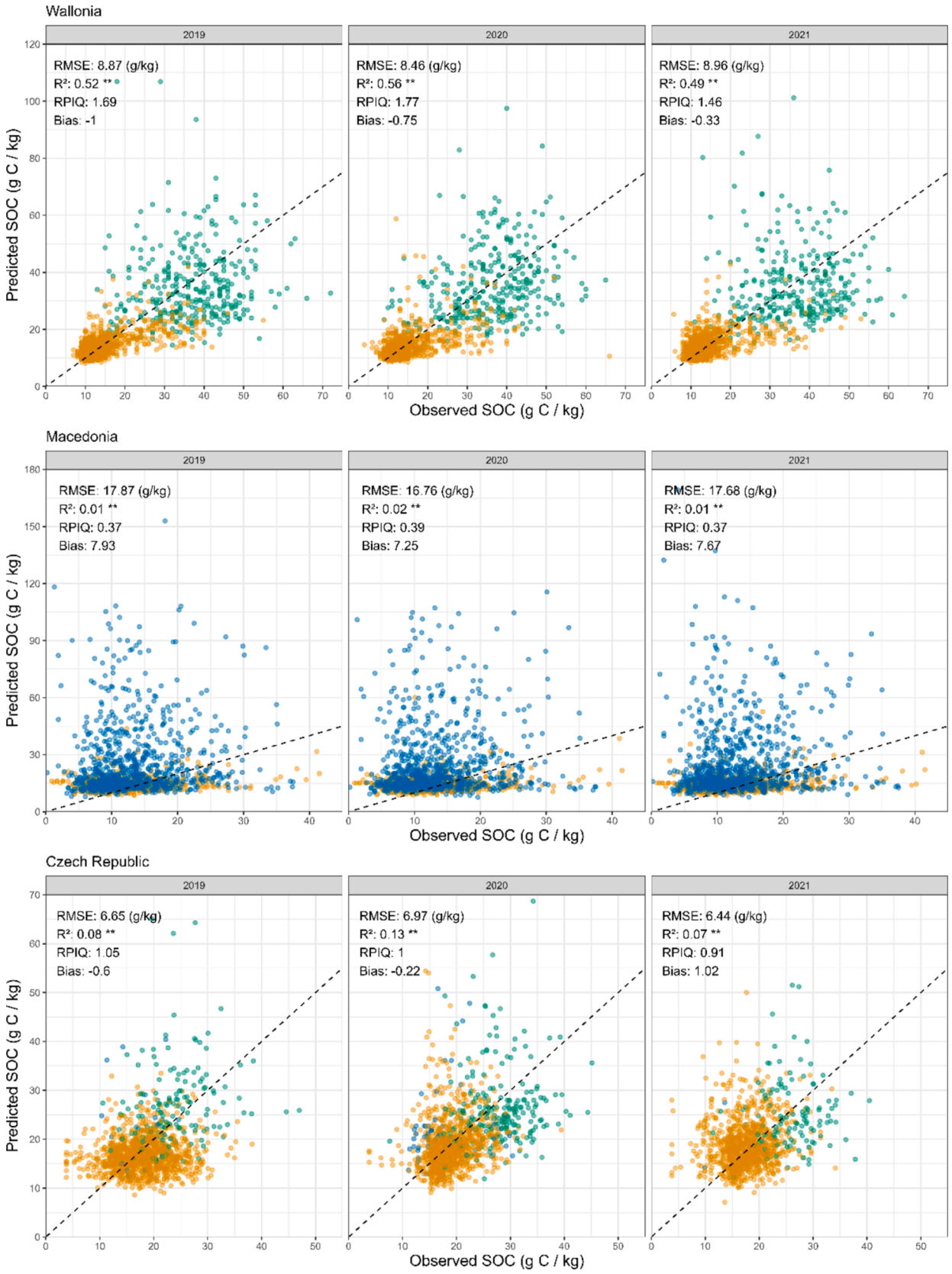


Fig. 11. Predicted versus observed SOC contents for agricultural soils, top panel: Wallonia, middle panel: Macedonia, bottom panel Czech Republic. The 1:1 line is given (broken line). Symbols: annual crops in orange, tree crops in blue and grasslands in green. The significance level of R^2 is indicated (ns: non significant, * $p < 0.05$; **$p < 0.01$).

(2024) used many Landsat TM and OLI images to predict the SOC content in the global Mollisol areas including northeast China, the United States, Ukraine and Russia and Argentina. The range in SOC content at 82 g C $kg^{-1}$is much smaller than for the LUCAS dataset (Fig. 4), resulting in a quite low RMSE at 4.84 g C kg^{-1} for the Mollisol area.

In the bare soil model, due to the high positive skewness of the SOC distribution, we augmented the dataset by selecting neighbouring samples for high SOC content samples. While the oversampling technique effectively increased the number of data points for high SOC content, it's important to acknowledge that even so there still is a positive skewness in the resulting distribution. However, this technique enabled the models to also predict the higher SOC contents, even though these are still somewhat underpredicted (Fig. 6). This performance is better compared to the first models that were trained on the original dataset without augmentation, which fail to produce predictions of high SOC content and only model the lower end of the distribution (results not shown). Future research could explore the impact of different oversampling strategies, such as varying the ratio of augmented to original samples, to better understand how this technique affects the model's ability to generalize to unseen data.

4.3. Uncertainty

Andries et al. (2021) carried out a survey amongst land managers and other stakeholders in order to investigate their preference for an uncertainty indicator. Most of the stakeholders suggested the percentage of observations that falls within the uncertainty interval of the model as an indicator and suggested a threshold of 90 % for the reliability of models to be used in decision making. The uncertainty of the Worldsoils system does not reach this 90 % threshold, as for only 31% of the validation sample points in the Czech Republic, 43 % in Wallonia and 54 % in Macedonia the SOC content falls within the P95-P5 (90% uncertainty) range of the prediction. Although the pixels (50*50 m) in the pilot regions are at an adequate scale to assess the within field variation in SOC content of most of the fields (Fig. 11), an aggregation of SOC contents to the average of an entire field results in a clearer image (Samarinas et al. 2023). After all, management practices such as tillage type or crop rotation have a rather uniform impact within the field. Unfortunately,

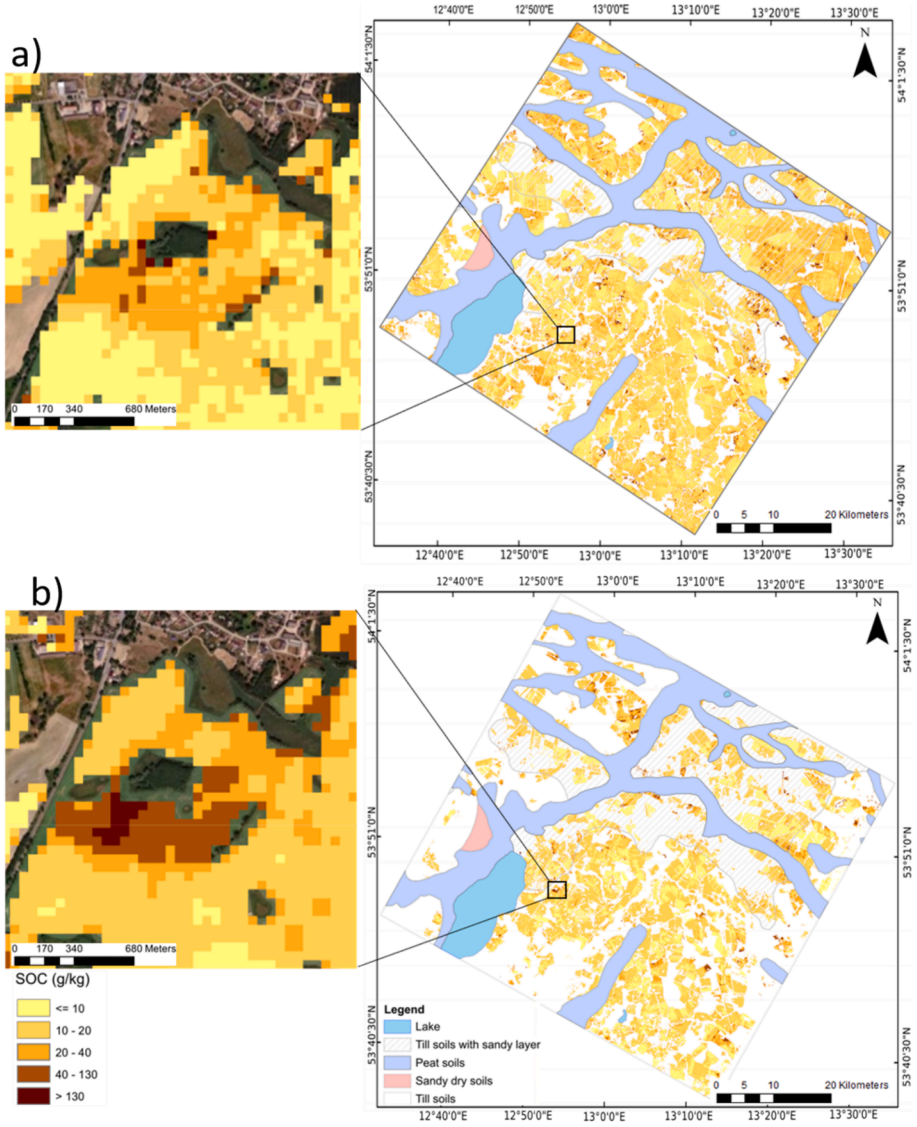


Fig. 12. Satellite-based topsoil SOC content predictions in croplands based on a) the Worldsoils system using the multispectral Sentinel 2 time series 2020–2022, resampled at 30 m and b) a regional prediction model based on four hyperspectral PRISMA acquisitions at 30 m in 2020–2021 (Ward et al., 2024). Left: Zoom around the village of Borrentin on a Google Earth background.

the Land parcel identification system (LPIS) is not generally available throughout Europe and the uncertainty values cannot simply be calculated by aggregating the pixel values to the field average. Moreover, the size of the pixels at European scale (100*100 m) is too large to characterize the SOC content for individual fields.

4.4. Detecting spatial patterns in SOC

We compared the SOC values and their spatial patterns of the Worldsoils monitoring system to the ones obtained from a regional SOC prediction system based on the latest generation of hyperspectral satellites (PRISMA: Fig. 12). The strong variations in SOC in the young morainic landscape of Northern Germany were picked up by the Worldsoils system, although the extreme high SOC contents in former glacial kettle holes were underestimated compared to the ones predicted by the PRISMA imagery (Ward et al., 2024). The SOC content of the Worldsoils product which is based on multispectral Sentinel 2 satellite data matches very well the predictions from the PRISMA hyperspectral satellite, and local spatial patterns are respected. Only in the high SOC content values (~>120 g C kg^{-1}), the Worldsoils product tends to underestimate SOC, as can be seen in a small area around the village of Borrentin, close to a kettle hole (see zooms Fig. 12). At these organic rich depressions SOC content can reach 150–200 g C kg^{-1} (Ward et al., 2020). The hummocky region shows a complex variability at a short distance of eroded summits to colluvial foot slopes resulting in strong variations of SOC content at small scale in the topsoils, which makes it an ideal site for the demonstration and validation of Earth Observation SOC products at different spatial scales. In general, absolute differences in SOC content between the two products are more than 60 g C kg^{-1}. This limitation was already seen in the validation of the Worldsoils bare soil model. However, it should be noted that the results shown in Fig. 12a have been derived using a European-wide model and the results of Fig. 12b are based on a regional model of the Demmin region. Having this in mind, the results from the European model are remarkably good and it is very likely that a regional model using Sentinel-2 data produces higher accuracies and lower uncertainties.

4.5. Limitations and perspectives

The overall accuracy of the monitoring system has been assessed in the three pilot regions, whereby it must be considered that this is not a strict validation as the sampling and analysis protocols were different between the LUCAS calibration data sets and the external validation data sets of the National Reporting Centres. A crucial step in the monitoring system is the correct attribution of the pixels to one of the two models (spectral vs digital soil mapping) and the window of bare soils during the Sentinel 2 over flights. These steps are not generally considered in the evaluation of Earth observation leveraged soil property mapping, see the review of Vaudour et al. (2022). We therefore suggest to not only focus on the global performance indicators of the algorithms such as R^2, RMSE, RPIQ, but also provided pixel-based uncertainty and proxies for the reliability of the composites such as bare soil count. Moreover, the bare soil frequency is a promising proxy to detect regenerative agriculture, even for the growing season.

As a result of the skewed distribution of SOC contents in Europe with a small number of values larger than 150 g C kg^{-1}, the algorithms tend to underestimate at such large SOC contents. The system provides the 90% uncertainty interval for the SOC predictions at pixel level, although the threshold of 90% of the SOC content in the sample points within the uncertainty range of the model is not yet reached Moreover, the system is stable as the SOC contents for the periods 2018–2020, 2019–2021 and 2020–2022 are nearly similar. It can thus be used as a baseline for SOC content and it's monitoring as longer time series of Sentinel 2 imagery become available.

Both for Macedonia and the Czech Republic, the range in SOC contents was rather small which had a serious impact on the evaluation performance, although the R^2 of the predicted/observed graphs was significant for all sites (Fig. 11). This shows that spatial SOC patterns are detectable by the WorldSoils systems and indicates that a further regionalisation of the models in European subregions could significantly improve the results. The WorldSoils monitoring system is a pre-operational modular system that is running on state-of-the-art cloud processing environments. In principle, it can be also used for areas outside Europe, if enough calibration data are available. Moreover, it is now ready for further methodological improvements and the mapping and monitoring of new soil parameters such as clay content can be implemented and tested allowing to monitor one of the soil descriptors i.e. SOC/clay ratio of the recently published soil monitoring law (Council of the European Union, 2024).

In the next decade, we can expect that hyperspectral-based soil products will be more regularly available from space, that will help to refine the regular mapping and monitoring of SOC and other soil properties such as clays, carbonates, iron oxides. Global mapping hyperspectral mission, such as the ESA Copernicus Hyperspectral Imaging Mission for the Environment (CHIME) (Celesti et al., 2022) and the NASA Surface Biology and Geology (SBG) (Turpie et al., 2023) are planned for 2028–2030. Meanwhile, new opportunities for soil mapping from current spaceborne imaging spectroscopy missions are being tested and demonstrated based on the PRISMA (Lopinto et al., 2022) or EnMAP (Chabrillat et al., 2024) satellites. These expectations are associated with increased accuracy compared to multispectral products due to the improved spectral coverage and spectral resolution that shall lead to improved quantification of disturbing factors (e.g. soil moisture, non-photosynthetic vegetation cover) directly pixel-wise from each image, and improved soil properties modelling. Nevertheless, hyperspectral systems will not reach the temporal revisit time of multispectral systems and do involve heavier computation. In this sense, multispectral-based SOC systems such as Worldsoils are already able to deliver maps based on the average SOC content over three years. While these maps developed from multispectral systems have some uncertainties, they offer a unique and relatively accurate source of information that is currently unavailable from other methods.

5. Conclusions

The Worldsoils monitoring system leverages Earth observation and uniform soil data sets to produce a continuous SOC map covering the European territory. The system uses separate models for croplands and permanently vegetated areas such as grasslands and forests. It strongly relies on compositing products from the Sentinel 2 multispectral satellite. The composites provide the mean reflectance for all valid bare soil pixels over a period of three years (from March to October). One of the benefits of the composite is that it does provide bare soil spectra for (nearly) all cropland pixels and it averages out spectral extremes (e.g. very moist and very dry soils). The mean reflectance composite and the bare soil reflectance composite is then the input to the SOC content prediction algorithm. Additionally, the bare soil frequency is computed, which may be used to infer information on soil management that influences the SOC contents in the long term. The derived uncertainty level (P95-P5), expressed as absolute value (for croplands P95-P5 = 2.5 g C kg^{-1}), was lower than the one required by the end users (5 g C kg^{-1}). As external validation, the performance of the Worldsoils monitoring system was evaluated in three pilot regions against national soil data sets For Wallonia (Belgium) the reflectance composites were able to accurately distinguish between croplands and grasslands. The SOC prediction in Wallonia was reasonable (R^2 0.51), although the accuracy of the models as expressed by the RMSE for the validation sets remains rather large at 7.6 g C kg^{-1}. For the Czech Republic and Macedonia (Greece), the validation results were less satisfactory with R^2 of around 0.11. A closer analysis in Macedonia demonstrated that the Worldsoils monitoring system tends to overestimate in Southern European orchards with low SOC due to frequent tillage below the tree canopies. Moreover, the

bare soil frequency in croplands is very low for this region. On the one hand this result could be an indicator for practices such as the use of conservation tillage, cover crops and grass leys, but on the other hand the short window for bare soil degrades the prediction performance of the algorithm.

CRediT authorship contribution statement

Bas van Wesemael: Writing - original draft, Writing – review & editing, Validation, Supervision, Funding acquisition, Conceptualization. **Asmaa Abdelbaki:** Writing – review & editing, Validation, Data curation. **Eyal Ben-Dor:** Writing – review & editing, Funding acquisition, Conceptualization. **Sabine Chabrillat:** Writing – review & editing, Validation, Supervision, Funding acquisition, Conceptualization. **Pablo d'Angelo:** Validation, Software, Methodology, Formal analysis, Data curation. **José A.M. Demattê:** . **Giulio Genova:** Writing – review & editing, Formal analysis, Data curation. **Asa Gholizadeh:** Validation, Funding acquisition, Conceptualization. **Uta Heiden:** Writing – review & editing, Validation, Supervision, Methodology, Funding acquisition, Data curation, Conceptualization. **Paul Karlshoefer:** Validation, Software, Methodology, Formal analysis, Data curation. **Robert Milewski:** Writing – review & editing, Validation, Data curation. **Laura Poggio:** Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Marmar Sabetizade:** Writing – review & editing, Validation, Formal analysis, Data curation. **Adrián Sanz:** Software, Project administration, Methodology, Funding acquisition. **Peter Schwind:** Methodology, Conceptualization. **Nikolaos Tsakiridis:** Writing – review & editing, Validation, Formal analysis, Data curation. **Nikolaos Tziolas:** Writing – review & editing, Validation, Formal analysis, Data curation. **Julia Yagüe:** Writing – review & editing, Project administration, Funding acquisition. **Daniel Žížala:** Validation, Formal analysis.

Declaration of competing interest

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.

Acknowledgements

In the context of WORLDSOILS contract, special guidance has been received from ESA-ESRIN Technical Officers Michael Berger, Zoltan Szantoi and Anke Schickling. Likewise, from Steering Committee - members: Catarina Bamps and Tim Lemmens (The EU Directorate-General for Defence Industry and Space), Rainer Baritz (European Environment Agency), Arwyn Jones (Joint Research Centre), Yi Peng and Yusuf Yigini (Food and Agriculture Organization) and Emmanuelle Vaudour (AgroParisTech). Access to validation regional soil data sets was provided by the three National Reporting Centres (NRC) on soil of the test regions: the Public Service of the Walloon Region (Belgium), the Soil and Water Resource Institute of Elkotimitra (Central Macedonia) and the Central Institute for Supervising and Testing in Agriculture -Úkzúk- (Czech Republic).

Author contributions

Conceptualization: BvW, UH, PS, EBD, SC, AG, LP, NTZ; Data curation MS, UH, PK, PdA, AA, GG, RM, LP, NTS, NTZ; Formal analysis MS, PK, PdA, GG, LP, NTS, NTZ: Funding acquisition BvW, UH, JY, AS, EBD, SC, AG, LP, NTZ; Methodology BvW, UH, PK, PdA, PS, AS, LP; Project administration JY, AS; Software PK, PdA, PS, AS, LP; Supervision BvW, UH, SC, AG, LP; Validation BvW, MS, UH, PK, PdA, AA, RM, SC, AG, DZ, NTS, JAMD; Writing – original draft: BvW; Writing – review and editing BvW, MS, UH, JY, AA, EBD, SC, JAMD, NTS, LP, GG.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.geoderma.2024.117113.

Data availability

The authors do not have permission to share data.

References

Andries, A., Morse, S., Murphy, R.J., Lynch, J., Mota, B., Woolliams, E.R., 2021. Can current earth observation technologies provide useful information on soil organic carbon stocks for environmental land management policy? Sustainability (Switzerland) 13 (21). https://doi.org/10.3390/su132112074.

Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., Bochtis, D., 2019. Remote sensing techniques for soil organic carbon estimation: A review. Remote Sensing 11 (6). https://doi.org/10.3390/rs11060676.

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. https://doi.org/10.1016/j.geoderma.2019.113912.

Batjes, N.H., Calisto, L., de Sousa, L.M., 2024. Providing quality-assessed and standardised soil data to support global mapping and modelling (WoSIS snapshot 2023). Earth Syst. Sci. Data Discuss. https://doi.org/10.5194/essd-2024-14 [preprint] in review.

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

Castaldi, F., 2021. Sentinel-2 and landsat-8 multi-temporal series to estimate topsoil properties on croplands. Remote Sensing 13 (17). https://doi.org/10.3390/rs13173345.

Castaldi, F., Buttafuoco, G., Bertinaria, F., Toscano, P., 2024. A geospatial approach for evaluating impact and potentiality of conservation farming for soil health improvement at regional and farm scale. Soil Tillage Res. 244, 106212. https://doi.org/10.1016/j.still.2024.106212.

Celesti, M., Rast, M., Adams, J., Boccia, V., Gascon, F., Isola, C., Nieke, J., 2022. The copernicus hyperspectral imaging mission for the environment (Chime): Status and planning. International Geoscience and Remote Sensing Symposium, 10.1109/IGARSS46834.2022.9883592.

Chabrillat, S., Ben-Dor, E., Cierniewski, J., Gomez, C., Schmid, T., van Wesemael, B., 2019. Imaging Spectroscopy for Soil Mapping and Monitoring. Surveys Geophys. 40 (3), 361–399. https://doi.org/10.1007/s10712-019-09524-0.

Chabrillat, S., Foerster, S., Segl, K., Beamish, A., Brell, M., Asadzadeh, S., Milewski, R., Ward, K., Brosinsky, A., Koch, K., Scheffler, D., Guillaso, S., Kokhanovsky, A., Roessner, S., Guanter, L., Kaufmann, H., Pinnel, N., Carmona, E., Storch, T., Hank, T., Berger, K., et al., 2024. The EnMAP spaceborne imaging spectroscopy mission: Initial scientific results two years after launch. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2024.114379.

Chang, C.C., Lin, C.J., 2001. Training *v*-support vector classifiers: theory and algorithms. Neural Comput 13 (9), 2119–2147. https://doi.org/10.1162/089976601750399335.

Chen, T., Guestrin, C., 2016. A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). Association for Computing Machinery, New York, NY, USA, pp. 785–794, 10.1145/2939672.2939785.

Cheng, L., Yan, M., Zhang, W., Guan, W., Zhong, L., Xu, J., 2024. Interpretable digital soil organic matter mapping based on geographical Gaussian process-generalized additive model (GGP-GAM). Agriculture (switzerland) 14 (9). https://doi.org/10.3390/agriculture14091578.

Council of the European Union, (2024). Proposal for a directive of the European Parliament and of the Council on soil monitoring and resilience (Soil Monitoring Law). Revised presidency compromise text (2023/0232(COD)). Brussels, May 3, 2024.

de Brogniez, D., Ballabio, C., Stevens, A., Jones, R.J.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 (1). https://doi.org/10.1111/ejss.12193.

Demattê, J.A.M., Fongaro, C.T., Rizzo, R., Safanelli, J.L., 2018. Geospatial Soil Sensing System (GEOS3): A powerful data mining procedure to retrieve soil spectral reflectance from satellite images. Remote Sens. Environ. 212, 161–175. https://doi.org/10.1016/J.RSE.2018.04.047.

Demattê, J.A.M., Safanelli, J.L., Poppiel, R.R., Rizzo, R., Silvero, N.E.Q., Mendes, W.S., Bonfatti, B.R., Dotto, A.C., Salazar, D.F.U., Mello, F.A.O., Sayão, V.M., Lisboa, C.J.S., 2020. Bare earth's surface spectra as a proxy for soil resource monitoring. Sci. Rep. 10 (1). https://doi.org/10.1038/s41598-020-61408-1.

Diek, S., Fornallaz, F., Schaepman, M.E., de Jong, R., 2017. Barest Pixel Composite for agricultural areas using landsat time series. Remote Sens. 9 (12). https://doi.org/10.3390/rs9121245.

Dvorakova, K., Heiden, U., Pepers, K., Staats, G., van Os, G., van Wesemael, B., 2023. Improving soil organic carbon predictions from a Sentinel–2 soil composite by assessing surface conditions and uncertainties. Geoderma 429. https://doi.org/10.1016/j.geoderma.2022.116128.

García, M.J.L., Caselles, V., 1991. Mapping burns and natural reforestation using thematic Mapper data. Geocarto Internat. 6, 31–37. https://doi.org/10.1080/10106049109354290.
Heiden, U., d'Angelo, P., Schwind, P., Karlshöoefer, P., Müller, R., Zepp, S., Wiesmeier, M., Reinartz, P., 2022. Soil reflectance composites-improved thresholding and performance evaluation. Remote Sensing 14 (18). https://doi.org/10.3390/rs14184526.
Hively, W.D., Lamb, B.T., Daughtry, C.S.T., Serbin, G., Dennison, P., Kokaly, R.F., Wu, Z., Masek, J.G., 2021. Evaluation of SWIR crop residue bands for the Landsat next mission. Remote Sensing 13 (18). https://doi.org/10.3390/rs13183718.
Ishwaran, H., 2015. The effect of splitting on random forests. Machine Learn. 99, 75–118. https://doi.org/10.1007/s10994-014-5451-2.
Iso 10694,, 1995. Soil quality -Determination of organic and total carbon after dry combustion (elementary analysis). International Organization for Standardization, Geneva, Switzerland.
Key, C.H., Benson, N.C., 2003. The normalized burn ratio (NBR): A Landsat TM radiometric measure of burn severityU.S. U.S. Department of the Interior, U.S. Geological Survey, Northern Rocky Mountain Science Center.
Kingma, P., Ba, J. (2014). Adam: A method for stochastic optimization. arXiv: 1412.6980v9 [cs.LG], https://doi.org/10.48550/arXiv.1412.6980.
Loiseau, T., Chen, S., Mulder, V.L., Román Dobarco, M., Richer-de-Forges, A.C., Lehmann, S., Bourennane, H., Saby, N.P.A., Martin, M.P., Vaudour, E., Lagacherie, P., Arrouays, D., 2019. Satellite data integration for soil clay content modelling at a national scale. Internat. J. Appl. Earth Observ. Geoinform. 82. https://doi.org/10.1016/j.jag.2019.101905.
Lopinto, E., Fasano, L., Longo, F., Sacco, P., 2022. Current Status of Prisma Mission. International Geoscience and Remote Sensing Symposium (IGARSS), 10.1109/IGARSS46834.2022.9884414.
Lundberg, S.M. & Lee, S,I., 2017. A unified approach to interpreting model predictions. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA. https://papers.nips.cc/paper_files/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.
Meinshausen, N., 2006. Quantile regression forests. J. Machine Learn. Res. 7, 983–999.
Meng, X., Bao, Y., Luo, C., Zhang, X., Liu, H., 2024. SOC content of global Mollisols at a 30 m spatial resolution from 1984 to 2021 generated by the novel ML-CNN prediction model. Remote Sens. Environ. 300. https://doi.org/10.1016/j.rse.2023.113911.
Minasny, B., McBratney, A., 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264, 301–311. https://doi.org/10.1016/j.geoderma.2015.07.017.
Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., Fernández-Ugalde, O., 2018. LUCAS Soil, the largest expandable soil dataset for Europe: a review. Eur. J. Soil Sci. 69 (1), 140–153. https://doi.org/10.1111/ejss.12499.
Poeplau, C., Don, A., Vesterdal, L., Leifeld, J., Van Wesemael, B., Schumacher, J., Gensior, A., 2011. Temporal dynamics of soil organic carbon after land-use change in the temperate zone - carbon response functions as a model approach. Global Change Biol. 17 (7), 2415–2427. https://doi.org/10.1111/j.1365-2486.2011.02408.x.
Poggio, L., Gimona, A., Brewer, M.J., 2013. Regional scale mapping of soil properties and their uncertainty with a large number of satellite-derived covariates. Geoderma 209–210, 1–14. https://doi.org/10.1016/j.geoderma.2013.05.029.
Poggio, L., De Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. SOIL 7 (1), 217–240. https://doi.org/10.5194/soil-7-217-2021.
Priori, S., Zanini, M., Falcioni, V., Casa, R., 2024. Topsoil vertical gradient in different tillage systems: An analytical review. Soil Tillage Res. 236. https://doi.org/10.1016/j.still.2023.105947.
Rogge, D., Bauer, A., Zeidler, J., Mueller, A., Esch, T., Heiden, U., 2018. Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). Remote Sens. Environ. 205, 1–17. https://doi.org/10.1016/j.rse.2017.11.004.
Rouse, J.W., Haas, R.H., Scheel, J.A., Deering, D.W., 1974. Monitoring Vegetation Systems in the Great Plains with ERTS. In: Proceedings, 3rd Earth Resource Technology Satellite (ERTS) Symposium, pp. 48–62.
Safanelli, J.L., Chabrillat, S., Ben-Dor, E., Demattê, J.A.M., 2020. Multispectral models from bare soil composites for mapping topsoil properties over Europe. Remote Sensing 12 (9). https://doi.org/10.3390/RS12091369.
Safanelli, J.L., Demattê, J.A.M., Chabrillat, S., Poppiel, R.R., Rizzo, R., Dotto, A.C., Silvero, N.E.Q., Mendes, W.D.S., Bonfatti, B.R., Ruiz, L.F.C., ten Caten, A., Dalmolin, R.S.D., 2021. Leveraging the application of Earth observation data for mapping cropland soils in Brazil. Geoderma 396. https://doi.org/10.1016/j.geoderma.2021.115042.
Samarinas, N., Tsakiridis, N.L., Kokkas, S., Kalopesa, E., Zalidis, G.C., 2023. Soil data cube and artificial intelligence techniques for generating national-scale topsoil thematic maps: a case study in lithuanian croplands. Remote Sensing 15 (22). https://doi.org/10.3390/rs15225304.
Sanderman, J., Hengl, T., Fiske, G.J., 2017. Soil carbon debt of 12,000 years of human land use. Proc. Natl. Acad. Sci. USA 114 (36), 9575–9580. https://doi.org/10.1073/pnas.1706103114.
Schmidinger, J., Heuvelink, G.B.M., 2023. Validation of uncertainty predictions in digital soil mapping. Geoderma 437, 116585. https://doi.org/10.1016/J.GEODERMA.2023.116585.
Schwalbe, G., Finzel, B., 2024. A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts. Data Min Knowl Disc 38, 3043–3101. https://doi.org/10.1007/s10618-022-00867-8.
Seliya, N., Abdollah Zadeh, A., Khoshgoftaar, T.M., 2021. A literature review on one-class classification and its potential applications in big data. J. Big Data 8, 122. https://doi.org/10.1186/s40537-021-00514-x.
Sentiwiki, 2024. S2 Processing. https://sentiwiki.copernicus.eu/web/s2-processing#S2Processing-Step2.1SnowFilter1:NormalisedDifferenceSnowIndex (NDSI) (accessed 10 October 2024).
Sousa, G.P.B., Belinaso, H., Rosas, J.T.F., de Mello, D.C. de, N.A. Rosin a,N.A., Amorim, M.T.A. , dos Angos Bartsch, B., B.dos,A. M.C. Cardoso, M.C., a, Mallah, S., Francelino, M.R., Falcioni, R., Alves, M.R., Demattê. José. A.M. (2024) Assessing Soil Degradation in Brazilian Agriculture by a Remote Sensing Approach to Monitor Bare Soil Frequency: Impact on Soil Carbon. Soil Advances, Volume 2, 100011. Available online 15 July 2024, 100011, In Press, Journal Pre-proof, https://doi.org/10.1016/j.soilad.2024.100011.
Tóth, G., Jones, A., Montanarella, L., 2013. The LUCAS topsoil database and derived information on the regional variability of cropland topsoil properties in the European Union. Environ. Monitor. Assess. 185 (9), 7409–7425.
Turpie, K.R., Casey, K.A., Crawford, C.J., Guild, L.S., Kieffer, H., Lin, G., Kokaly, R., Shrestha, A.K., Anderson, C., Ramaseri Chandra, S.N., Green, R., Hook, S., Lukashin, C., Thome, K., 2023. Calibration and Validation for the Surface Biology and Geology (SBG) mission concept: recommendations for a multi-sensor system for imaging spectroscopy and thermal imagery. J. Geophys. Res.: Biogeosci. 128. https://doi.org/10.1029/2023JG007452.
Tziolas, N., Tsakiridis, N., Ben-Dor, E., Theocharis, J., Zalidis, G., 2019. A memory-based learning approach utilizing combined spectral sources and geographical proximity for improved VIS-NIR-SWIR soil properties estimation. Geoderma 340, 11–24. https://doi.org/10.1016/J.GEODERMA.2018.12.044.
Vaudour, E., Gomez, C., Loiseau, T., Baghdadi, N., Loubet, B., Arrouays, D., Ali, L., Lagacherie, P., 2019. The impact of acquisition date on the prediction performance of topsoil organic carbon from Sentinel-2 for croplands. Remote Sensing 11 (18). https://doi.org/10.3390/rs11182143.
Vaudour, E., Gholizadeh, A., Castaldi, F., Saberioon, M., Borůvka, L., Urbina-Salazar, D., Fouad, Y., Arrouays, D., Richer-De-forges, A.C., Biney, J., Wetterlind, J., Van Wesemael, B., 2022. Satellite imagery to map topsoil organic carbon content over cultivated areas: an overview. Remote Sensing 14 (12). https://doi.org/10.3390/rs14122917.
Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. Remote Sens. Environ. 185, 46–56. https://doi.org/10.1016/j.rse.2016.04.008.
Wadoux, A.-M.-J.-C., Román Dobarco, M., Malone, B., Minasny, B., McBratney, A.B., Searle, R., 2023. Baseline high-resolution maps of organic carbon content in Australian soils. Scientific Data 10 (1). https://doi.org/10.1038/s41597-023-02056-8.
Walkley, A.J., Black, I.A., 1934. Estimation of soil organic carbon by the chromic acid titration method. Soil Sci. 37, 29–38.
Wang, J., Chen, Y., Wu, Z., Wei, Y., Zhang, Z., Wang, X., Huang, J., Shi, Z., 2024. On the effectiveness of multi-scale landscape metrics in soil organic carbon mapping. Geoderma 449, 117026. https://doi.org/10.1016/J.GEODERMA.2024.117026.
Ward, K.J., Chabrillat, S., Brell, M., Castaldi, F., Spengler, D., Foerster, S., 2020. Mapping soil organic carbon for airborne and simulated EnMAP imagery using the LUCAS soil database and a local PLSR. Remote Sens. 12, 3451. https://doi.org/10.3390/rs12203451.
Ward, K.J., Chabrillat, S., Foerster, S., 2024. Estimating soil organic carbon using multitemporal PRISMA imaging spectroscopy data. Geoderma 450, 117025. https://doi.org/10.1016/j.geoderma.2024.117025.
Wright, M.N., Ziegler, A., 2017. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J. Stat. Software 77, 1–17. https://doi.org/10.18637/jss.v077.i01.