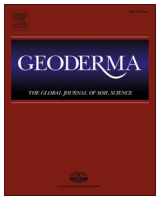
Geoderma 444 (2024) 116850

Contents lists available at ScienceDirect

Geoderma

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



Using local ensemble models and Landsat bare soil composites for large-scale soil organic carbon maps in cropland

Tom Broeg [a,b,*], Axel Don [c], Alexander Gocht [a], Thomas Scholten [b], Ruhollah Taghizadeh-Mehrjardi [b], Stefan Erasmi [a]

[a] Thünen Earth Observation (ThEO), Thünen Institute of Farm Economics, Bundesallee 63, 38116 Braunschweig, Germany
[b] Department of Geosciences, Soil Science and Geomorphology, University of Tübingen, 72070 Tübingen, Germany
[c] Thünen Institute of Climate-Smart Agriculture, Bundesallee, 65, 38116 Braunschweig, Germany

ARTICLE INFO

Handling Editor: Budiman Minasny

Keywords:
Soil Reflectance Composite
Digital Soil Mapping
Soil Organic Carbon
Cropland
Ensemble Modeling
Pedometrics

ABSTRACT

National soil organic carbon (SOC) maps are essential to improve greenhouse gas accounting and support climate-smart agriculture. Large-scale SOC models based on wall-to-wall soil information from remote sensing remain a challenge due to the high diversity of natural soil conditions and the difficulty of accounting for the spatial location of the soil samples. In this study, we tested if the implementation of local ensemble models (LEM) can be used to improve the SOC predictions from Landsat-based soil reflectance composites (SRC) for Germany. For this, we divided the research area into 30 times 30 km tiles and calculated local generalized linear models (GLM) based on random, nearby observations. Based on the GLMs, local SOC maps were predicted and aggregated using a moving window approach. The local variable importance was analyzed to identify spatial dependencies in the correlation between the SRC and SOC. For the final SOC map, a Random Forest (RF) model was trained using the aggregated local SOC predictions, the SRC, and a full set of training samples from the agricultural soil inventory. The results show that the LEM was able to improve the accuracy ($R^2 = 0.68$; RMSE = 5.6 g kg^{-1}), compared to the maps based on a single, global model ($R^2 = 0.52$; RMSE = 6.8 g kg^{-1}). The local variable importance of the spectral bands showed clear spatial patterns throughout the research area. Differences can be explained by the local soil conditions, influencing the correlation between SOC and the spectral properties. Compared to the widely adopted integration of distance covariates such as geographical coordinates, the LEM was able the reduce the spatial autocorrelation to a greater extent and to improve the prediction accuracy, especially for underrepresented SOC values. The LEM presents a new method to integrate spatial information and increase the interpretability of DSM models.

1. Introduction

Accurate and spatially explicit information on soil organic carbon (SOC) in cropland soils is necessary to improve greenhouse gas accounting (Roe et al., 2019) and to design and verify climate-smart agricultural measures at high resolution (Paustian et al., 2016). Digital soil mapping (DSM) can help in assessing cropland soil conditions and provide comprehensive information on SOC stocks (Amelung et al., 2020; Minasny et al., 2013). Recently, soil reflectance composites (SRC) based on remote sensing data have been used for unbiased and large-scale predictions of soil properties (Diek et al., 2017; Heiden et al., 2022; Roberts et al., 2019; Rogge et al., 2018). This is possible since most cropland soils are regularly exposed and methods of soil spectroscopy can be applied to multispectral bare soil observations (Ben-Dor et al., 2009; Demattê et al., 2018). Multiple studies have proven the effectiveness of SRC in generating cropland SOC maps (Dvorakova et al., 2021; Urbina-Salazar et al., 2023; Vaudour et al., 2021; Zepp et al., 2021; Žížala et al., 2022). In a review by Vaudour et al. (2022), however, it was concluded that only limited research has been conducted at large-scale > 100,000 km^2, e.g. for national models. Safanelli et al. (2020) predicted multiple soil properties across Europe using Landsat data and over 7000 soil samples but reported mixed results. The authors conclude that the high spatial, as well as temporal variability of SOC represents a significant constraint for large-scale SOC predictions. This is supported by results from soil spectroscopy, showing that the relationships between SOC and visible (VIS), near-infrared (NIR), and shortwave

* Corresponding author at: Thünen Earth Observation (ThEO), Thünen Institute of Farm Economics, Bundesallee 63, 38116 Braunschweig, Germany.
E-mail address: tom.broeg@thuenen.de (T. Broeg).

https://doi.org/10.1016/j.geoderma.2024.116850
Received 30 September 2023; Received in revised form 12 February 2024; Accepted 4 March 2024
Available online 19 March 2024

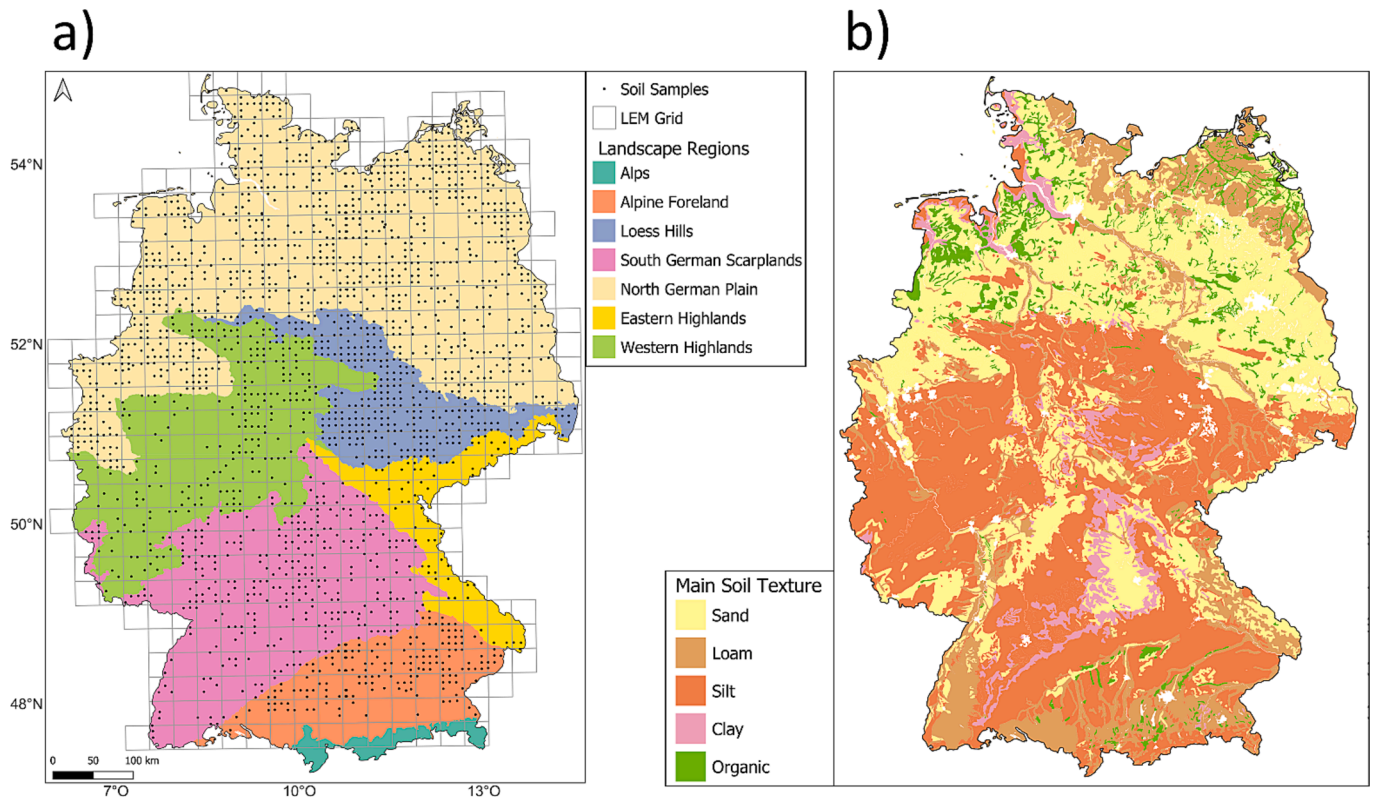


Fig. 1. Overview of the research area: a) Major landscape regions of Germany. The tiling scheme of the LEM and the soil samples (BZE-LW) used to train the SOC models are shown in black; b) Main soil texture based on BGR (2007). White areas are urban areas or mining areas.

infrared (SWIR) are distinct but depend on local factors like the soil texture, moisture, or mineralogy (Chabrillat et al., 2019; Demattê et al., 2018; Stenberg et al., 2010).

The spatial dependency between soil reflectance and SOC can be problematic for large-scale models, when common machine learning (ML) algorithms, like Random Forests (RF), are applied which do not account for the spatial positions of the soil samples. Methods have been developed to integrate spatial information into the DSM framework and reduce the spatial autocorrelation of the predictions. The most common ones are based on different distance measures like Euclidean distance fields (Behrens et al., 2018), oblique geographic coordinates (Møller et al., 2020), or buffer distances (Hengl et al., 2018), which are converted into distance covariates and used for the model training. Because of their simplicity and easy integration, these methods have been commonly used in combination with remote sensing data (Meyer et al., 2019; Urbina-Salazar et al., 2023; Žížala et al., 2022). However, concern has been raised over the missing pedological relevance of distance covariates, potentially masking out the importance of other model variables, such as remote sensing data (Wadoux et al., 2020). This was illustrated by Meyer et al. (2019), showing that the integration of distance covariates can lead to prediction artifacts, even though the model accuracy is improving.

Spatial ML methods like Random Forest Spatial Interpolation (Sekulić et al., 2020), Geographical Random Forests (Georganos et al., 2021), or Spatial Ensemble Learning (Jiang et al., 2017) do not depend on distance covariates but have not yet been adopted by DSM. Spatial Ensemble Learning is based on the idea that a heterogeneous research area can be split into multiple, more homogenous sub-areas to reduce the model complexity. For each geographically distinct area, a local sub-model can be optimized to improve the overall prediction accuracy. In heterogeneous research areas, the sub-models can therefore account for local patterns that are potentially masked out with a single, global model. Similarly, Georganos et al. (2021) trained local models for each observation, only including a defined number of nearby observations as local training data. The results show that the local variable importance varies across the research area, highlighting the importance of including spatial information in the modeling process.

For soil spectroscopy, it has been shown that SOC models calibrated with local samples generally outperform regional models in geologically heterogeneous areas (Stenberg, 2010; Stenberg et al., 2010; Udelhoven et al., 2003). As soil properties are spatially autocorrelated, methods like variograms have been commonly used as analytic and predictive tools (Webster and Oliver, 1992). It has been shown that the integration of local variograms, in contrast to global ones, can improve the model performance (Walter et al., 2001). A similar concept has been adopted for soil spectroscopy, illustrating that Geographically Weighted Regressions (GWR) can be used to account for spatial dependency between spectral signal and the soil properties (Song et al., 2021). These results illustrate that it is not only important to integrate spatial information to improve prediction accuracy but also to understand spatial patterns and increase the model interpretability (Arrouays et al., 2020; Chen et al., 2022; Padarian et al., 2020; Wadoux and Molnar, 2022).

Considering these findings, we adapted and applied a framework, called local ensemble model (LEM), to predict cropland SOC in Germany. A total of 1966 samples from the national agricultural soil inventory of Germany were used as ground truth for model calibration (Poeplau et al., 2020a). In the first step, an SRC was created based on Landsat data and a dynamic thresholding approach to include the full range of soil conditions. We divided the research area into 469 sub-models, using a 30 times 30 km regular grid, to account for the spatial position of the soil samples without additional distance covariates. For each of the sub-models, a generalized linear model (GLM) was calibrated with a random selection of 60 nearby samples to capture the local soil conditions. Spatial variable importance was analyzed to improve the model interpretability and identify spatial dependency in the correlation between the SRC and SOC. To generate the national SOC map, the local

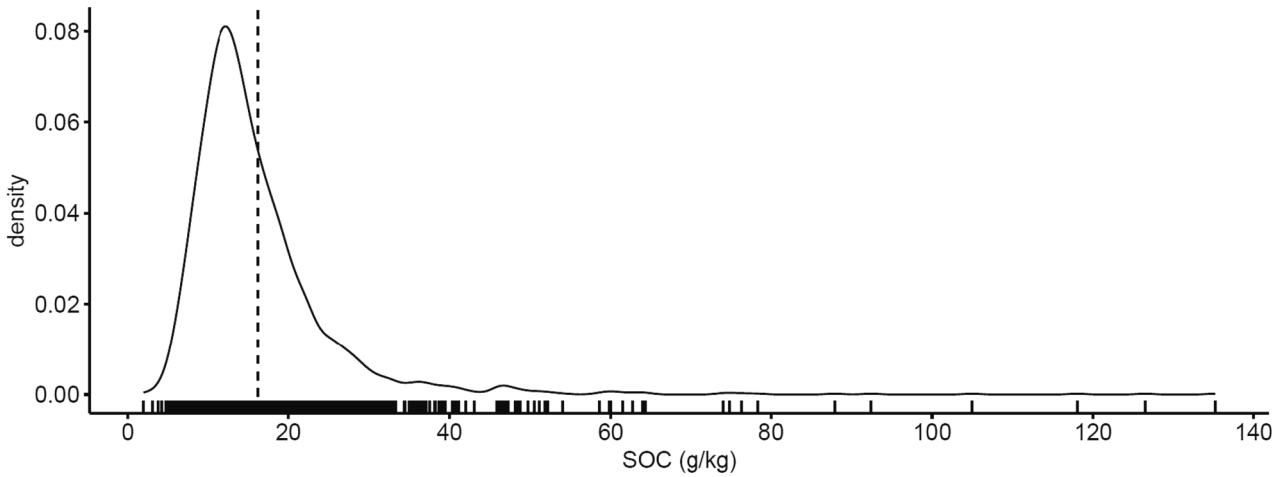


Fig. 2. a) Density plot of the SOC content (g kg^{-1}) of 1966 soil samples across German croplands used for model training and validation (dotted line = mean value). Individual observations are shown above the x-axis; b) Mean SOC contents (g kg^{-1}) of the soil samples within the LEM tiles (white = no samples available).

predictions of the GLMs were aggregated with a moving window approach and used as a covariate in a final RF model.

In total, three hypotheses were tested:

1) The integration of the LEM is improving the accuracy of large-scale SOC predictions in comparison to a global RF model.
2) The LEM can reduce the spatial autocorrelation of the prediction residuals to a higher degree than the implementation of distance covariates.
3) The correlation between SOC and the spectral bands of the SRC is dependent on the local soil conditions and varies across the research area.

2. Material and methods

2.1. Research area

Germany has a total area of 357,592 km^2 of which roughly 36 % is dedicated to cropland (Destatis, 2022). The natural conditions are heterogeneous and can be divided into seven major landscape regions (Thünen-Institut, 2022) (Fig. 1a). The altitude generally follows a North-South gradient, beginning with the North German Plain, followed by the Loess Hills, Eastern, and Western Highlands, the South German Scarplands, and the Alps in the south. The soil conditions are diverse and mainly influenced by geomorphological and geological processes (BGR, 2020, 2007) (Fig. 1b). The North German Plain was formed by glacial and periglacial processes during the Pleistocene. The surface is mostly flat and covered by glacial and fluvial sediments in different weathering stages. It is characterized by mostly sandy and acidic conditions, inducing the formation of bleached soils like Podzols, and Albeluvisols, as well as other sandy soil types like Arenosols. Throughout the North German Plain, organic soils are present in lowland areas and widely drained for agricultural purposes (Fig. 1b). In the northeast, close to the national border, the river Oder induced the formation of a large floodplain (Oderbruch). The loess band in Central East Germany is characterized by a large deposit of loess as aeolian sediment from the Pleistocene. Its high silt content is increasing soil quality and carbon accumulation in some regions, inducing the formation of Chernozems, Kastanozems, and Luvisols. The Eastern and Western Uplands consist of multiple low mountain ranges below 1,500 m a.s.l. In the western part, sedimentary materials from the Devonian and Triassic are present, inducing the formation of Cambisols and other soil types. The Eastern Highlands are mostly built from metamorphosed and plutonic rocks and are characterized by sandy and loamy soils (Fig. 1b). Further west, the South German Scarplands are built by strata of the Triassic and Jurassic.

Table 1
Descriptive statistics of SOC contents (g kg^{-1}) for each landscape region (Fig. 1). SD = standard deviation; IQR = interquartile range.

Landscape Regions	Samples	Min	Median	Mean	Max	SD	IQR
Alpine Foreland	165	7.2	15.5	20.6	126.4	17.7	7.1
Loess Hills	272	6.1	14.5	15.7	48.6	5.4	5.7
South German Scarplands	312	4.2	14.2	15.5	78.4	6.6	6.8
North German Plain	926	2.0	12.9	15.9	135.2	10.5	9.5
Western Highlands	229	5.6	13.5	14.8	47.0	5.0	5.5
Eastern Highlands	62	8.4	18.8	20.3	37.5	6.6	9.1
Total	1966	2.0	14.0	16.2	135.2	9.8	7.7

Here, the soil conditions are very diverse (Fig. 1b). In general, Leptosols are common in regions with limestone while Cambisols and Podzols are connected to sandstone. In the west, high groundwater levels of the Rhine Valley led to the accumulation of carbon and the formation of Phaeozems in some regions. The Alpine Foreland is dominated by the glaciofluvial deposits of the Tertiary Molasse Basin. In this region, loamy Cambisols are the most common (Fig. 1b). Larger areas of organic soil are present in the lowlands of the Pre-Alpes which are fed by rivers originating in the mountains. Approaching the Alps in the south, the density of cropland areas decreases with increasing altitude.

2.2. Soil samples

The soil samples used to train and validate the SOC models were collected by the first German Agricultural Soil Inventory (BZE-LW), conducted between 2011 and 2018 (Jacobs et al., 2018). Details about the sampling design and the laboratory analyses are described in Poeplau et al., (2020b). A total of 3104 soil profiles have been sampled across different agricultural land use classes of which 2254 were sampled on cropland. The selection of sampling sites was based on an 8 times 8 km regular grid across Germany. For each location, a 1 m profile was dug, and a distributed composite sample (1 kg) was taken in five fixed depth increments (0–10 cm, 10–30 cm, 30–50 cm, 50–70 cm, and 70–100 cm) for chemical and textural analyses. The samples were oven-dried (40 °C), sieved (2 mm), and analyzed for SOC (g kg^{-1}) using dry combustion. Measurements for the depths between 0 and 10 cm and 10–30 cm were combined using a weighted average to obtain mean values for the cropland topsoil (0–30 cm). The SOC models were limited

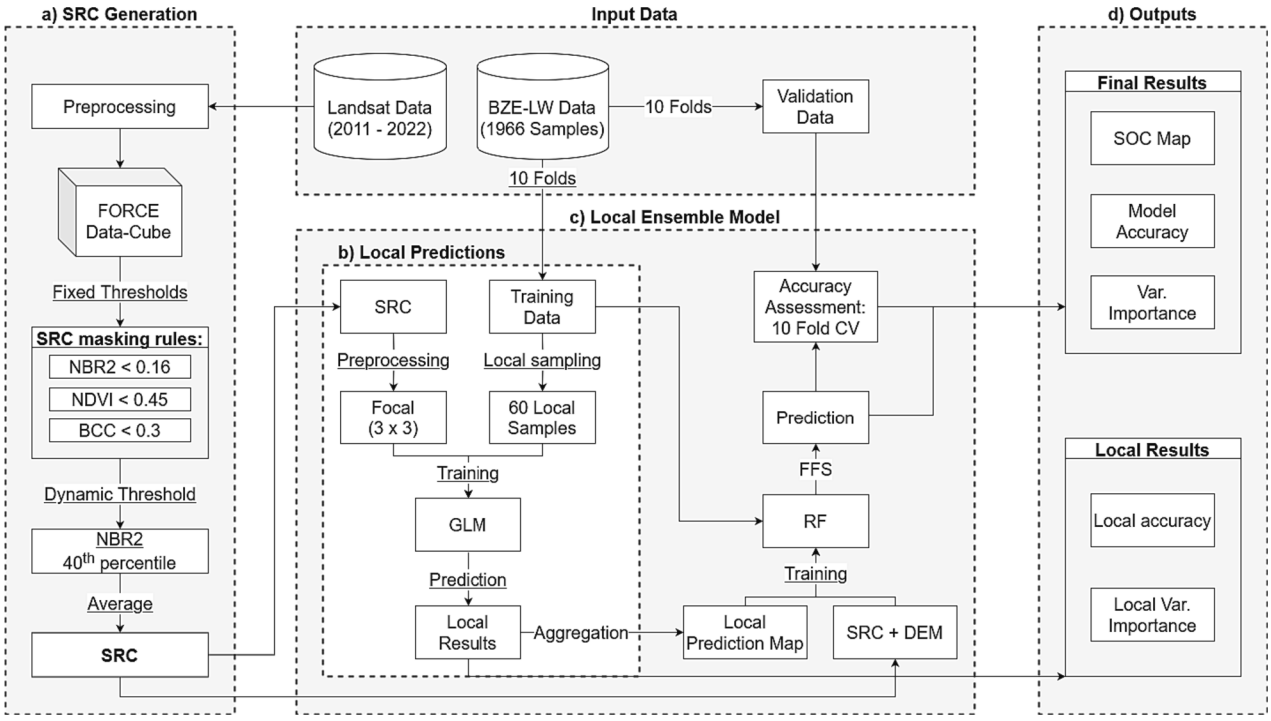


Fig. 3. Flow chart of the Local Ensemble Model: a) Generation of the SRC based on Landsat data (section 2.4); b) Training and prediction of the local GLMs, based on the SRC and local training samples; c) Aggregation of the local maps and RF model for the final SOC prediction, based on all training samples; d) Outputs of the LEM: Final SOC map (RF) and the local model results (GLMs).

to the cropland samples, as soils with permanent vegetation are not represented in the SRC (section 2.4). Based on the extent of the SRC, a total of 1966 soil samples were used for the final SOC models (Fig. 2 + Table 1). The remaining cropland sampling points were masked out from the model area, as no information on the soil reflectance was available (e. g. fallow land or cropland used for grass production). A large proportion of cropland soil showed relatively low SOC concentrations (mean = 16 g kg^{-1}), and most measurements were below 40 g kg^{-1} (Fig. 2). Extreme values on organic soils (>87 g kg^{-1}) were unrepresented but included to cover the full range of cropland SOC values.

2.3. Remote sensing data

The optical satellite data were preprocessed using the Framework for Operational Radiometric Correction for Environmental Monitoring (FORCE) (Frantz, 2019). A data cube was generated based on Landsat images from 2011 to 2022 using the FORCE level 2 processing system. The period was selected to include Landsat 5, 7, and 8 data since the beginning of the BZE-LW sampling campaign in 2011, and more recent Landsat 9 images, starting from 2021. All available Landsat images with a cloud coverage below 70 % were considered. This includes observations from all seasons to improve the representation of different soil management types (e.g. summer and winter crops) and to maximize the number of bare soil observations for each pixel. The final composite was created with a spatial resolution of 30 m, using the six spectral bands that are shared throughout the different Landsat generations (B, G, R, NIR, SWIR1, SWIR2). The FORCE level 2 preprocessing includes multiple steps to create and organize analysis-ready bottom-of-atmosphere reflectance data (Frantz et al., 2016). Each Landsat scene is automatically downloaded, corrected, and tiled into a uniform 30 km by 30 km grid. For this purpose, the scenes are split into multiple image chips that are lined up with the underlying FORCE grid and can be stored in a data cube format for further processing. Cloud masking was performed using an updated version of the Fmask code (Frantz et al., 2018; Zhu and Woodcock, 2012). Radiometric corrections include aerosol optical depth (Royer et al., 1988), topographic correction (Kobayashi and Sanga-Ngoie, 2008), as well as nadir BRDF adjustments (Roy et al., 2016). The radiometrically corrected images were projected to the ETRS89-LAEA (EPSG:3035). In addition to the optical imagery, elevation data from the Shuttle Radar Topography Mission (SRTM) was included. The SRTM digital elevation model (DEM) is available at a resolution of 30 m and was projected to match the FORCE data cube (Farr and Kobrick, 2000).

2.4. Soil reflectance composite

The general workflow for the generation of the SRC is presented in Fig. 3a. Spectral indices were used to automatically detect bare soil observations, based on the Landsat data. We used a combination of fixed and dynamic thresholds to maximize the soil coverage within Germany, while only including optimal observations. The selected bare soil observations were averaged to create the final SRC.

2.4.1. Bare soil indices

A collection of spectral indices was used to detect bare soil observations in the Landsat data. First, the Normalized Difference Vegetation Index (NDVI) was calculated based on the red and NIR bands (Tucker, 1979).

$$NDVI = \frac{NIR - red}{NIR + red} \quad (1)$$

The NDVI ranges between −1 and 1 and uses absorption peaks in the red band (630–690 nm) to detect photosynthetic active vegetation. It has been shown that low NDVI values of 0.25 and less correspond to bare soil observations in most cases (Demattê et al., 2018). For dark and carbon-rich soils in North West Germany, however, we observed that higher NDVI thresholds are necessary to cover all cropland areas (Fig. A1). Based on visual interpretation, we defined an NDVI threshold of 0.45 to increase the total area of the SRC and to include the full range of SOC values in cropland soils.

The NDVI is not sufficient to detect bare soil, as the presence of non-

photosynthetic vegetation (NPV) like crop residues can reduce the signal (Demattê et al., 2018). Previous studies have shown that the Normalized Burn Ratio 2 (NBR2) (Van Deventer et al., 1997) can be used to reduce the effects of NPV and soil moisture (Castaldi et al., 2019; Demattê et al., 2018; Dvorakova et al., 2022):

$$NBR2 = \frac{SWIR1 - SWIR2}{SWIR1 + SWIR2} \quad (2)$$

Absorption peaks in the SWIR2 band (2080–2350 nm) can be related to the presence of Lignin and Cellulose and therefore be used to detect NPV, like crop residues (Demattê et al., 2018). Similar is the case of the water content, with lower NBR2 values generally corresponding to lower levels of soil moisture. This is important as the soil water can decrease the soil reflectance and affect the correlation to SOC (Stenberg, 2010). Lowering the NBR2 thresholds has therefore been shown to improve the model performance (Castaldi et al., 2019; Dvorakova et al., 2022; Vaudour et al., 2021). However, this also leads to a trade-off, as lower thresholds reduce the extent of the mapping area and exclude certain soil conditions. On one hand, an NBR2 threshold of 0.05 has been identified as optimal for separating dry bare soil observations from unfavorable conditions like crop residues and soil moisture (Dvorakova et al., 2022). On the other hand, it was shown that NBR2 thresholds of up to 0.15 are necessary to achieve a full representation of the soil surface (Demattê et al., 2018; Safanelli et al., 2020). As our goal was to maximize the extent of the SOC map, an NBR2 threshold of 0.16 was identified to include the full range of SOC conditions within Germany. This was done with a visual interpretation, specifically looking for soils with naturally high NBR2 values, like floodplains with high soil moisture contents.

The Blue Chromatic Coordinate (BCC), based on the blue, green, and red band, was used to filter artificial structures and improve the soil mask (Gillespie et al., 1987):

$$BCC = \frac{blue}{red + green + blue} \quad (3)$$

This includes objects like cropland foliage, photovoltaic systems, or greenhouses with absorption peaks in the blue band. Similarly, an index based on the difference between the green and blue bands has been used to improve the soil mask in previous studies (Demattê et al., 2018; Fiorio and Demattê, 2009). Based on visual interpretation, observations with BCC values above 0.3 were masked out.

Multiple studies reported a trade-off between the extent of the SRC and the model accuracy (Castaldi et al., 2019; Dvorakova et al., 2022; Vaudour et al., 2021). It has been shown that lower NBR2 thresholds improve the quality of the SRC while also being selective towards certain soil types, thus reducing the overall mapping area. This is especially problematic in large research areas with diverse soil conditions, potentially masking out significant parts of cropland soils. To reduce this bias, we deployed dynamic NBR2 thresholds to select the final observations for the SRC. First, the fixed NBR2 threshold (0.16) was used to define the extent of the SRC. This threshold is higher than the recommended value but is necessary to include all soil types (Demattê et al., 2018) (Fig. A1). Next, we implemented dynamic NBR2 thresholds for each pixel to improve the quality of the considered bare soil observations without reducing the extent of the SRC. This was done by assessing the NBR2 statistics of each pixel and only including observations below the 40th percentile. Using this method, the NBR2 values of the SRC are lowered and thresholds can be dynamically assigned to different soil conditions (Fig. A2). The 40th percentile was selected as a cutting point as lower values only marginally affected the NBR2 values of the SRC while reducing the number of bare soil observations. If observations with low NBR2 are available, the final dynamic threshold will be significantly lower than the fixed threshold (Fig. A2). This can help to improve the soil reflectance quality, as shown by Dvorakova et al., (2022). If no observations with low NBR2 are available, it can be assumed that the natural bare soil NBR2 values are closer to the fixed threshold. The final SRC used for the prediction models is shown in Fig. 4.

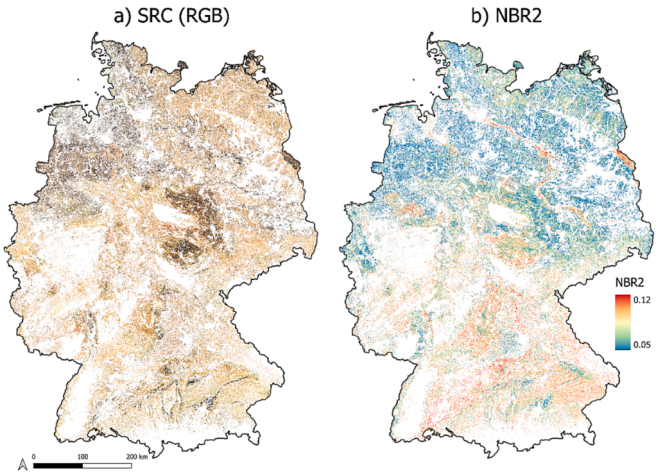


Fig. 4. a) Final SRC based on Landsat data from 2011 to 2022, RBG = Red-Blue-Green composite; b) Number of bare soil observations for each pixel.

Table 2
Overview of the covariates used for the SOC models: G = Global Model, G + D = Global + Distance, L = LEM, L + D = LEM + Distance.

Variable	Abbreviation	Wavelength (μm)	Models
SRC Band 1 — Blue	Blue	0.45–0.51	G / G + D / L / L + D
SRC Band 2 — Green	Green	0.53–0.59	G / G + D / L / L + D
SRC Band 3 — Red	Red	0.64–0.67	G / G + D / L / L + D
SRC Band 4 — Near Infrared	NIR	0.85–0.88	G / G + D / L / L + D
SRC Band 5 — Shortwave Infrared 1	SWIR1	1.57–1.65	G / G + D / L / L + D
SRC Band 6 — Shortwave Infrared 2	SWIR2	2.11–2.29	G / G + D / L / L + D
SRTM — Digital Elevation Model	DEM	–	G / G + D / L / L + D
Y Coordinates	Distance	–	G + D / L + D
X Coordinates	Distance	–	G + D / L + D
Euclidean distance — upper left	Distance	–	G + D / L + D
Euclidean distance — upper right	Distance	–	G + D / L + D
Euclidean distance — lower left	Distance	–	G + D / L + D
Euclidean distance — lower right	Distance	–	G + D / L + D
Aggregated local SOC predictions	Local_model	–	L / L + D

2.4.2. *Additional masking*

Additional masking was performed to remove bare soil observations outside cropland areas. This includes land use types such as forests (e.g. after clear-cuts), urban areas (e.g. roads and fallow ground), and permanent crops such as vineyards, or orchards. A digital landscape model (ATKIS Basis-DLM) (AdV, 2023) was used to mask out all areas, which were not classified either as agricultural cropland or grassland. Misclassified pixels inside permanent grassland areas were rare and mostly included single observations. It has been shown that the predictive power of the SRC can be improved by excluding pixels with not enough bare soil observations (Dvorakova et al., 2022). Based on this, only pixels with a minimal number of six observations were kept.

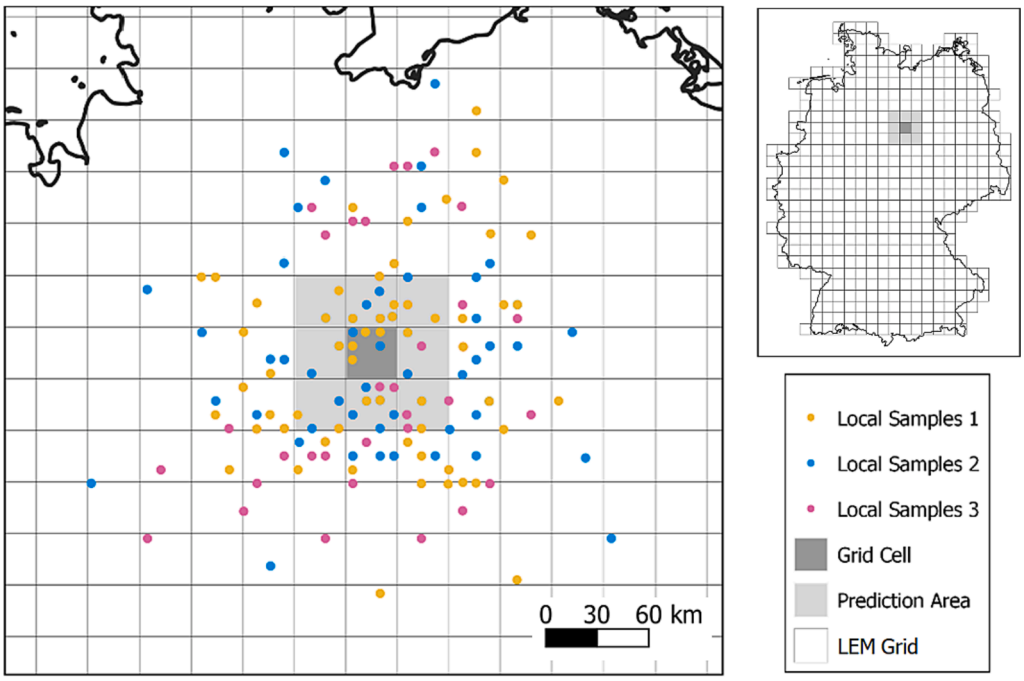


Fig. 5. Sampling and prediction process for the local models: For each LEM grid tile (dark grey), 60 nearby samples were randomly selected. This process was repeated three times (colored dots) to obtain three local training sets. Each local training set was used to train a GLM and predict the SOC in nine tiles (light grey). In the end, 27 predictions were available for each grid tile and averaged to obtain the covariate layer, used for the final SOC model.

2.5. Soil organic carbon prediction models

Models were trained using the SRC, the DEM, and 1966 samples from the soil inventory BZE-LW to predict the SOC of cropland soils in Germany. We compared the results of the Local Ensemble Model (LEM), based on the prediction and aggregation of multiple local models, to the results using a single model for the whole research area (Global Model). Additionally, we tested the effect of the integration of distance covariates on both models (LEM + Distance; Global + Distance). A summary of the covariates is provided in Table 2.

2.5.1. Global model

Model building was conducted in R (R Core Team, 2022), using the packages mlr3 (Lang et al., 2019) and ranger (Wright and Ziegler, 2017). For the global models, a single Random Forest (RF) model was built to predict the SOC across the research area (Breiman, 2001). Two cases were tested: (1) Only the six spectral bands of the SRC, as well as the elevation data from the DEM, were used for model training (Global Model). (2) Distance covariates were added to test the effect on the model performance and spatial autocorrelation (Global Model + Distance). For this, six Euclidian distance fields (EDF) were calculated, as described by Behrens et al. (2018): Two for the X and Y coordinates and one, using the Euclidian distances to the four coordinates at the corners of the research area. For both models, RF regressions were conducted using the ranger package and the following hyperparameters: num.trees = 500; mtry = square root of the no. variables; min.node.size = 5, max. depth = unlimited; min.bucket = 1. A forward feature selection (FFS) was conducted with mlr3 (Lang et al., 2019) as proposed by Meyer et al. (2019). Additional hyperparameter tuning was tested but only achieved marginal accuracy improvements, following the results of the FFS. Because of this, all RF models were trained with the above-mentioned parameters.

2.5.2. Local Ensemble model

We tested the LEM as a new framework to integrate spatial information into DSM and account for regional effects and spatial heterogeneity. An overview of the modeling steps is provided in Fig. 3. The LEM was conducted in three parts: First, multiple geographically distinct sub-models were trained and used for local SOC predictions (Fig. 3b). This is done by randomly selecting nearby samples based on their distance to the local models. The results of the sub-models are then aggregated into a single covariate layer. Lastly, the local predictions are combined with the SRC and the DEM and used to predict the final SOC map (Fig. 3c). In the following, the three steps are explained in detail.

1) Local sample selection

To define the boundary of the local models, the research area was divided into a 30 times 30 km grid (Fig. 5). This structure was selected to match the processing units of FORCE (Frantz, 2019) and to speed up the computational time of the LEM. Comparable to Georganos et al. (2021), only nearby observations were used to train the local models. For each grid tile, a total of three local training sets were randomly selected from the full data set with replacements (Fig. 5). The size of the local training data was limited to 60 samples. This number was selected to be small enough to represent the local soil conditions while also being large enough to prevent overfitting from insufficient training data. Instead of using a fixed maximum distance to select nearby samples for the local model, a weighted random selection was conducted based on the relative distance of the samples to the LEM tiles (Fig. 5). To ensure that nearby samples are more likely to be drawn for the local models and that the probability decreases rapidly with increasing distance, the distance weights were multiplied by the power of ten. Compared to a selection based on fixed distances, this method can be easily applied to select multiple randomized local training sets in regions with varying sampling densities. As illustrated in Fig. 5, not all the local samples are located within the prediction area. For SOC models trained on SRCs, it has been shown that additional samples can improve prediction accuracy, even if they come from outside the research area (Broeg et al., 2023).

2) Local prediction and aggregation

In the next step, local models were calibrated for each grid tile based on the randomly selected local samples and the SRC (Fig. 3b). The local predictions were conducted using GLMs instead of RF models to improve the spatial transferability of the results. This is necessary because the local models were only calibrated with a relatively small subset of local samples (60) and tree-based learners, like RF models, are unable to

predict SOC values outside the range of training samples. Instead, GLMs were trained to fit local regressions that can be applied for spatial extrapolation and be used to analyze local dependencies between the SRC and SOC samples. In contrast to linear regression, GLMs do not assume that the residuals are normally distributed (Gaussian) but rather follow different exponential distributions (Gamma, Poisson, Bernoulli, etc.) that can vary for each observation. The varying distributions are then connected to the covariates using a link function, which allows the regressions to be generalized and possible non-normal responses to be modeled. Before applying the GLMs, the SRC was preprocessed using a moving window average of 3x3 pixels as proposed by Chabrillat et al. (2019). This was done to reduce the noise at the pixel level and to prevent the risk of overfitting, introduced by the small number of training samples in the local models. The local predictions were carried out in the corresponding grid tile and the surrounding eight neighboring tiles (Fig. 5). This overlap is necessary to ensure that the aggregated local SOC predictions result in a seamless map, without artifacts at the tile borders. Using this method, a total of 27 local predictions (nine tiles + three repeats) were obtained for each inner tile, surrounded by eight other tiles. Accordingly, this number will be lower for tiles at the edge of the research area that are surrounded by less than eight tiles. These local predictions were then averaged for each tile to generate a single local prediction covariate layer to be used for the final RF models.

3) Final prediction

An RF model was used to predict the final SOC map of the LEM. The aggregated local prediction map was used as a covariates layer and combined with the SRC, the DEM, and all available soil samples. This step was conducted for multiple reasons: First, the local GLMs predicted values outside the training data, which led to unrealistic SOC values in some cases (e.g. SOC below 0). Using an RF model, the local predictions were adjusted to always fall within the range of SOC values in the training data. Secondly, a final SOC model trained with all soil samples can add information on general correlations, that are not recognized by the local models. A forward feature selection was conducted to identify and select the optimal set of covariates. As with the global model, additional distance covariates were included to assess the effect on the model performance (LEM + Distance).

2.6. Variable importance

The local variable importance of the GLMs was assessed for each grid tile. The absolute values of the GLM regression coefficients were used to calculate the relative importance of each band. The results of the three model repetitions were averaged to obtain the final local variable importance.

The variable importance of the final RF models was assessed using the importance mode "impurity", as implemented in the ranger package (Wright and Ziegler, 2017). The method is based on impurity reduction to calculate the relative importance of each variable for the regression trees (Louppe, 2015). Starting from the first split, the purity of the training samples is calculated for each decision node. Variables leading to purer nodes are considered more important and better predictors. The impurity reduction of each split can therefore be aggregated and used as a proxy for variable importance.

2.7. Model accuracy and uncertainty

Model accuracy and variable importance are based on 10-fold cross-validation. The data was split randomly to ensure samples in each fold were uniformly distributed across the research area. The validation data of each fold was excluded from the LEM and only used for the final accuracy assessment (Fig. 3). This is important to prevent potential bias, introduced by the fact that local GLMs and final RF models are trained with the same set of soil samples. Because of this, the full workflow of the LEM was conducted ten times, aggregating the results for the final accuracy assessment. Based on the validation data, R^2, Root Mean Square Error (RMSE), and Lin's concordance correlation coefficient (CCC) were calculated, as proposed by Hengl et al. (2018). The Ratio of Performance to Deviation (RPD) and Ratio of Performance to Interquartile Range (RPIQ) were included to improve the comparability with previous studies, even though it has been shown that the results are redundant to the R^2 (McBratney and Minasny, 2013). Uncertainty maps were derived from the RF models, as described by Hengl et al. (2018). For this, quantile regression forests (QRF) were calculated to estimate the upper and lower limits of the prediction intervals. The resulting maps provide information on the model error (g kg^{-1}) within one standard deviation (±1 σ) of the SOC predictions. To assess the spatial autocorrelation of the cross-validation residuals, variograms were calculated using the gstat package in R (Gräler et al., 2016).

Table 3
Results of the final RF models. Accuracies were calculated using all 1966 samples and a 10-fold cross-validation.

Model	R^2	RMSE (g kg^{-1})	CCC	RPIQ	RPD
Global Model	0.52	6.79	0.68	1.14	1.44
Global + Distance	0.57	6.4	0.72	1.2	1.53
LEM	0.67	5.63	0.80	1.37	1.74
LEM + Distance	0.68	5.55	0.80	1.39	1.76

3. Results

3.1. Soil organic carbon model accuracy

The accuracies of the different SOC models are presented in Table 3. The global model, trained with a single RF model and all soil samples, showed the overall lowest model performance (R^2 = 0.52; RMSE = 6.79). The implementation of the LEM, based on the aggregation of local sub-models, was able to improve the overall prediction accuracy (R^2 = 0.67; RMSE = 5.63). Both models reacted differently to the addition of the distance covariates: While the prediction accuracy of the global model improved (R^2 = 0.57; RMSE = 6.4), only a marginal change was visible for LEM (R^2 = 0.68; RMSE = 5.55).

Scatterplots and regression lines of the cross-validated SOC predictions are shown in Fig. 6a. The models produced similar results and generally underpredicted the SOC contents above 30 g kg^{-1}. For comparison, the local regression lines of all models are shown in Fig. 6b. Here, it is visible that all models showed a similar regression for low SOC values (below 30 g kg^{-1}) but that the LEM generally produced better results for high SOC values. This is indicated by the fact that the local regression lines of the global models are deviating further from the 1:1 line. For the LEM and the global model, the addition of distance covariates only marginally influenced the results (Fig. 6b).

Variograms were calculated to test the degree of spatial autocorrelation in the cross-validation residuals (Fig. 6c). If unexplained spatial structures remain in the SOC predictions, the semivariance of the residuals increases with sample distance, indicating the presence of spatial autocorrelation (Hengl et al., 2018; McBratney et al., 2003). This is mainly the case for the global model, as illustrated by the increasing semivariance for distances higher than 300 km. A similar, but generally lower spatial autocorrelation was visible for the global + distance model. Contrary, the cross-validation residuals of the LEM show no spatial autocorrelation (Fig. 6c). In this case, the addition of distance covariates only marginally affected the results.

The spatial distribution of the prediction residuals is shown in Fig. 7. In the case of the global model, it is visible that underpredictions are present throughout Germany but clustered towards the northwest and the south. Contrarily, overpredictions are mostly apparent in the north and northeast. A similar pattern is visible in the residuals of the global + distance model. In contrast to the global model, the number and magnitude of the residuals decreased with the integration of the LEM. In direct comparison, underpredictions are less common in the northwest

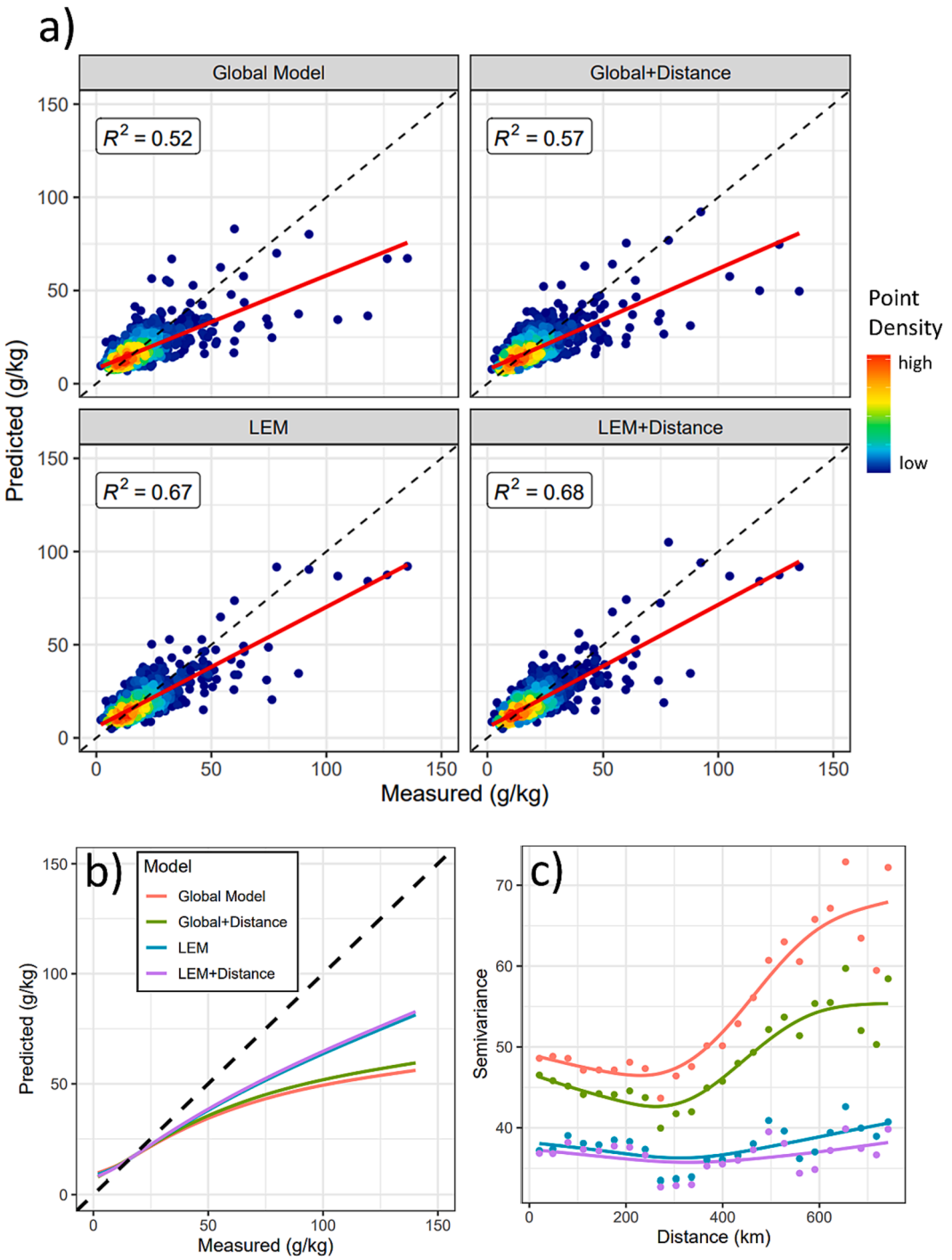


Fig. 6. a) Scatter plots and R^2 values of the cross-validated predictions and the measured topsoil SOC contents (g kg^{-1}). Red = regression lines, black = 1:1 lines; b) Direct comparison of the local regression lines (LOESS), based on the predictions of each model. Black = 1:1 line; c) Variograms, based on the cross-validation residuals of each model.

and south of Germany. Instead, the number of overpredictions slightly increased in these regions. Overpredictions in the northeast are far less common than in the residuals of the global model.

3.2. Final soil organic carbon predictions

The final SOC map of the LEM is presented in Fig. 8. The distribution of SOC varies across the research area and the highest values were predicted in northwest, south, and central Germany. Large areas with lower SOC values are mainly present in northeast and west Germany. Differences between the predictions of the LEM and the two global models are shown in Fig. 9. Here, it is visible that the LEM generally predicted higher SOC contents in the south and northwest of Germany. This tendency was more pronounced in the global model, especially in the northwest of Germany. In contrast, the global model predicted higher SOC values in the east, where generally lower SOC values are present. Small-scale differences between the models are illustrated in Fig. 9a + b. The LEM predicted higher extreme values and increased the local range of SOC contents (Fig. 9a). The integration of distance generally increased the predictions in some regions, leading to a smaller

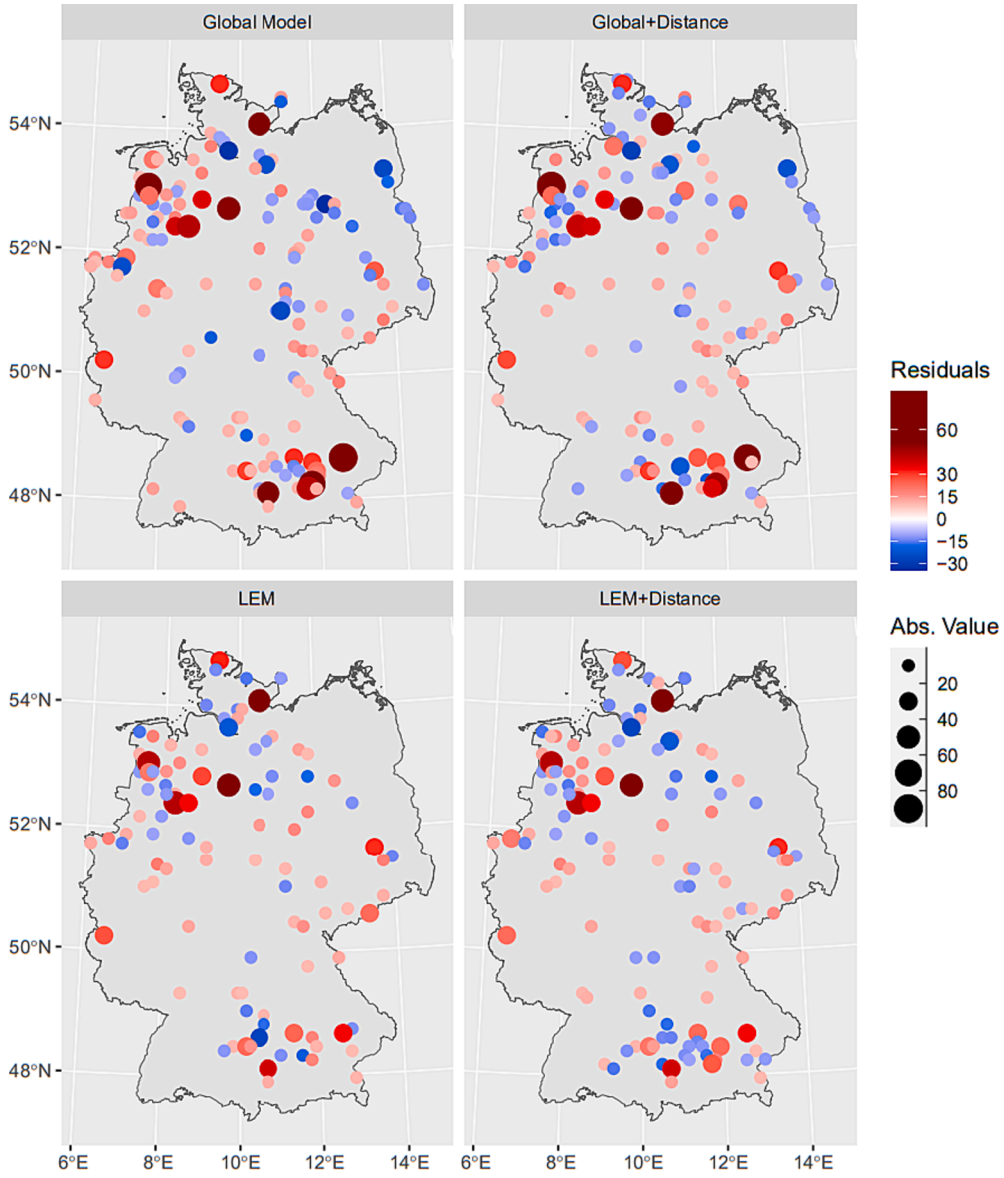


Fig. 7. Spatial distribution of the cross-validation residuals of the topsoil SOC content (g kg^{-1}). Only residuals with absolute values above 10 g kg^{-1} are displayed. Red = underpredictions, Blue = overpredictions.

range of local SOC values in comparison to the LEM (Fig. 9b). Uncertainty maps of the SOC predictions are shown in Fig. 10. The prediction intervals show a similar spatial pattern in all models and are strongly connected to the SOC content (Fig. 8) and the NBR2 values of the SRC (Fig. A2). In comparison to the global model, the LEM produced lower uncertainties throughout the research area. However, large prediction intervals remain present in regions with high SOC contents, especially in northwest and south Germany.

3.3. Variable importance of the final soil organic carbon models

The relative variable importance of the final RF models is shown in Fig. 11. The red bad was by far the most important variable for the global model (>30 %). The integration of distance covariates slightly increased the importance of the green band and decreased the importance of the other bands in favor of the distance covariates. The results from the local models were by far the most important variable for the LEM (>40 %). Here, the green, red, and SWIR bands showed similar importance around 10 %. The integration of the distance covariates decreased the importance of the most important bands but did not affect the importance of the local predictions. In this case, the distance covariates were the second most important with a combined importance of over 15 %.

3.4. Local model accuracy

The accuracies of the local sub-models are shown in Fig. 12. The results of the local GLMs were averaged to calculate the R^2 and RMSE of each grid tile. Both measures show clear but opposing spatial patterns across the research area. For the R^2, local values between 0.1 and 0.4 were observed. The highest values around 0.4 are present in north, south, and central Germany but are restricted to relatively small areas. In most parts of Germany, however, relatively low R^2 values between 0.2 and 0.3 were observed. In contrast, the lowest RMSE values are present in a large area ranging from northeast to southwest Germany, while the highest values are present in the southeast. The contrasting spatial patterns of the local R^2 and RMSE are underlined in Fig. 12c, illustrating a significant positive correlation between both measures.

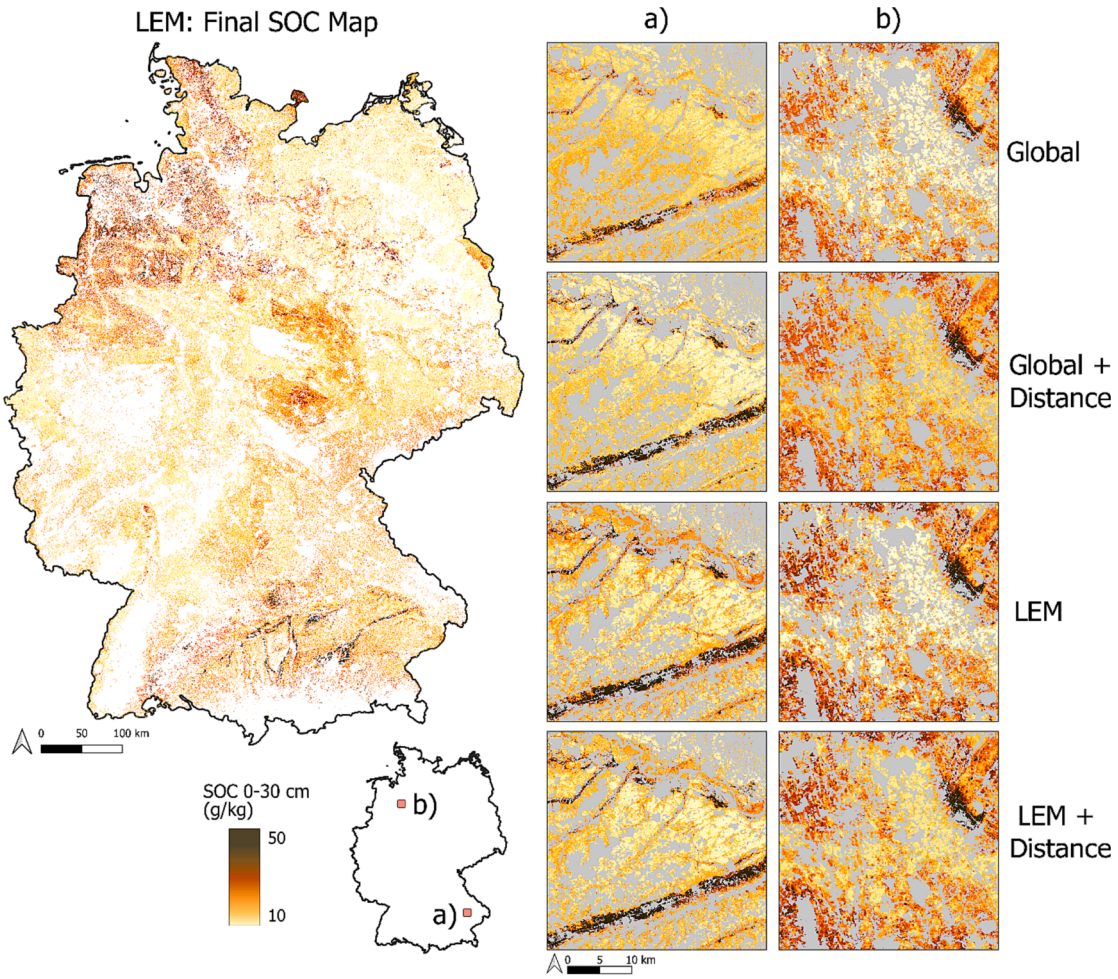


Fig. 8. Left: Final SOC map of the LEM; Right: Direct comparison between examples (30x30 km tiles) for the SOC predictions of the LEM and the global models. The sections were enlarged to showcase small-scale differences between the maps.

3.5. Local variable importance

Based on the GLMs, the local variable importance of the SRC was calculated for each grid tile (Fig. 13). The spectral bands show varying model importance and clear spatial patterns across the research area. In general, the green band was the most important for the local GLMs (Fig. 13c) with high values throughout Germany (Fig. 13a). The red and blue bands had similarly high local importance in some regions but showed an opposing spatial dependency. For the blue band, a clear gradient is visible throughout the research area: Starting with high local importance in the north, the values are steadily decreasing towards the south (Fig. 13a). The opposite is true for the red band, showing the highest importance in south and central Germany. The SWIR bands had an overall lower local importance and were most relevant in central Germany. Local results of the blue, red, and SWIR bands were combined into a ternary composition plot to highlight the spatial pattern (Fig. 13b). Based on this, three general model regions can be identified: north (Blue > Red/SWIR), south (Red > Blue/SWIR), and central (SWIR > Red/Blue) Germany.

4. Discussion

4.1. General model performance

In a comprehensive review of recent DSM publications, Chen et al. (2022) reported a median R^2 value of 0.49 for large-scale SOC maps (>10,000 km^2). A similar R^2 value of 0.52 was found for the global model, underlining the results of the review (Table 3). The implementation of the LEM improved the prediction accuracy and increased the R^2 by 31 % (R^2 = 0.68). Considering the large size of the research area (>350,000 km^2) and the results of the review, the LEM was able to produce a national cropland SOC map with an overall high accuracy (RMSE = 5.55 g kg^{-1}). Sakhaee et al. (2022) also predicted SOC for Germany using a traditional DSM approach without remote sensing data and reported an RMSE of 9.1 g kg^{-1} for a model trained with BZE-LW samples below 87 g kg^{-1}. Compared to this, our results highlight the benefit of using remote sensing data and SRCs to predict cropland SOC in large-scale research areas.

The difference between the results of the LEM and the global model can be attributed to multiple factors. As shown in Fig. 2, training samples above 40 g kg^{-1} were generally underrepresented in the global model, leading to poor prediction for high SOC values (Fig. 6b). With the integration of the LEM, it was possible to reduce this underrepresentation by dividing the research area into multiple sub-models and only considering a subset of local samples to train the models. This led to overall better predictions of SOC values above 40 g kg^{-1} (Fig. 6b). As shown in Fig. 9, the LEM predicted higher SOC values in south and northwest Germany and generally reduced the number of under-predictions (Fig. 7). Training the local models with a subset of the soil samples can therefore improve predictions in regions that would be otherwise underrepresented in the global model. This is also the case for regions with suboptimal conditions for the SRC, like the Oderbruch floodplain in northeast Germany. Here, the high soil moisture contents decrease the overall soil reflectance and the spectral SOC signal (Fig. 4)

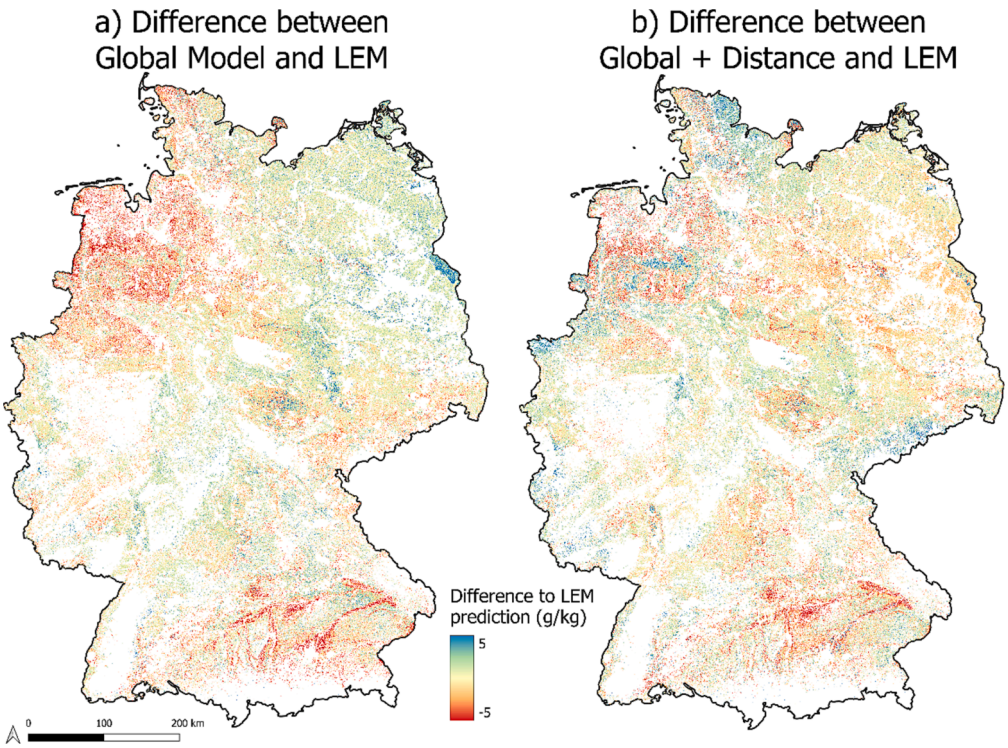


Fig. 9. a) Difference (g kg^{-1}) between the SOC predictions of the LEM and the global model; b) Difference (g kg^{-1}) between the SOC predictions of the LEM and the global + distance model; Red = lower predictions of the global models compared to LEM; Blue = higher prediction of the global models compared to LEM; Yellow = no difference

(Castaldi et al., 2019; Nocita et al., 2013; Stenberg et al., 2010). Consequently, the global model showed a strong tendency to overpredict the SOC concentrations in this region (Fig. 7). This was not the case for the LEM, as the range of SOC values used to train the local models was limited to nearby soil samples. Misclassifications between moist and carbon-rich soils were less likely and the LEM generally predicted lower SOC values (Fig. 9) leading to fewer overpredictions in the Oderbruch (Fig. 7). These differences between the predictions of the LEM and the global can also be illustrated with the uncertainty maps. As shown in Fig. 10, the LEM produced lower uncertainties in regions with high NBR2 values (Fig. A2), highlighting the ability of the local models to account for varying soil conditions.

Another factor contributing to the success of the LEM is the sampling design and extent of the soil inventory. Because of the regular sampling grid and the high spatial coverage, it was possible to repeatedly draw random samples for the local models, without the risk of oversampling certain areas. As illustrated in Fig. 5, spatial information was provided by using the distance between the local models and the soil samples as weights. Similar to bagging, this method of repeated random resampling is essential to reduce overfitting in ensemble models, especially because the aggregated local results were used as input for the final predictions (Sagi and Rokach, 2018; Taghizadeh-Mehrjardi et al., 2021).

4.2. Results of the local models

As illustrated by Georganos et al., (2021), geospatial machine learning can not only help to improve the model accuracy but can be also used as an explanatory tool to improve the understanding of spatial dependencies. In recent years, model interpretability has been highlighted as one of the most important challenges to enhance knowledge discovery in DSM (Arrouays et al., 2020; Chen et al., 2022; Padarian et al., 2020; Wadoux and Molnar, 2022). Besides interpretability, plausibility and explainability were emphasized as the main points that should be addressed when evaluating DSM models (Wadoux et al. 2020).

Plausibility

The plausibility is related to the fact that models should not only be accurate but also reflect the current state of scientific knowledge (Lipton, 2018; Wadoux et al., 2020). In soil spectroscopy, several studies tried to identify the most important spectral regions to predict the SOC in agricultural soils (Stenberg et al., 2010). It was concluded that the SOC absorptions in the VIS-NIR spectrum are diverse and depend on several soil factors, like texture and mineralogy. This is especially the case in cropland soils, as the SOC contents are generally lower than in other land use types and the absorption peaks can be masked out by the strong spectral effects of the soil matrix (Stenberg et al., 2010). For studies that used soil spectroscopy on a landscape level, it was concluded that SOC can be modeled based on the VIS-NIR spectrum but that the predictions are limited to geologically homogenous areas (Song et al., 2021; Udelhoven et al., 2003). This regional dependency between SOC and the VIS-NIR bands is supported by the local variable importance (Fig. 13). The results of the LEM show that the different spectral bands expressed clear spatial patterns throughout the research area and underline the necessity to adjust the SOC models to the local soil conditions.

Interpretability

The fundamental difference between the process of the global model and the LEM is illustrated by the variable importance of the final RF models (Fig. 11). For the global model, the red band was by far the most important predictor of SOC, masking most of the influence of the other bands. This highlights the fact that the global model was not able to sufficiently adjust to the heterogenous soil conditions. Instead, the red band was identified as the "lowest common denominator" to predict SOC across the research area. The dominance of a single band in the variable importance suggests that the general soil darkness was the main mechanism for SOC prediction in the global model. This hypothesis is backed by several studies, showing that soil darkness can be a basic predictor but is limited to soils with homogenous conditions (Hummel et al., 2001; Stenberg et al., 2010; Udelhoven et al., 2003). Contrary to

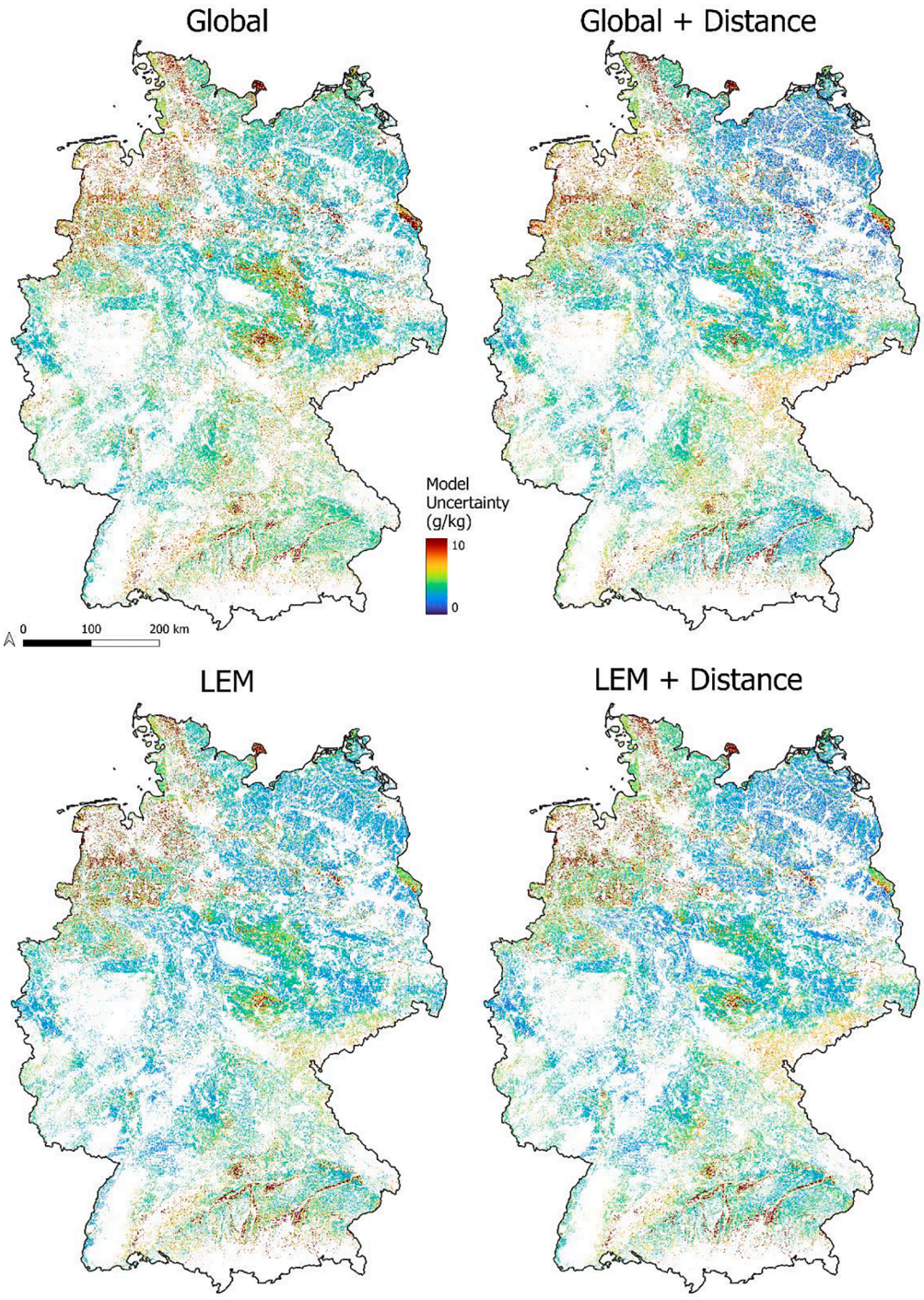


Fig. 10. Uncertainty maps of the final RF predictions: Model uncertainty (g kg^{-1}) within one standard deviation of the prediction interval.

the global model, the variable importance of the different spectral bands was more homogenous for the LEM (Fig. 11 + Fig. 13c) and expressed clear spatial patterns across the research area (Fig. 13a + b). Instead of relying on a single model to predict across all soil conditions, the local models were optimized to account for varying correlations between SOC and the SRC.

The results of the LEM revealed a spatial dependency between the RMSE and the R^2 of the local models (Fig. 12c). Similar results have been observed in soil spectroscopy, showing that both, RMSE and R^2, are increasing with the standard deviation of the training samples (Stenberg et al., 2010). This spatial dependency of the local model results highlights the importance of analyzing the soil heterogeneity within the research area (Fig. 12). The broad range of SOC values in south and northwest Germany increased the R^2 values of the GLMs, however, it also increased the RMSE of the local SOC predictions. High model errors in these regions could be attributed to the fact that the GLMs are unable to produce accurate predictions if the standard deviation of the training samples is too high. Considering the relatively low number of training samples used for the local models, it is important to take into account the distribution of SOC values, as outsiders could significantly influence the regressions and predictions of the GLMs. This is also visible in Fig. 10, underlying that the spatial uncertainty of the LEM shows a similar pattern to the RMSE values of the local GLMs.

Explainability

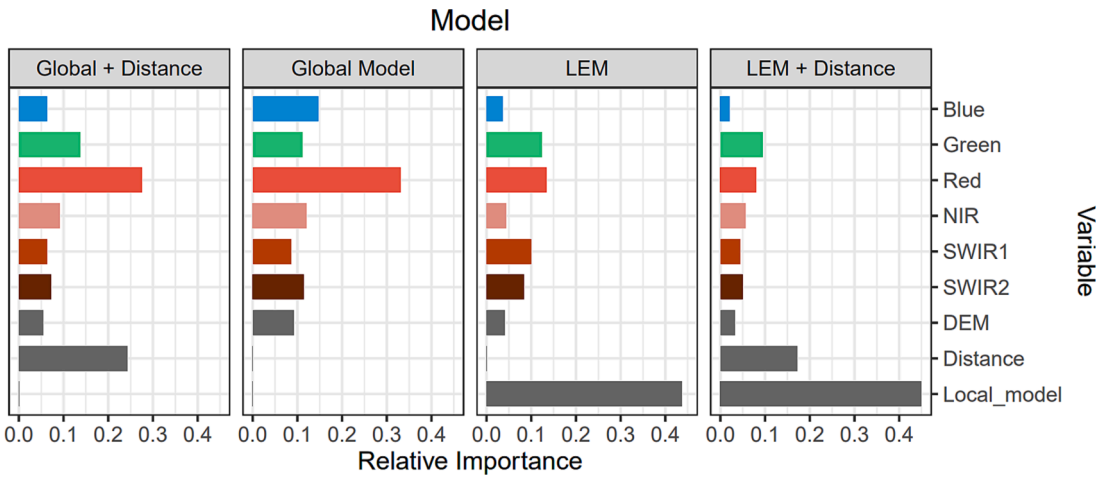


Fig. 11. Relative variable importance of the final SOC models; Blue/Green/Red/NIR/SWIR1/SWIR2 = SRC bands; Distance = aggregated importance of the distance covariates; Local_model = aggregated predictions of the local GLMs; DEM = digital elevation model.

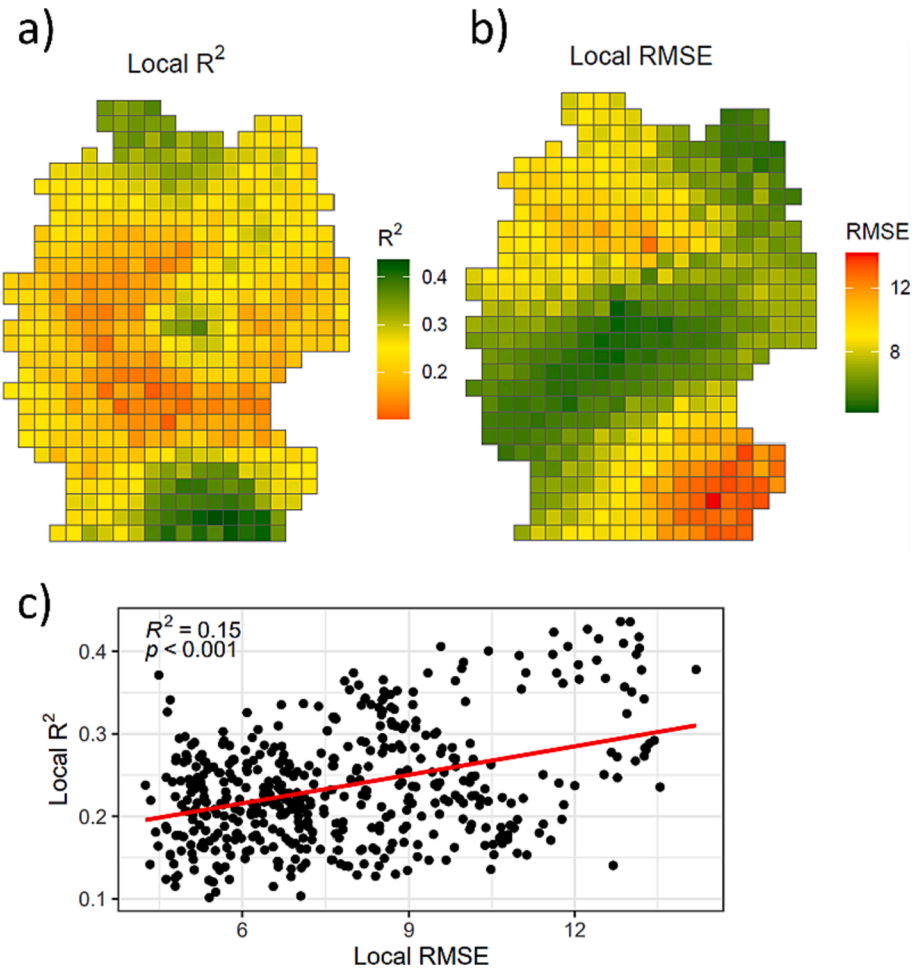


Fig. 12. Model accuracies of the local GLMs. Mean values of all model repetitions were calculated for each tile: a) Local R^2; b) Local RMSE; c) Correlation between the local R^2 and local RMSE. Red = regression line.

It has been shown that factors like soil texture, moisture, or soil chemical properties can alter the correlation between SOC and the spectral signal (Chabrillat et al., 2019; Demattê et al., 2018; Stenberg et al., 2010). The results of LEM can be combined with pedological knowledge to understand the underlying dependencies between the SRC and SOC. As shown in the local variable importance analysis, the blue, red, and SWIR bands expressed different spatial patterns across the research area (Fig. 13). For the sandy soil of the North German Plain (Fig. 1), the blue band was the most important feature in predicting SOC. Studies showed that SOC is the main absorbent in the VIS spectrum of sandy soils (Stenberg et al., 2010). High quartz content can therefore lead to high reflectance values, indicating why the blue band was the

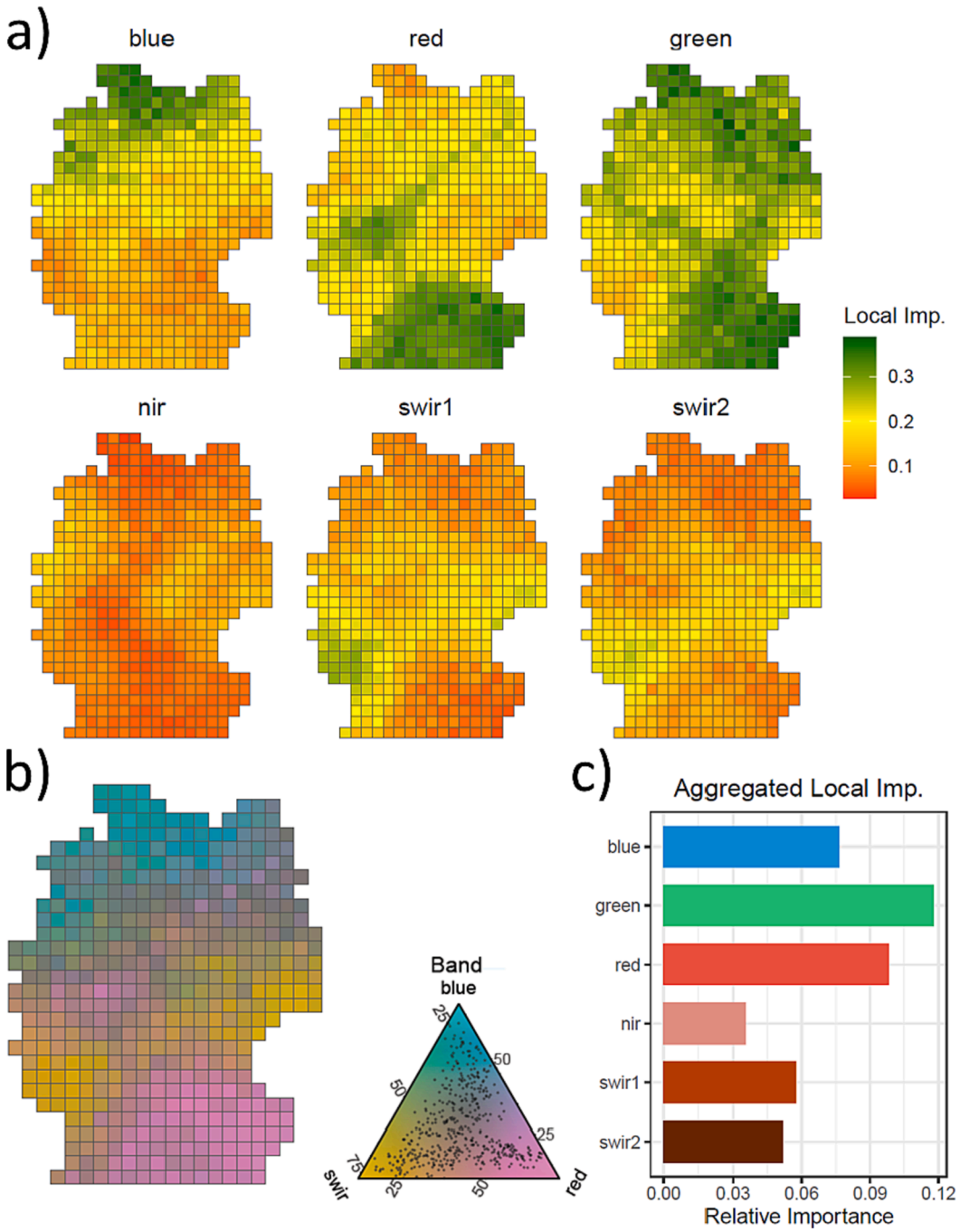


Fig. 13. Results of the local variable importance, based on the GLMs: a) Relative local importance of each spectral band; b) Ternary plot of the relative local importance of the blue, red, and SWIR bands. SWIR1 and SWIR2 bands were aggregated for the illustration. c) Aggregated local importance of all grid tiles.

most important to predict SOC in sandy, low-carbon soils. In contrast to quartz, clay minerals, like Kaolinite, express distinct absorption features in the VIS-NIR, as well as in the SWIR spectrum (Demattè et al., 2018; Stenberg et al., 2010). Previous studies proved the capacity of multi-spectral satellite data to predict the clay content in cropland topsoils (Gholizadeh et al., 2018; Gomez et al., 2022; Žížala et al., 2022). Accordingly, in our study, the highest local importance of the SWIR bands was found in central Germany, where loamy and clayey soils are more abundant (Fig. 1 + Fig. 13). The results show that the clay minerals are affecting the spectral signal and therefore influence the correlation between the SWIR bands and the SOC.

4.3. *Influence of the distance covariates*

Distance covariates have been commonly integrated into the DSM framework to improve model performance (Behrens et al., 2018; Møller et al., 2020). The main benefits arise from the reduction of the spatial autocorrelation of the prediction residuals (Hengl et al., 2018). This is supported by the results of the global model but does not apply to the results of the LEM (Table 3; Fig. 6). In the first case, the distance covariates improved both the overall model performance, as well as the spatial autocorrelation of the predictions. As illustrated in Fig. 11, they were important predictors and partially masked out the relevance of the spectral bands (Fig. 11). A similar concern was raised by Wadoux et al. (2020), who highlighted that the integration of distance covariates can

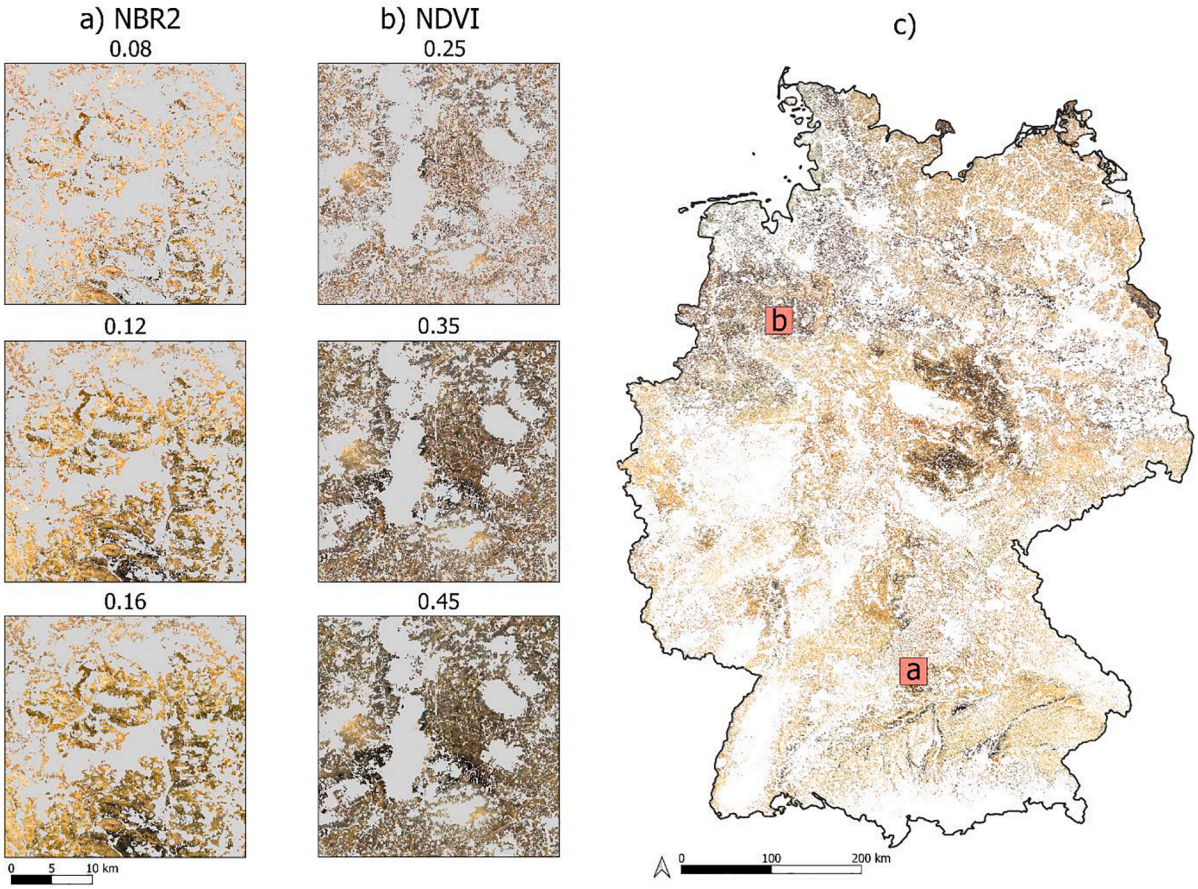


Fig. A1. Reference tiles highlighting the visual selection of NBR2 and NDVI thresholds for the generation of the SRC: Low thresholds (NBR2 = 0.08; NDVI = 0.25) significantly reduce the extent of the SRC by excluding soils with higher soil moisture and SOC contents. Higher thresholds (NBR2 = 0.16; NDVI = 0.45) were selected to remove bias and to include the full range of cropland soils.

be counterproductive for DSM models because of the missing pedological relevance. This conclusion is supported by the results of Meyer et al. (2019), showing that prediction artifacts can be introduced while improving the model accuracy at the same time. Fig. 8b illustrates that the inclusion of distance covariates generally increased SOC predictions in regions with high contents. Distance covariates are only able to describe general, large-scale patterns and therefore reduce the local range of SOC predictions. This is especially problematic for soil conditions that are underrepresented in the training data. In this case, the distance covariates could lead to overfitting, reducing the importance of the spectral bands and the ability to predict in unknown and underrepresented regions. In contrast to the global model, no spatial autocorrelation was present in the residuals of the LEM (Fig. 6b). The distance covariates only marginally improved the results but showed high variable importance for the final SOC predictions (Fig. 6+Fig. 11). These findings support the concern of Wadoux et al. (2020), showing that the distance covariates mask out the importance of the pedologically relevant covariates and therefore decrease the prediction quality. In general, the LEM was able to reduce the spatial autocorrelation to a greater extent than the implementation of distance covariates (Fig. 6c). The results illustrate that it is possible to integrate spatial information into the DSM framework, without relying upon additional distance covariates.

4.4. Influence of the soil reflectance composite

One of the main goals of this study was to generate comprehensive and large-scale information on cropland SOC in Germany. To achieve this, we decided to use relatively high NDVI and NBR2 thresholds to increase the extent of the SRC (Fig. A1). However, the presence of various soil conditions and management types remains a challenge, as they need to be represented with a single map. It is well known that soil properties can influence the generation of the SRC and the model accuracy, especially when high soil moisture contents are present (i.e. high NBR2 values) (Castaldi et al., 2019; Dvorakova et al., 2022; Vaudour et al., 2021). Instead of lowering the NBR2 threshold and removing these soils from the model completely, we decided to implement dynamic NBR2 thresholds to improve the quality of the SRC while maximizing its extent. As shown in Fig. A2, this method greatly reduced the NBR2 values of the SRC for most cropland soils in Germany. Regions in which the NBR2 values remained high after the dynamic thresholds were applied, for example floodplains or lowlands, are most affected by high soil moisture contents. Here, the model showed higher uncertainties since the spectral signal of the SOC is influenced by the presence of soil water (Fig. 10). Compared to the global model, the LEM was able to reduce the uncertainty in these regions by accounting for the local soil conditions. However, a potential trade-off between low model uncertainties (low NBR2/NDVI thresholds) and high model coverage (high NBR2/NDVI thresholds) remains present.

To maximize the number of bare soil scenes, the SRC was generated for a relatively long period (2011–2022), including observations throughout all seasons. This was done to improve the representation of different management types (e.g. summer/winter crops) and to reduce the influence of outliers when averaging the soil spectrum. As illustrated in Fig. 4, the number of bare soil observations is mainly influenced by the Landsat footprints, cropland management, and soil conditions, but shows no visible influence on model predictions and uncertainty (Fig. 8 + 10). However, multiple studies have highlighted a connection

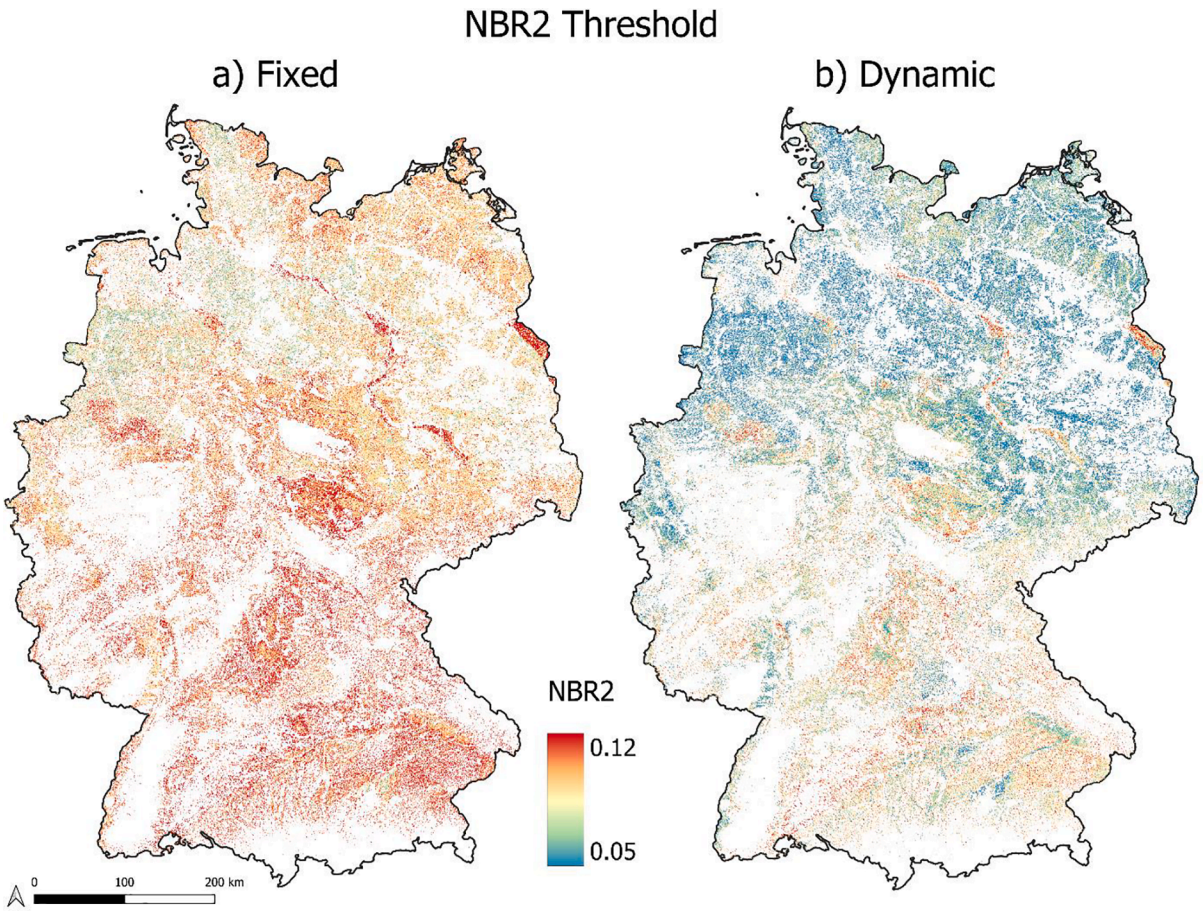


Fig. A2. Mean NBR2 values of the SRC before (a) and after applying the dynamic NBR2 thresholds (b).

between the model performance and the number of cloud-free, bare soil scenes per pixel (Castaldi et al., 2023; Dvorakova et al., 2022). Furthermore, it has been shown that a focus on specific management seasons (e.g. sowing in spring and plowing in autumn) can enhance the quality of the SRC (Castaldi et al., 2023; Urbina-Salazar et al., 2023). Additional research is necessary to test if such methods apply to improve SOC predictions in large research areas with diverse cropland and soil conditions.

5. Conclusion

We tested if the implementation of a LEM can improve the accuracy of large-scale SOC models based on SRCs. Our results show that the calculation and aggregation of multiple local sub-models can outperform the predictions of a single, global model. This improvement is attributed to multiple factors: First, the local models were able to account for spatial dependencies in the correlation between SOC and the SRC. The variable importance of the different spectral bands varies across the research area and spatial patterns can be explained by the local soil properties. Second, the LEM was able to adjust the models to differences in the local range of SOC contents. This led to better predictions for SOC values that were otherwise underrepresented in the global model and improved the overall accuracy. Last, the LEM reduced the spatial autocorrelation of the residuals by accounting for the spatial location of the soil samples. It represents a new method to integrate spatial information into the DSM framework, especially for large-scale research areas with heterogeneous soil conditions. Considering these findings, the LEM approach can improve the accuracy of national SOC maps while increasing the interpretability of the underlying modeling process.

CRediT authorship contribution statement

Tom Broeg: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft. **Axel Don:** Data curation, Investigation, Resources, Validation, Writing – review & editing, Funding acquisition. **Alexander Gocht:** Conceptualization, Funding acquisition, Investigation, Resources, Validation, Writing – review & editing. **Thomas Scholten:** Conceptualization, Investigation, Methodology, Supervision, Validation, Writing – review & editing. **Ruhollah Taghizadeh-Mehrjardi:** Conceptualization, Investigation, Methodology, Validation, Writing – review & editing. **Stefan Erasmi:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – review & editing.

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.

Data availability

The soil data used in this study is publicly available and can be downloaded from https://doi.org/10.3220/DATA20200203151139 (Poeplau et al., 2020a).

Acknowledgements

The project "KlimaFern" is funded by the Federal Ministry of Food and Agriculture as part of the German Climate Protection Programme 2022.

Appendix A

See Figs. A1 and A2.

References

AdV, 2023. Amtliches Topographisch-Kartographisches Informationssystem (ATKIS) [WWW Document]. URL https://www.adv-online.de/AdV-Produkte/Geotopographie/ATKIS/ (accessed 9.11.23).

Amelung, W., Bossio, D., de Vries, W., Kögel-Knabner, I., Lehmann, J., Amundson, R., Bol, R., Collins, C., Lal, R., Leifeld, J., Minasny, B., Pan, G., Paustian, K., Rumpel, C., Sanderman, J., van Groenigen, J.W., Mooney, S., van Wesemael, B., Wander, M., Chabbi, A., 2020. Towards a global-scale soil climate mitigation strategy. Nat. Commun. 11, 5427. https://doi.org/10.1038/s41467-020-18887-7.

Arrouays, D., McBratney, A., Bouma, J., Libohova, Z., Richer-de-Forges, A.C., Morgan, C. L.S., Roudier, P., Poggio, L., Mulder, V.L., 2020. Impressions of digital soil maps: the good, the not so good, and making them ever better. Geoderma Reg. 20, 00255. https://doi.org/10.1016/j.geodrs.2020.e00255.

Behrens, T., Schmidt, K., Viscarra Rossel, R.A., Gries, P., Scholten, T., MacMillan, R.A., 2018. Spatial modelling with Euclidean distance fields and machine learning. Eur. J. Soil Sci. 69, 757–770. https://doi.org/10.1111/ejss.12687.

Ben-Dor, E., Chabrillat, S., Demattê, J.A.M., Taylor, G.R., Hill, J., Whiting, M.L., Sommer, S., 2009. Using Imaging Spectroscopy to study soil properties. Remote Sens. Environ. 113, 38–55. https://doi.org/10.1016/j.rse.2008.09.019.

BGR, 2007. Bodenarten der Böden Deutschlands [WWW Document]. URL https://www.bgr.bund.de/DE/Themen/Boden/Informationsgrundlagen/Bodenkundliche_Karten_Datenbanken/Themenkarten/BOART1000OB/boart1000ob_node.html (accessed 9.11.23).

BGR, 2020. Bodenübersichtskarte 1:200.000 (BÜK200) [WWW Document]. URL https://www.bgr.bund.de/DE/Themen/Boden/Projekte/Informationsgrundlagen-laufend/BUEK200/BUEK200.html (accessed 9.11.23).

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

Broeg, T., Blaschek, M., Seitz, S., Taghizadeh-Mehrjardi, R., Zepp, S., Scholten, T., 2023. Transferability of covariates to predict soil organic carbon in cropland soils. Remote Sens. (Basel) 15, 876. https://doi.org/10.3390/rs15040876.

Castaldi, F., Chabrillat, S., Don, A., van Wesemael, B., 2019. Soil organic carbon mapping using LUCAS topsoil database and sentinel-2 data: an approach to reduce soil moisture and crop residue effects. Remote Sens. (Basel) 11, 2121. https://doi.org/10.3390/rs11182121.

Castaldi, F., Halil Koparan, M., Wetterlind, J., Žydelis, R., Vinci, I., Özge Savaş, A., Kıvrak, C., Tunçay, T., Volungevičius, J., Obber, S., Ragazzi, F., Malo, D., Vaudour, E., 2023. Assessing the capability of Sentinel-2 time-series to estimate soil organic carbon and clay content at local scale in croplands. ISPRS J. Photogramm. Remote Sens. 199, 40–60. https://doi.org/10.1016/j.isprsjprs.2023.03.016.

Chabrillat, S., Ben-Dor, E., Cierniewski, J., Gomez, C., Schmid, T., van Wesemael, B., 2019. Imaging spectroscopy for soil mapping and monitoring. Surv. Geophys. 40, 361–399. https://doi.org/10.1007/s10712-019-09524-0.

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

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.

Destatis, 2022. Land- und Forstwirtschaft, Fischerei - Bodenfläche nach Art der tatsächlichen Nutzung [WWW Document]. URL https://www.statistischebibliothek.de/mir/servlets/MCRFileNodeServlet/DEHeft_derivate_00071957/2030510217004.pdf (accessed 9.11.23).

Diek, S., Fornallaz, F., Schaepman, M.E., 2017. Barest pixel composite for agricultural areas using landsat time series. Remote Sens. (Basel) 9, 1245. https://doi.org/10.3390/rs9121245.

Dvorakova, K., Heiden, U., van Wesemael, B., 2021. Sentinel-2 exposed soil composite for soil organic carbon prediction. Remote Sens. (Basel) 13, 1791. https://doi.org/10.3390/rs13091791.

Dvorakova, K., Heiden, U., Pepers, K., Staats, G., van Os, G., van Wesemael, B., 2022. Improving soil organic carbon predictions from a Sentinel–2 soil composite by assessing surface conditions and uncertainties. Geoderma 116128. https://doi.org/10.1016/j.geoderma.2022.116128.

Farr, T.G., Kobrick, M., 2000. Shuttle radar topography mission produces a wealth of data. Eos Trans. AGU 81, 583–585. https://doi.org/10.1029/EO081i048p00583.

Fiorio, P.R., Demattê, J.A.M., 2009. Orbital and laboratory spectral data to optimize soil analysis. Sci. agric. (Piracicaba, Braz.) 66, 250–257. https://doi.org/10.1590/S0103-90162009000200015.

Frantz, D., 2019. FORCE—landsat + sentinel-2 analysis ready data and beyond. Remote Sens. (Basel) 11, 1124. https://doi.org/10.3390/rs11091124.

Frantz, D., Röder, A., Stellmes, M., Hill, J., 2016. An operational radiometric landsat preprocessing framework for large-area time series applications. IEEE Trans. Geosci. Remote Sens. 54, 3928–3943. https://doi.org/10.1109/TGRS.2016.2530856.

Frantz, D., Haß, E., Uhl, A., Stoffels, J., Hill, J., 2018. Improvement of the Fmask algorithm for Sentinel-2 images: separating clouds from bright surfaces based on parallax effects. Remote Sens. Environ. 215, 471–481. https://doi.org/10.1016/j.rse.2018.04.046.

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., Kalogirou, S., 2021. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto Int. 36, 121–136. https://doi.org/10.1080/10106049.2019.1595177.

Gholizadeh, A., Žižala, D., Saberioon, M., Borůvka, L., 2018. Soil organic carbon and texture retrieving and mapping using proximal, airborne and Sentinel-2 spectral imaging. Remote Sens. Environ. 218, 89–103. https://doi.org/10.1016/j.rse.2018.09.015.

Gillespie, A.R., Kahle, A.B., Walker, R.E., 1987. Color enhancement of highly correlated images. II. Channel ratio and "chromaticity" transformation techniques. Remote Sens. Environ. 22, 343–365. https://doi.org/10.1016/0034-4257(87)90088-5.

Gomez, C., Vaudour, E., Féret, J.-B., de Boissieu, F., Dharumarajan, S., 2022. Topsoil clay content mapping in croplands from Sentinel-2 data: influence of atmospheric correction methods across a season time series. Geoderma 423, 115959. https://doi.org/10.1016/j.geoderma.2022.115959.

Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-temporal interpolation using gstat. R J. 8, 204–218.

Heiden, U., d'Angelo, P., Schwind, P., Karlshöfer, P., Müller, R., Zepp, S., Wiesmeier, M., Reinartz, P., 2022. Soil reflectance composites—improved thresholding and performance evaluation. Remote Sens. (Basel) 14, 4526. https://doi.org/10.3390/rs14184526.

Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B.M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ 6, 5518. https://doi.org/10.7717/peerj.5518.

Hummel, J.W., Sudduth, K.A., Hollinger, S.E., 2001. Soil moisture and organic matter prediction of surface and subsurface soils using an NIR soil sensor. Comput. Electron. Agric. 32, 149–165. https://doi.org/10.1016/S0168-1699(01)00163-6.

Jacobs, A., Flessa, H., Don, A., 2018. Landwirtschaftlich genutzte Böden in Deutschland: Ergebnisse der Bodenzustandserhebung. Thünen-Report 64. https://doi.org/10.3220/REP1542818391000.

Jiang, Z., Li, Y., Shekhar, S., Rampi, L., Knight, J., 2017. Spatial ensemble learning for heterogeneous geographic data with class ambiguity. In: Hoel, E. (Ed.), Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, New York, NY, pp. doi: 1–10. 10.1145/3139958.3140044.

Kobayashi, S., Sanga-Ngoie, K., 2008. The integrated radiometric correction of optical remote sensing imageries. Int. J. Remote Sens. 29, 5957–5985. https://doi.org/10.1080/01431160701881889.

Lang, M., Binder, M., Richter, J., Schratz, P., Pfisterer, F., Coors, S., Au, Q., Casalicchio, G., Kotthoff, L., Bischl, B., 2019. mlr3: A modern object-oriented machine learning framework in R. J. Open Sour. Softw. https://doi.org/10.21105/joss.01903.

Lipton, Z.C., 2018. The Mythos of Model Interpretability: in machine learning, the concept of interpretability is both important and slippery. Queue 16, 31–57. https://doi.org/10.1145/3236386.3241340.

Louppe, G., 2015. Understanding Random Forests: From Theory to Practice. doi: 10.48550/arXiv.1407.7502.

McBratney, A., Minasny, B., 2013. Why you don't need to use RPD. Pedometron 33, 14–15.

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

Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications – Moving from data reproduction to spatial prediction. Ecol. Model. 411, 108815 https://doi.org/10.1016/j.ecolmodel.2019.108815.

Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Chapter One - Digital mapping of soil carbon. In: Sparks, D.L. (Ed.), Advances in Agronomy, Advances in Agronomy. Academic Press, pp. 1–47. https://doi.org/10.1016/B978-0-12-405942-9.00001-3.

Møller, A.B., Beucher, A.M., Pouladi, N., Greve, M.H., 2020. Oblique geographic coordinates as covariates for digital soil mapping. Soil 6, 269–289. https://doi.org/10.5194/soil-6-269-2020.

Nocita, M., Stevens, A., Noon, C., van Wesemael, B., 2013. Prediction of soil organic carbon for different levels of soil moisture using Vis-NIR spectroscopy. Geoderma 199, 37–42. https://doi.org/10.1016/j.geoderma.2012.07.020.

Padarian, J., Minasny, B., McBratney, A.B., 2020. Machine learning and soil sciences: a review aided by machine learning tools. Soil 6, 35–52. https://doi.org/10.5194/soil-6-35-2020.

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G.P., Smith, P., 2016. Climate-smart soils. Nature 532, 49–57. https://doi.org/10.1038/nature17174.

Poeplau, C., Don, A., Flessa, H., Heidkamp, A., Jacobs, A., Prietz, R., 2020a. Erste Bodenzustandserhebung Landwirtschaft – Kerndatensatz. doi: 10.3220/DATA20200203151139.

Poeplau, C., Jacobs, A., Don, A., Vos, C., Schneider, F., Wittnebel, M., Tiemeyer, B., Heidkamp, A., Prietz, R., Flessa, H., 2020b. Stocks of organic carbon in German

agricultural soils—Key results of the first comprehensive inventory. J. Plant Nutr. Soil Sci. 183, 665–681. https://doi.org/10.1002/jpln.202000113.
R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Roberts, D., Wilford, J., Ghattas, O., 2019. Exposed soil and mineral map of the Australian continent revealing the land at its barest. Nat. Commun. 10, 5297. https://doi.org/10.1038/s41467-019-13276-1.
Roe, S., Streck, C., Obersteiner, M., Frank, S., Griscom, B., Drouet, L., Fricko, O., Gusti, M., Harris, N., Hasegawa, T., Hausfather, Z., Havlík, P., House, J., Nabuurs, G.-J., Popp, A., Sánchez, M.J.S., Sanderman, J., Smith, P., Stehfest, E., Lawrence, D., 2019. Contribution of the land sector to a 1.5 °C world. Nat. Clim. Chang. 9, 817–828. https://doi.org/10.1038/s41558-019-0591-9.
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.
Roy, D.P., Zhang, H.K., Ju, J., Gomez-Dans, J.L., Lewis, P.E., Schaaf, C.B., Sun, Q., Li, J., Huang, H., Kovalskyy, V., 2016. A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. Remote Sens. Environ. 176, 255–271. https://doi.org/10.1016/j.rse.2016.01.023.
Royer, A., Charbonneau, L., Teillet, P.M., 1988. Interannual landsat-MSS reflectance variation in an urbanized temperate zone. Remote Sens. Environ. 24, 423–446. https://doi.org/10.1016/0034-4257(88)90017-X.
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 Sens. (Basel) 12, 1369. https://doi.org/10.3390/rs12091369.
Sagi, O., Rokach, L., 2018. Ensemble learning: a survey. WIREs Data Min. Knowl. Discovery 8, e1249.
Sakhaee, A., Gebauer, A., Ließ, M., Don, A., 2022. Spatial prediction of organic carbon in German agricultural topsoil using machine learning algorithms. Soil 8, 587–604. https://doi.org/10.5194/soil-8-587-2022.
Sekulić, A., Kilibarda, M., Heuvelink, G.B.M., Nikolić, M., Bajat, B., 2020. Random Forest Spatial Interpolation. Remote Sens. (Basel) 12, 1687. https://doi.org/10.3390/rs12101687.
Song, Y., Shen, Z., Wu, P., Viscarra Rossel, R.A., 2021. Wavelet geographically weighted regression for spectroscopic modelling of soil properties. Sci. Rep. 11, 17503. https://doi.org/10.1038/s41598-021-96772-z.
Stenberg, B., 2010. Effects of soil sample pretreatments and standardised rewetting as interacted with sand classes on Vis-NIR predictions of clay and soil organic carbon. Geoderma 158, 15–22. https://doi.org/10.1016/j.geoderma.2010.04.008.
Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Chapter Five - Visible and near infrared spectroscopy in soil science. In: Sparks, D.L. (Ed.), Advances in Agronomy. Academic Press, pp. 163–215. https://doi.org/10.1016/S0065-2113(10)07005-7.
Taghizadeh-Mehrjardi, R., Hamzehpour, N., Hassanzadeh, M., Heung, B., Ghebleh Goydaragh, M., Schmidt, K., Scholten, T., 2021. Enhancing the accuracy of machine learning models using the super learner technique in digital soil mapping. Geoderma 399, 115108. https://doi.org/10.1016/j.geoderma.2021.115108.
Thünen-Institut, 2022. Forstliche Großlandschaften (2011) [WWW Document]. URL https://atlas.thuenen.de/layers/wgwb:geonode:wgwb_forstl_gl_2011 (accessed 9.11.23).
Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0.
Udelhoven, T., Emmerling, C., Jarmer, T., 2003. Quantitative analysis of soil chemical properties with diffuse reflectance spectrometry and partial least-square regression: a feasibility study. Plant and Soil 251, 319–329. https://doi.org/10.1023/A:1023008322682.
Urbina-Salazar, D., Vaudour, E., Richer-de-Forges, A.C., Chen, S., Martelet, G., Baghdadi, N., Arrouays, D., 2023. Sentinel-2 and sentinel-1 bare soil temporal mosaics of 6-year periods for soil organic carbon content mapping in central France. Remote Sens. (Basel) 15, 2410. https://doi.org/10.3390/rs15092410.
Van Deventer, A.P., Ward, A.D., Gowda, P.M., Lyon, J.G., 1997. Using thematic mapper data to identify contrasting soil plains and tillage practices. Photogramm. Eng. Remote Sens. 63, 87–93.
Vaudour, E., Gomez, C., Lagacherie, P., Loiseau, T., Baghdadi, N., Urbina-Salazar, D., Loubet, B., Arrouays, D., 2021. Temporal mosaicking approaches of Sentinel-2 images for extending topsoil organic carbon content mapping in croplands. Int. J. Appl. Earth Obs. Geoinf. 96, 102277 https://doi.org/10.1016/j.jag.2020.102277.
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 Sens. (Basel) 14, 2917. https://doi.org/10.3390/rs14122917.
Wadoux, A.-M.-J.-C., Minasny, B., McBratney, A.B., 2020. Machine learning for digital soil mapping: applications, challenges and suggested solutions. Earth Sci. Rev. 210, 103359 https://doi.org/10.1016/j.earscirev.2020.103359.
Wadoux, A.-M.-J.-C., Molnar, C., 2022. Beyond prediction: methods for interpreting complex models of soil variation. Geoderma 422, 115953. https://doi.org/10.1016/j.geoderma.2022.115953.
Walter, C., McBratney, A.B., Douaoui, A., Minasny, B., 2001. Spatial prediction of topsoil salinity in the Chelif Valley, Algeria, using local ordinary kriging with local variograms versus whole-area variogram. Soil Res. 39, 259–272. https://doi.org/10.1071/sr99114.
Webster, R., Oliver, M.A., 1992. Sample adequately to estimate variograms of soil properties. J. Soil Sci. 43, 177–192. https://doi.org/10.1111/j.1365-2389.1992.tb00128.x.
Wright, M.N., Ziegler, A., 2017. ranger: A fast implementation of random forests for high dimensional data in C++ and R. J. Stat. Softw. 77, 1–17. https://doi.org/10.18637/jss.v077.i01.
Zepp, S., Heiden, U., Bachmann, M., Wiesmeier, M., Steininger, M., van Wesemael, B., 2021. Estimation of soil organic carbon contents in croplands of Bavaria from SCMaP soil reflectance composites. Remote Sens. (Basel) 13, 3141. https://doi.org/10.3390/rs13163141.
Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sens. Environ. 118, 83–94. https://doi.org/10.1016/j.rse.2011.10.028.
Žížala, D., Minařík, R., Skála, J., Beitlerová, H., Juřicová, A., Reyes Rojas, J., Penížek, V., Zádorová, T., 2022. High-resolution agriculture soil property maps from digital soil mapping methods, Czech Republic. CATENA 212, 106024. https://doi.org/10.1016/j.catena.2022.106024.