Remote Sensing-Based Monitoring of Cropland Soil Organic Carbon

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Tom Brög
aus Hamburg

Tübingen
2025

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	20.11.2025
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Thomas Scholten
2. Berichterstatter:	Prof. Dr. Volker Hochschild
3. Berichterstatter:	Prof. Dr. Christopher Conrad

Abstract

Agricultural soils are essential for food security, but are increasingly affected by soil degradation and the impacts of climate change. The sequestration of soil organic carbon (SOC) is recognized as an effective strategy to improve soil quality, mitigate climate change, and increase resilience to droughts and extreme weather events. EU policies such as the Carbon Removals and Carbon Farming (CRCF) regulation aim to promote climate-smart agricultural practices but require robust and cost-efficient systems to verify SOC changes. Remote sensing offers the potential to derive large-scale information on SOC status and trends. However, its application for soil monitoring is currently limited by the lack of validation for the temporal model accuracy. This dissertation investigates the potential of multispectral soil reflectance composites (SRC), derived from Sentinel-2 and Landsat time series, for large-scale soil mapping and spatiotemporal SOC modeling. The results show that SRCs are strong predictors of cropland soil properties and enable the generation of high-resolution SOC maps. Correlations between SRC bands and soil properties such as SOC and clay are consistent with findings from soil spectroscopy, supporting the calibration of robust and transferable models. Using high-quality reference samples from the German Agricultural Soil Inventory, it was demonstrated that the selection of spectral indices and thresholding values significantly affects the spectral quality and bias of the generated SRC. In areas with heterogeneous soil conditions, the prediction accuracy can be further improved by calibrating local sub-models to account for regional variability in the soil signal. Based on harmonized Landsat and Sentinel-2 time series from 1986 to 2022, a spatiotemporal model was developed to assess the potential of remote sensing data to detect SOC changes in cropland soils. For the first time, SOC trend predictions were validated using repeated measurements from long-term soil monitoring sites. While model accuracy was not sufficient to replace direct sampling, the validation showed a low confusion rate between increasing and decreasing SOC trends, underlining the potential of satellite-based models for long-term soil monitoring. The findings suggest that remote sensing can improve the cost-efficiency of SOC monitoring by enabling targeted sampling at locations where predicted trends diverge from expected effects of climate-smart management. Further research should aim to (1) improve the validity of modeled SOC dynamics by integrating additional soil data on bulk density and vertical SOC distribution, and (2) further advance the methods for the retrieval of soil reflectance to increase the signal-to-noise ratio and shorten the required time frames to detect significant SOC changes.

Zusammenfassung

Landwirtschaftliche Böden sind für die Ernährungssicherheit von entscheidender Bedeutung, sind jedoch zunehmend von Bodendegradation und den Auswirkungen des Klimawandels betroffen. Die Speicherung von organischem Bodenkohlenstoff (Soil organic carbon; SOC) gilt als effektive Strategie zur Verbesserung der Bodenqualität und zur Erhöhung der Resilienz gegenüber Dürren und extremen Wetterereignissen. Politische Initiativen wie die EU-Verordnung über Carbon Removals and Carbon Farming (CRCF) zielen darauf ab, Klimaschutzmaßnahmen zu fördern, erfordern jedoch belastbare und kosteneffiziente Systeme zur Verifizierung von SOC-Veränderungen. Fernerkundung bietet die Möglichkeit, großräumige Informationen über den Zustand und die Veränderung von SOC abzuleiten. Die Anwendbarkeit für das Bodenmonitoring wird jedoch derzeit durch eine fehlende Validierung der temporalen Modellgenauigkeit begrenzt. Diese Dissertation untersucht das Potenzial multispektraler Bodenreflektanzkomposite (BRK), abgeleitet aus Sentinel-2- und Landsat-Zeitreihen, für die Erstellung großflächiger Bodenkarten sowie dynamischer SOC-Modelle. Die Ergebnisse zeigen, dass BRK eine hohe Vorhersagekraft für die Erstellung hochauflösender SOC-Karten besitzen. Korrelationen zwischen den spektralen Bändern und Bodenparametern wie SOC und Tongehalt stimmen mit Ergebnissen aus der Bodenspektroskopie überein und ermöglichen die Kalibrierung belastbarer und übertragbarer Modelle. Anhand von Referenzdaten aus der Bodenzustandserhebung Landwirtschaft konnte gezeigt werden, dass die Auswahl spektraler Bodenindizes und Schwellenwerte einen signifikanten Einfluss auf die spektrale Qualität der BRK hat. In Untersuchungsgebieten mit heterogenen Bodenbedingungen kann die Modellgenauigkeit mit der Kalibrierung lokaler Submodelle weiter verbessert werden, um regionale Unterschiede im Bodensignal zu berücksichtigen. Auf Basis harmonisierter Landsat- und Sentinel-2-Zeitreihen von 1986 bis 2022 wurde ein dynamisches Modell entwickelt, um das Potenzial zur Detektion von SOC-Veränderungen in Ackerböden zu bewerten. Hierbei ist es erstmals gelungen, Vorhersagen von SOC-Trends anhand wiederholter Messungen aus Bodendauerbeobachtungsflächen zu verifizieren. Zwar zeigt die Validierung, dass die Modellgenauigkeit nicht ausreicht, um direkte Probenahmen zu ersetzen, jedoch bestehen nur geringe Verwechslungsraten zwischen zunehmenden und abnehmenden SOC-Trends. Die Ergebnisse unterstreichen somit das Potenzial fernerkundungsbasierter Modelle für ein großskaliges Bodenmonitoring. Auf Grundlage der resultierenden Karten ist es möglich, die Kosteneffizienz zu steigern, indem Probenahmen gezielt auf Standorte gerichtet werden, an denen die vorhergesagten SOC-Trends von den erwarteten Effekten der Klimaschutzmaßnahmen abweichen. Zusätzliche Forschung ist notwendig, um (1) die Validität der modellierten SOC-Dynamiken durch Integration weiterer Bodendaten (Lagerungsdichte, vertikale SOC-Verteilung, etc.) zu verbessern und (2) die Methoden zur Extraktion der Bodenreflektanz weiterzuentwickeln, um die Modellgenauigkeit zu erhöhen und den Zeitraum zur Detektion signifikanter SOC-Veränderungen zu verkürzen.

Table of contents

List of Figures

List of Tables

List of Acronyms

AF	Alpine Foreland
ANN	Artificial Neural Network
ARD	Analysis-Ready Data
BCC	Blue Chromatic Coordinate
BGL	Bodengroßlandschaften (Soil-landscape units)
BNIR	Broad Near-Infrared
BRDF	Bidirectional Reflectance Distribution Function
BSI	Bare Soil Index
BW	Baden-Wuerttemberg
BY	Bavaria
BZE-LW	Bodenzustandserhebung Landwirtschaft (German Agricultural Soil Inventory)
CAP	Common Agricultural Policy
CCC	Concordance Correlation Coefficient
CRCF	Carbon Removals Certification Framework
CSO	Clear-Sky Observations
CTM	Crop Type Map
CV	Cross-Validation
DEM	Digital Elevation Model
DSM	Digital Soil Mapping
DWD	Deutscher Wetterdienst (German Meteorological Service)
EDF	Euclidean Distance Fields
EH	Eastern Highlands
EnMAP	Environmental Mapping and Analysis Program
EO	Earth Observation

EU	European Union
FFS	Forward Feature Selection
FORCE	Framework for Operational Radiometric Correction for Environmental Monitoring
GCC	Green Chromatic Coordinate
GDAL	Geospatial Data Abstraction Library
GDR	German Democratic Republic
GHG	Greenhouse Gas
GLM	Generalized Linear Model
HDB	Humus Database
IACS	Integrated Administration and Control System
IPCC	Intergovernmental Panel on Climate Change
KSG	Klimaschutzgesetz (Federal Climate Action Act)
LEM	Local Ensemble Model
LH	Loess Hills
LOESS	Locally Estimated Scatterplot Smoothing
LTM	Long-Term Monitoring
LUCAS	Land Use and Cover Area Frame Survey
LULUCF	Land Use, Land Use Change, and Forestry
MESS	Multivariate Environmental Similarity Surfaces
ML	Machine Learning
NBR2	Normalized Burn Ratio 2
NDTI	Normalized Difference Tillage Index
NDVI	Normalized Difference Vegetation Index
NIR	Near-Infrared
NP	Northern Plain
NPV	Non-Photosynthetic Vegetation

OOB	Out-Of-Bag error
PCA	Principal Component Analysis
PV	Photosynthetic Vegetation
PV+IR2	Combined Vegetation and Shortwave Infrared Index
QRF	Quantile Regression Forests
RCC	Red Chromatic Coordinate
RF	Random Forest
RGB	Red-Green-Blue Composite
RMSE	Root Mean Square Error
SCMaP	Soil Composite Mapping Processor
SCORPAN	Soil, Climate, Organisms, Relief, Parent material, Age, Spatial position
SML	Soil Monitoring Law
SMOTE	Synthetic Minority Over-sampling Technique
SOC	Soil Organic Carbon
SOCS	Soil Organic Carbon Stocks
SRC	Soil Reflectance Composites
SS	Southern Scarplands
STM	Spectral-Temporal Metrics
SVM	Support Vector Machine
SWIR	Shortwave Infrared
TWI	Topographic Wetness Index
VBF	Valley Bottom Flatness
VIS	Visible Spectrum
VNIR	Visible and Near-Infrared
WH	Western Highlands
W_p	Wasserstein Distance

List of Publications

Accepted and published

- **Manuscript 1**: Broeg, T., Don, A., Gocht, A., Scholten, T., Taghizadeh-Mehrjardi, R., Erasmi, S., 2024. Using local ensemble models and Landsat bare soil composites for large-scale soil organic carbon maps in cropland. Geoderma 444, 116850. DOI:10.1016/j.geoderma.2024.116850
- **Manuscript 2**: Broeg, T., Don, A., Scholten, T., Erasmi, S., 2025. Reducing bias in cropland soil organic carbon and clay predictions using Sentinel-2 composites and data balancing. Remote Sensing of Environment 333, 115109. DOI:10.1016/j.rse.2025.115109
- **Manuscript 3**: 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 Sensing 15, 876. DOI:10.3390/rs15040876
- **Manuscript 4**: Broeg, T., Don, A., Wiesmeier, M., Scholten, T., Erasmi, S., 2024. Spatiotemporal monitoring of cropland soil organic carbon changes from space. Global Change Biology 30, e17608. DOI:10.1111/gcb.17608
- **Manuscript 5**: Lobert, F., Schwieder, M., Alsleben, J., Broeg, T., Kowalski, K., Okujeni, A., Hostert, P., Erasmi, S., 2025. Unveiling year-round cropland cover by soil-specific spectral unmixing of Landsat and Sentinel-2 time series. Remote Sensing of Environment 318, 114594. DOI:10.1016/j.rse.2024.114594

Submitted

- **Manuscript 6**: Blickensdörfer, L., Broeg, T., Lobert, L., Schwieder, M., Hostert, P., Erasmi, S., 2025. Leveraging multidecadal satellite time series to estimate grassland age on national scale. International Journal of Applied Earth Observation and Geoinformation (under review)

Chapter 1

Introduction

The degradation of agricultural land is a major global challenge, posing a risk to food security and threatening essential ecosystem functions that are provided by soils (Foley et al., 2005). As reported by the Intergovernmental Panel on Climate Change (IPCC), global cereal prices are projected to rise by up to 29% due to climate-induced extreme weather events such as droughts, floods, and altered precipitation patterns (IPCC, 2022). Combined with the expected population growth during the 21st century, these developments will intensify pressure on cropland soils and highlight the need for resilient and sustainable soil management and strategies (Paustian et al., 2016).

1.1 Soil organic carbon

Soil organic carbon (SOC) is a key component of the terrestrial carbon dynamics, forming the interface between the biosphere and atmosphere, and influencing global greenhouse gas (GHG) fluxes (Smith et al., 2008). The SOC stocks of agricultural soils are not static but controlled by the balance between carbon inputs (e.g., plant residues, root exudates, organic amendments) and losses (e.g., mineralization, leaching, erosion) (Don et al., 2024b). Land use is a key driver of this dynamic, determining whether SOC stocks accumulate and support the removal of atmospheric CO_2 or whether they decline and contribute to the net GHG emissions (Poeplau and Don, 2013).

Following centuries of agricultural intensification, cropland soils have experienced a substantial depletion of SOC, with estimates suggesting that up to 50% of the original stocks have been lost (Lal, 2003). Most of this observed decline can be associated with the historical conversion of forest and grassland to cropland, which typically results in lower organic inputs and higher mineralization rates, leading to decreasing SOC contents over time (Janzen, 2006). This is supported by long-term studies, showing that such trends can persist for several decades and must be considered when analyzing present-day SOC dynamics (Emde et al., 2024).

The deficit between the potential and actual carbon sequestration in agricultural soils is recognized as a mitigation opportunity that could substantially contribute to reaching the goal of climate neutrality (Minasny et al., 2017). In Germany, cropland soils are estimated to store approximately 2.5 petagrams (Pg) of carbon in the upper meter, which corresponds to nearly eleven times the country's annual CO_2 emissions (Poeplau et al., 2020b). In addition to serving as the largest terrestrial carbon pool, SOC is strongly linked to key soil functions such as water retention, aggregate stability, and nutrient cycling (Murphy, 2015). These functions not only significantly contribute to soil fertility but also support the resilience against climate pressures such as droughts, extreme weather events, or soil erosion (Webb et al., 2017).

Increasing the SOC in agricultural soils is therefore critical not only for climate mitigation purposes but also for sustaining long-term soil health (Johnston et al., 2009). Given that many soil functions scale with carbon content, increasing SOC trends can be used as indicators for improving soil quality (Don et al., 2024a). On the contrary, soils with decreasing trends are considered at risk for degradation, even if high initial SOC stocks have been observed. Advanced monitoring systems must therefore be able to assess both the status and changes of the SOC stocks in cropland soils to support climate mitigation and sustainable agriculture (De Rosa et al., 2024).

1.2 Political background

To reduce the adverse effects of climate change and limit global warming to 1.5 °C, the Paris Agreement was adopted in 2015 and signed by 194 countries worldwide (UNFCCC, 2015). As contributing parties, Germany and the European Union (EU) are required to submit national pathways to reduce the total GHG emissions and meet defined climate targets. Germany's Federal Climate Protection Act (Klimaschutzgesetz, KSG) aims to reduce national GHG emissions by 88% by 2040 (Bundesrepublik Deutschland, 2021). A key role is attributed to the Land Use, Land Use Change, and Forestry (LULUCF) sector, which is not only a source of emissions but also offers substantial potential for carbon sequestration in soils and biomass (Umweltbundesamt, 2023). According to the KSG, annual net CO_2 removals in the LULUCF sector are expected to increase to 40 million tons, making it a key component in achieving climate neutrality.

Recent estimations indicate that the GHG emissions in the LULUCF sector have increased significantly over the past years, reducing the net carbon sequestration and risking the KSG targets (Fuß, 2025; Gensior et al., 2025). While soil-related emissions are strongly driven by the degradation of peatland, it is estimated that carbon losses from organic and mineral soils contribute around 50 million tons of CO_2-equivalent to Germany's annual emissions (Gensior et al., 2025). Data from recent soil inventories shows that cropland soils in Germany are losing SOC at a rate of approximately $0.14\,\mathrm{t\,ha^{-1}\,y^{-1}}$, which sums up to over seven million tons of carbon each year (Fuß, 2025). As a result, while the emissions in Germany have been declining significantly since 1990, the contribution of cropland soils to the total GHGs is steadily increasing (Figure 1.1).

These developments emphasize the need for comprehensive monitoring to assess both the soil health as well as the SOC dynamics (Moinet et al., 2023). As carbon stock changes usually occur slowly, long-term programs are essential to track the impact of efforts to enhance SOC sequestration (Prechtel et al., 2009). The newly adopted EU Soil Monitoring Law (SML) therefore requires Member States to assess the soil health using a harmonized indicator catalogue (European Commission, 2023). Under the new regulation, repeated soil surveys must be conducted at least every five years to track progress toward the long-term goal of achieving healthy soil conditions by 2050.

Furthermore, international initiatives such as 4 per 1000 have gained attention for promoting carbon sequestration in agricultural soils as strategies for climate mitigation and soil restoration (Minasny et al., 2017). The initiative estimates that an annual increase of the global SOC stocks by 0.4% would be sufficient to offset current anthropogenic GHG emissions. However, various biophysical and socio-economic challenges have been identified to limit the implementation of this approach at larger scales (Rumpel, 2023).

One mechanism to incentivize the SOC sequestration in cropland soils is the adoption of nature-based solutions such as carbon farming, which promote the implementation of measures such as cover cropping, residue retention, or reduced tillage (Paul et al., 2023). Under the recently adopted Carbon Removals Certification Framework (CRCF), the EU aims to provide a standardized and transparent CO_2 certification system (European Union, 2024). Instead of rewarding the implementation of climate-smart management, the CRCF proposes a results-based

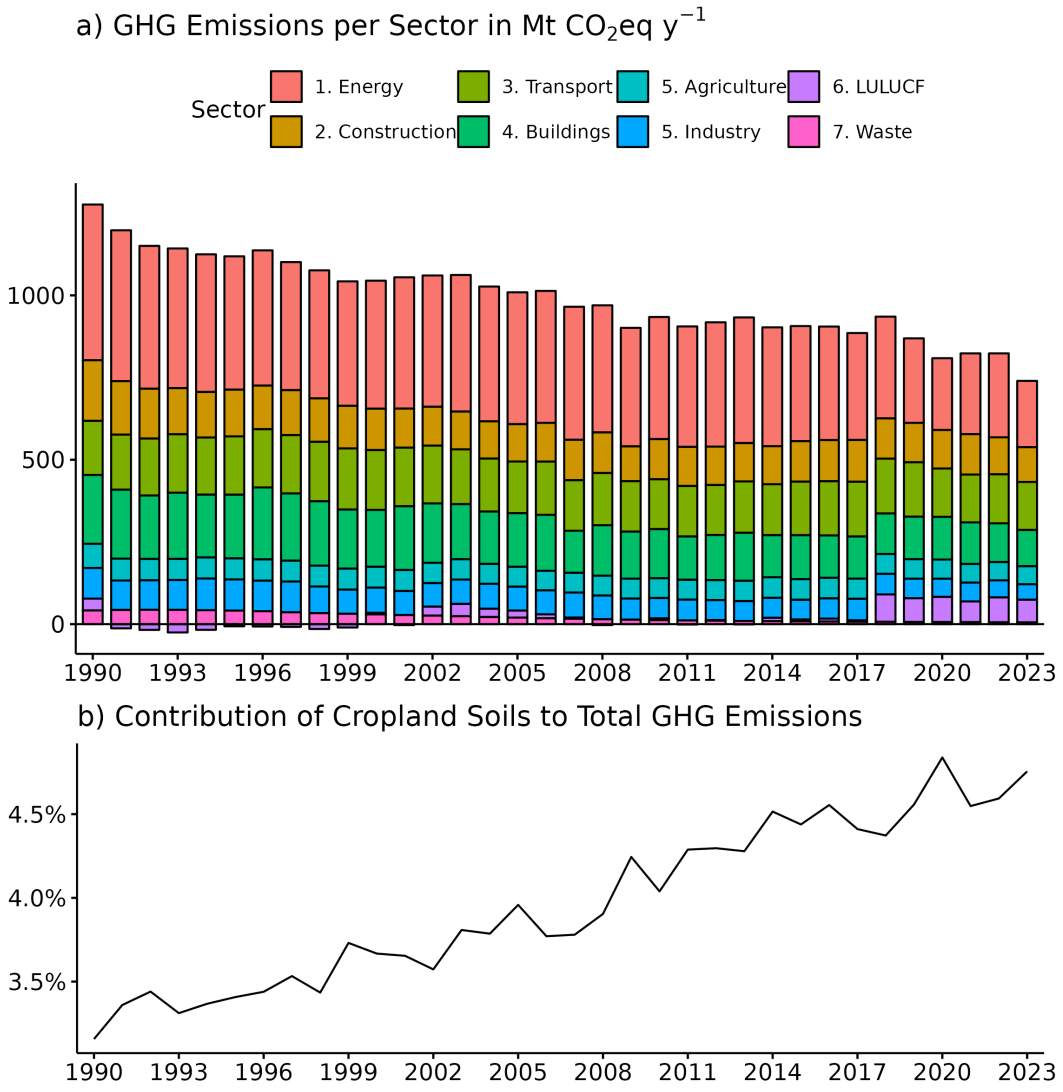

Figure 1.1: a) Total GHG emissions in Germany, colored by sector. b) Contribution of cropland soils to the total GHG emissions in Germany, derived from the sectors 5 (Agriculture; *Agricultural Soils*) and 6 (LULUCF; *Cropland*). Data from Umweltbundesamt (2025).

system that requires the verification of actual SOC stock changes. While such an approach improves the credibility of the carbon removals and prevents greenwashing, it also strongly increases the demand for soil monitoring data. As a result, the cost-effectiveness of the CRCF regulation could be significantly reduced if the verification costs to track SOC changes exceed the financial benefits that farmers can expect from the implementation of carbon farming.

To resolve the trade-off between robustness and cost of soil monitoring, the CRCF regulation and SML promote the potential of Earth Observation (EO) as a scalable approach to generate soil data and complement field-based samples. The growing availability of high-resolution satellite imagery from the Copernicus Sentinel missions provides new opportunities for EO-based soil maps (Tziolas et al., 2021). While spatiotemporal models offer a promising pathway the increase the cost-effectiveness of soil monitoring, they require the collection of soil reference data and the development of new methods to meet the requirements set by the emerging soil policies (Paustian et al., 2019; Smith et al., 2020).

1.3 Large-scale soil monitoring

In response to the increasing demand for high-quality soil data in space in time, large-scale sampling campaigns must be carried out and repeated regularly to detect long-term trends and evaluate the effectiveness of climate protection measures (Prechtel et al., 2009). One example is the Land Use and Coverage Area Frame Survey (LUCAS), which was launched in 2009 and currently includes over 250,000 sites across the EU (European Commission, Joint Research Centre, 2020). Every three to five years, approximately 20,000 soils are resampled and evaluated to assess changes in key soil properties. In Germany, the first Agricultural Soil Inventory (*Bodenzustandserhebung Landwirtschaft*; BZE-LW) was conducted between 2011 and 2018 and includes over 3000 cropland and grassland sites that have been sampled on an 8×8km grid across the country (Poeplau et al., 2020b). Even though the samples for the second BZE-LW are currently being collected and analyzed, the availability of high-quality data on long-term soil dynamics is low, especially before the start of the LUCAS campaign in 2009. A total of 387 long-term monitoring sites (LTM) have been established in Germany since 1985, which are continuously sampled at various intervals between 5 and 15 years (Umweltbundesamt, 2020). While the results from the LTM can provide some

information on the general soil trends within the past 30 to 40 years, the data density is currently too low to meet the demands of the CRCF regulation.

Large-scale field surveys such as LUCAS and BZE-LW are the foundation of soil monitoring systems, but their implementation can be cost-intensive and limited in the number of sampling locations. Only in very few cases can information on long-term soil dynamics be retrieved for past periods if no LTM sites have been established. To increase the spatial and temporal information density, methods of geostatistics have been widely adopted to map soil properties based on reference samples (Trangmar et al., 1986). Using spatiotemporal models based on EO data, it is possible to extrapolate the predictions to unknown regions and time frames (Heuvelink and Webster, 2001).

1.4 Digital soil mapping

Over recent decades, digital soil mapping (DSM) has become an established approach for generating spatially explicit information on soil properties across different scales (Behrens and Scholten, 2006a; Chen et al., 2022; Hengl et al., 2017; Poggio et al., 2021). With the growing availability of harmonized soil datasets, such as the LUCAS survey and the BZE-LW, and advances in machine learning and feature engineering, large-scale DSM products are increasingly applied in agricultural and policy contexts (Kidd et al., 2020).

The conventional DSM framework, which was popularized by Mcbratney et al. (2003), builds on the *SCORPAN* model, which describes soil formation as a function of seven environmental factors: soil (S), climate (C), organisms (O), relief (R), parent material (P), age (A), and spatial position (N). These factors are represented by covariates that capture the spatial variability of the landscape and are used to calibrate quantitative functions between point-based soil samples and the environmental predictors. To ensure model robustness and applicability across the research area, covariates should be selected based on their relevance to at least one of the *SCORPAN* factors (Wadoux et al., 2020b).

Commonly used predictors include terrain metrics derived from digital elevation models (DEM) (R), climatic parameters such as precipitation and temperature (C), vegetation indices and land use information (O), lithological maps or legacy soil data (P, S), and location-based variables (N) (Behrens et al., 2010;

Hengl et al., 2018). Remote sensing data are increasingly used as covariates in DSM models, as they provide information on key proxies such as aboveground vegetation, cropland management intensity, topsoil properties, and bare soil reflectance (Richer-de-Forges et al., 2023; Vaudour et al., 2022). Unlike other covariate groups derived from interpolated point data (e.g., climate parameters, legacy soil maps), satellite imagery offers high spatial resolution and information density, which can improve the robustness of the DSM models (Chabrillat et al., 2019).

Machine learning algorithms, such as *Random Forest* (Breiman, 2001) and *Support Vector Machines* (Cortes and Vapnik, 1995), are commonly used in DSM due to their capacity to model nonlinear relationships and handle high-dimensional feature spaces. Advanced methods, including *Artificial Neural Networks* and *Deep Learning*, have been demonstrated to improve the prediction performance in some cases, while also increasing overall model complexity (Behrens et al., 2005; Taghizadeh-Mehrjardi et al., 2022a). In general, DSM applications often adopt a results-oriented approach, focusing on prediction accuracy and relying on black-box models that limit the interpretability (Taghizadeh-Mehrjardi et al., 2021; Wadoux et al., 2020a). Consequently, many DSM models are trained on large sets of covariates, which may improve performance but reduce explainability while also increasing the risk of overfitting (Hengl et al., 2021; Nussbaum et al., 2018).

Statistical models, such as *Partial Least Squares Regression*, have been widely applied in remote sensing-based DSM (Angelopoulou et al., 2019). These approaches can offer greater model transferability, especially when calibrated with covariates that are strongly correlated with the target soil properties (e.g., soil reflectance) (Ward et al., 2020). However, they generally yield lower prediction accuracies compared to machine learning methods such as *Random Forest* or *Neural Networks* (Meng et al., 2022). Open questions remain regarding the optimal model configurations and environmental covariates that can improve map accuracy while maintaining interpretability and ensuring robust predictions both within (interpolation) and outside (extrapolation) the research area (Zhao et al., 2020).

1.5 Remote sensing-based soil mapping

Following the increasing availability of multi- and hyperspectral satellite images, methods to derive soil information based on remote sensing data have evolved significantly over the past decades (Chabrillat et al., 2019). Early studies have shown that the detection of aboveground biomass based on indices such as the *Normalized Difference Vegetation Index* (NDVI) can be linked to certain properties such as soil texture and water availability (Mulder et al., 2011). Other indices, such as the *Normalized Difference Tillage Index* (NDTI), can be used as proxies to detect cropland management information and support SOC monitoring (Van Deventer et al., 1997). Further methodological developments, such as *spectral unmixing*, have improved the capacity to characterize changes in soil cover over time (Bartholomeus et al., 2011).

While spectral indices can serve as proxies for certain soil conditions, it has been shown that the most accurate soil maps can be derived by observing the bare soil reflectance directly (Ben-Dor et al., 2008; Stevens et al., 2008). Laboratory and field-based soil spectroscopy have demonstrated that soil properties influence characteristic absorption features in the visible (VIS), near-infrared (NIR), and shortwave infrared (SWIR) regions of the electromagnetic spectrum (Ben-Dor et al., 2009; Stenberg et al., 2010). These interactions are specific to each property and are controlled by the chemical composition of the soil (Rossel and Behrens, 2010). For instance, the SOC content predominantly affects reflectance in the VIS–NIR range due to the absorption of CH, CO, and NH bonds, whereas clay minerals show distinct absorption features in the SWIR region that are associated with OH bonds (Ben-Dor et al., 2008) (Figure 1.2 a).

In contrast to laboratory-based soil spectroscopy, commonly used multispectral satellites such as Landsat or Sentinel-2 only provide a limited number of spectral bands (Figure 1.2 b). While hyperspectral satellites have a much higher spectral resolution, the quality of the soil signal is often affected by interfering factors such as surface roughness, vegetation cover, and soil moisture (Chabrillat et al., 2019). It has been shown that even small fractions of surface cover, such as biomass or crop residues, significantly influence the bare soil signal and reduce the accuracy of the resulting soil maps (Bartholomeus et al., 2007). In consequence, many studies either restrict their analyses to arid regions with minimal surface cover or apply spectral indices to identify fields with bare soil conditions

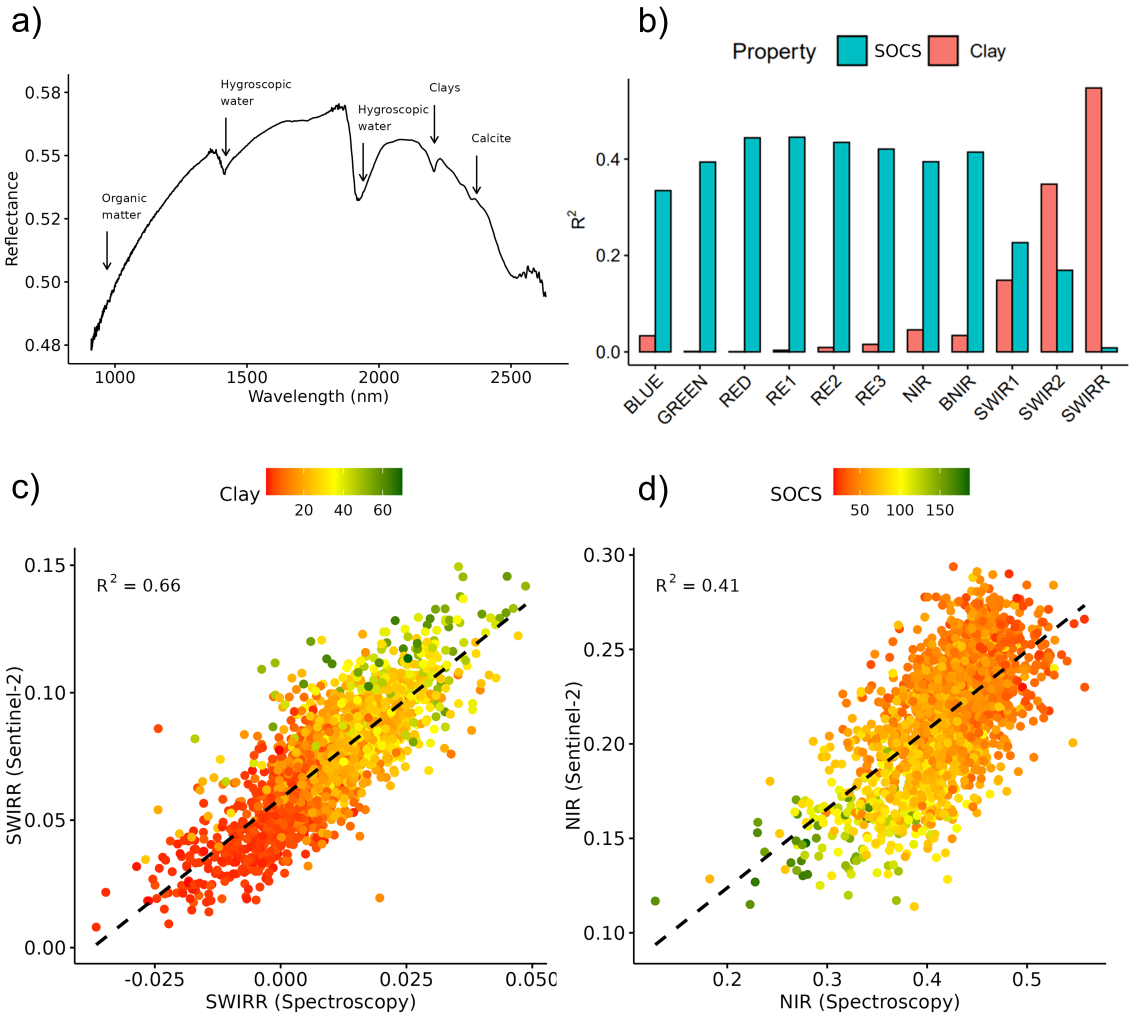

Figure 1.2: a) Example of a soil spectrum (Jaconi et al., 2019) with notations of the major absorption features (Ben-Dor et al., 2008). b) Band-wise correlations between the Sentinel-2 soil reflectance and the BZE samples (see *manuscript 2*); SWIRR = Ratio between SWIR1 and SWIR2. c+d) Relationship between Sentinel-2-derived soil reflectance (*manuscript 1*) and laboratory spectroscopy (Jaconi et al., 2019) of the BZE samples. Colors indicate measured topsoil properties; SOCS = soil organic carbon stock (Mg ha^{-1}).

(Chabrillat et al., 2002).

While such methods can improve model performance, they significantly reduce the overall mapping area, especially in temperate regions with intensive agriculture (Demattê et al., 2009). Due to factors such as cropping cycles, residue management, and precipitation patterns, exposed cropland soils are often spatially and temporally fragmented (Zepp et al., 2023). Consequently, DSM models calibrated on bare soil reflectance can only be applied to a limited fraction of agricultural land when based on single-date satellite observations. To address this limitation, two strategies have been proposed, which are based on the analyses of multitemporal time series data. Gerighausen et al. (2012) showed that using a sequence of hyperspectral observations from different years and seasons, multiple single-date predictions can be merged to generate a soil property mosaic. Although this approach can be used to increase the coverage of soil maps in intensively managed regions, it introduces challenges related to temporal variability in soil conditions, such as moisture content, surface roughness, and crop residue presence. As an alternative, multitemporal bare soil observations can be aggregated to generate reflectance composites, improving both the spatial extent as well as the spectral quality of the soil signal (Diek et al., 2016).

1.6 Soil reflectance composites

Early remote sensing studies focused on transferring laboratory-based soil spectroscopy to hyperspectral images, which offer higher spectral resolution than conventional multispectral sensors (Lagacherie et al., 2008). Although additional spectral bands can improve the prediction of soil properties, the low availability and temporal resolution of hyperspectral satellite data currently limit the applicability for large-scale soil mapping at national or continental scales (Ward et al., 2020; Zhao et al., 2025). In response, several studies have suggested that the lower spectral resolution of multispectral imagery can be compensated for by computing band-wise spectral averages over multiple bare soil observations to reduce the influence of interfering factors such as vegetation cover, soil moisture, and surface roughness (Demattê et al., 2016; Diek et al., 2016; Rogge et al., 2018). Analyzing more than three decades of Landsat imagery, Diek et al. (2017) demonstrated that approximately five years are required to generate complete soil reflectance composites (SRC) at the national scale, which support the prediction

of large-scale soil maps (Figure 1.3). Since then, further studies reported that multitemporal SRC consistently outperform models based on single-date observations in terms of prediction accuracy and map extent (Shi et al., 2022; Silvero et al., 2021). Demattê et al. (2018) not only showed that Landsat-based SRCs correlate with measurements from laboratory soil spectroscopy (see Figure 1.2 c&d), but also emphasized that the selection of bare soil indices and thresholds affects the purity of the resulting spectral soil signal.

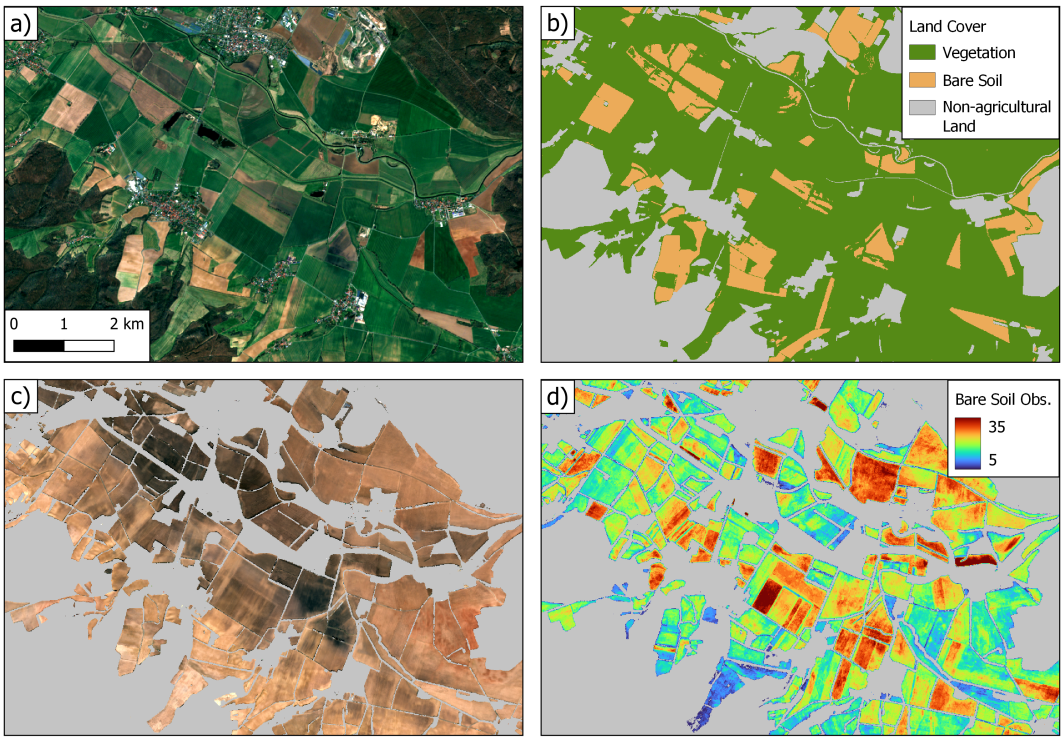

Figure 1.3: a) Single-date Sentinel-2 scene (RGB). b) Detection of bare soil, based on the methods described in *manuscript 2*. c) Bare soil composite (RGB), based on a Sentinel-2 time series from 2015 to 2024. d) Number of bare soil observations, used for the generation of the composite.

Different spectral indices, including the *Normalized Burn Ratio 2* (NBR2), *Bare Soil Index* (BSI), and vegetation indices such as NDVI, have been applied to identify bare soil observations for the generation of SRC (Delaney et al., 2025). Different studies have argued that by combining multiple indices and thresholding values, the quality of the resulting soil reflectance can be maximized (Castaldi et al., 2019a; Demattê et al., 2018). For example, while the NDVI is strongly associated with green vegetation cover, indices such as NBR2 are more sensitive

to crop residues and soil moisture (Dvorakova et al., 2022).

In response, numerous SRC frameworks have been developed to predict soil properties at various spatial scales and validated in terms of spectral quality and predictive power for DSM (Vaudour et al., 2022). However, open questions remain regarding the optimal parameterization for generating accurate and unbiased soil maps at large scales that are compatible with the requirements of current soil monitoring and policy targets.

1.7 Spatiotemporal models

While conventional DSM has been successfully used to generate static soil maps, the methods have since been adapted to predict changes in soil properties using spatiotemporal modeling approaches (Helfenstein et al., 2024; Heuvelink et al., 2021). Dynamic models integrate multitemporal soil observations with covariates such as vegetation indices or climate data to produce spatiotemporal overlays that enable mapping across both space and time (Heuvelink and Webster, 2001). Remote sensing data is frequently used as the primary input for these models due to the availability of long-term satellite time series, high spatial resolution, and strong predictive power (Meng et al., 2024; Tian et al., 2024; Venter et al., 2021). Consequently, several studies have produced dynamic maps that show large-scale trends in soil properties such as SOC (Hengl et al., 2025; Van Wesemael et al., 2024). While these models are often calibrated using extensive soil datasets at continental or global scales, they rarely include repeated measurements at the same locations, limiting the direct validation of the predicted SOC changes (Tian et al., 2024).

The lack of temporal validation has been identified as one of the major challenges for the operational use of dynamic soil maps in policy contexts, including the CRCF regulation (Wageningen Environmental Research, 2024). Although previous studies have noted that SOC change detection is constrained by factors such as low signal-to-noise ratios in satellite-derived soil maps and the limited temporal variability of cropland SOC, there is a lack of empirical validation to support these claims (Croft et al., 2012; Heuvelink et al., 2021). Despite the growing body of literature on the use of remote sensing for dynamic soil mapping, additional research is necessary to further assess the potential and challenges regarding the usage of spatiotemporal models for large-scale soil monitoring.

Chapter 2

Scope of this Dissertation

Motivated by the methodological and practical challenges outlined in Chapter 1, this dissertation addresses the pressing question: *To what extent can remote sensing data be used to monitor changes in cropland SOC?* Framed by this overarching issue, five main objectives are defined, analyzing the potential of multispectral imagery and bare soil composites to generate static and dynamic soil maps for large-scale SOC monitoring (Figure 2.1):

1. **Generation of multispectral soil reflectance composites**
    - RQ1: How does the selection of the satellite input data affect the generation of soil reflectance composites for soil mapping?
    - RQ2: What is the influence of the bare soil indices and thresholds on the resulting soil reflectance composites?
    - RQ3: Is there a trade-off between the extent and quality of the soil reflectance composites?
2. **Integration of soil reflectance composites for conventional digital soil mapping**
    - RQ4: What correlations can be observed between the soil reflectance and different topsoil properties?
    - RQ5: How does the predictive power of the soil reflectance compare to other covariate groups?
    - RQ6: How are the relationships between the soil reflectance and training data affected by the landscape conditions?
3. **Development of enhanced digital soil mapping frameworks**
    - RQ7: To what extent can local ensemble models be used to improve the accuracy and explainability of large-scale soil maps?
    - RQ8: How does the accuracy and distribution of the predicted soil properties depend on the balancing of the training data?
4. **Integration of bare soil time series for spatiotemporal soil mapping**
    - RQ9: What are the temporal consistency and signal-to-noise ratio of long-term SOC predictions based on soil reflectance composites?
    - RQ10: Is it possible to validate predicted SOC trends and changes

based on long-term soil monitoring data?

5. **Usage of bare soil time series to derive further agricultural activity data**
    - RQ11: How can cropland and grassland management data be derived from satellite time series to support soil monitoring?

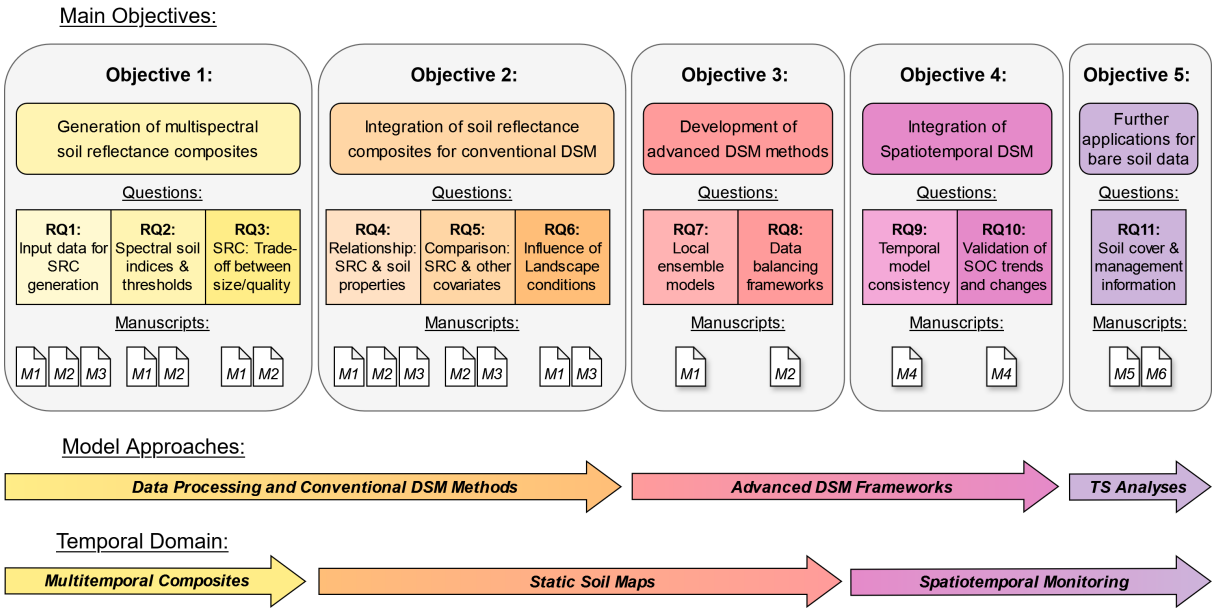

Figure 2.1: Overview of the five main objectives, the corresponding research questions, and the included manuscripts.

First, *Objective 1* addresses the methodological foundations for generating high-quality bare soil maps from multispectral satellite data. Next, the use of SRC as primary covariates for static DSM is evaluated in *Objective 2*, focusing on predictive performance, model transferability, and the relationship between spectral signal and soil properties. Building on this, *Objective 3* explores advanced modeling approaches, such as the use of local ensemble models and data balancing frameworks, to improve the accuracy and explainability of large-scale soil maps. *Objective 4* evaluates the potential of spatiotemporal modeling to detect and validate SOC changes using long-term EO data and repeated field measurements, addressing one of the central challenges for the operational use of EO-based soil monitoring in policy frameworks. Finally, *Objective 5* examines further applications of bare soil time series for deriving agricultural activity data to complement integrated SOC monitoring systems.

In Chapter 3, each of these objectives is addressed through a series of research

questions, which are explored across the six manuscripts included at the end of this dissertation (Chapters 7 to 12). Each subsection in Chapter 3 concludes with a summary of the key implications derived from the respective research questions. Chapter 4 presents a synthesis that responds to the overarching question of this dissertation: *To what extent can remote sensing data be used to monitor changes in cropland SOC?* Chapter 5 discusses current limitations and methodological challenges, while Chapter 6 provides a summary of the policy and practical considerations emerging from the findings of this work.

Chapter 3

Synthesis of the Findings

3.1 Generation of multispectral soil reflectance composites

Over the past decade, an increasing number of studies have focused on the generation of multispectral SRC for soil mapping applications (Pouladi et al., 2023). While threshold-based approaches remain the most popular method for bare soil identification, various strategies have been explored to improve the spectral soil signal and predictive power of SRC-based soil models. A key focus has been set on optimizing the selection of spectral indices and corresponding thresholding values, used to increase the SRC quality by identifying the most suitable bare soil observations.

In the *manuscripts 1* to *4*, different frameworks were evaluated for the generation of SRC in the context of static and dynamic SOC mapping. Based on the findings, three central aspects were identified that influence the accuracy, spatial extent, as well as bias of the resulting soil maps: (1) The satellite data used as input for the SRC generation; (2) the parameterization of the SRC, including the selection of bare soil indices and corresponding thresholds; and (3) the trade-off between spectral quality and spatial coverage.

RQ1: How does the selection of the satellite input data affect the generation of soil reflectance composites for soil mapping?

Different multispectral input data were tested for the SRC generation, including Sentinel-2 imagery (*manuscript 2*), Landsat data (*manuscripts 1 & 3*), and a combined time series of all available Sentinel-2 and Landsat observations (*manuscript 4*). In general, all evaluated configurations resulted in similar model accuracies for the prediction of SOC in Germany (Broeg et al., 2026; Broeg et al., 2024a) and Bavaria (Broeg et al., 2024b, 2023), with observed R^2 values

ranging between 0.61 and 0.68. These findings are in line with existing literature, suggesting that both Sentinel-2 and Landsat images are suitable for the SRC generation; however, they offer distinct benefits that can depend on the overall aim of the studies (Castaldi, 2021; Silvero et al., 2021).

In comparison to Landsat, Sentinel-2 images offer a higher spatial resolution and improve the analysis of the field-scale soil variability. The positive influence of the enhanced spectral resolutions, including four additional bands in the red-edge (698 to 793 nm) and NIR (855 to 875 nm) region, significantly depends on the spectral characteristics of the soil properties. For the SOC models in *manuscript 2*, it was found that the explanatory power of the red-edge bands is slightly higher than the red band, influencing the overall prediction accuracy (Broeg et al., 2026). In contrast, the clay models strongly depend on the bands in the SWIR region and are only weakly affected by the enhanced spectral resolution.

For large-scale applications, Landsat images with slightly lower spatial resolution can be used to reduce the computational demand for SRC generation, while still providing good results for the prediction of SOC (Broeg et al., 2024a). In addition, the Landsat archive provides a significantly longer temporal record, supporting the analyses of long-term soil dynamics. As demonstrated with the signal-to-noise ratio in *manuscript 4*, the temporal availability of Sentinel-2 is not sufficient to predict significant SOC trends under typical cropland conditions (Broeg et al., 2024b). For this reason, a combined time series of Sentinel-2 and Landsat imagery was used to increase the temporal observation density, enabling the analysis of the SOC dynamics for almost four decades (1986 to 2022).

To harmonize these datasets, it is crucial to conduct spectral adjustments to generate SRC that are consistent across time and different sensor generations (Okujeni et al., 2024). The adjustments, as well as the necessary preprocessing of the satellite data (radiometric correction, cloud masking, etc.), were conducted using the open-source software *Framework for Operational Radiometric Correction for Environmental monitoring* (FORCE), offering transparent and reproducible workflows that have proven to be optimal for SRC generation (Frantz, 2019).

Key implications of RQ1:

- Both Landsat and Sentinel-2 data are suitable for SRC generation, and the selection of input data should be guided by the spatial and temporal

resolution, scale, and objectives of the DSM applications.

- To increase the temporal density of bare soil observations, both data sources should be combined when deriving long-term soil dynamics.

RQ2: What is the influence of the bare soil indices and thresholds on the resulting soil reflectance composites?

In the *manuscript 1* and *2*, different combinations of bare soil indices and thresholding values were tested for SRC generation and evaluated in terms of the extent and accuracy of the resulting soil maps. In general, the SRC parameters must be optimized to (1) determine which pixels are included in the maps (i.e., bare soil mask) and (2) which bare soil observations should be considered to calculate the mean soil reflectance (i.e., spectral quality).

While both of the objectives are usually addressed using a single set of parameters, the results in *manuscript 2* demonstrated that the ability of different spectral indices to generate unbiased soil masks and to select optimal observations can vary significantly. The NDVI and NBR2, which are often used in SRC frameworks, are sensitive to factors interfering with the soil reflectance (vegetation, soil moisture, crop residues, etc.) and can be used to improve the overall quality of the SRC. As both of these indices are also affected by the soil reflectance itself, however, they are less suitable for generating unbiased bare soil masks and can lead to the systematic exclusion of certain soil properties when implemented incorrectly (Broeg et al., 2026).

To address these issues, three SRC frameworks were evaluated in the *manuscripts 1 & 2*: (1) Using high thresholds (e.g., NDVI < 0.35 & NBR2 < 0.16), it is generally possible to minimize the exclusion of valid bare soil pixels, increasing the overall extent of the SRC. While this approach reduces the bias of the resulting soil maps, it may also decrease the spectral quality of the SRC by including soil observations that are influenced by vegetation, crop residues, and other interfering factors. (2) Alternative soil indices, such as the PV+IR2 (Heiden et al., 2022), can be used to improve the SRC quality while minimizing the systematic exclusion of soil properties. The results in *manuscript 2* demonstrated that the PV+IR2 is less sensitive to soil reflectance than the NDVI and can be used to implement low SRC thresholds (e.g., 0.24), increasing the spectral quality and accuracy of the resulting soil models. (3) Lastly,

dynamic thresholds can be used to separate the bare soil detection and masking to improve the quality of the soil signal without systematically excluding specific soil properties. In *manuscript 1*, high SRC thresholds were first applied to define the bare soil mask, while in the second step, only the observations with the lowest NBR2 values per pixel were selected for calculating the soil reflectance (Broeg et al., 2024a).

Key implications of RQ2:

- Spectral indices and thresholds for SRC generation must be optimized regarding the bare soil mask and quality of the soil signal.
- Disregarding the interactions between the spectral indices and soil properties can lead to the generation of biased SRC and limit the accuracy and extent of the resulting maps.

RQ3: Is there a trade-off between the extent and quality of the soil reflectance composites?

Multiple studies have reported a trade-off between the spatial extent and spectral quality of SRC, directly influencing the prediction accuracy of the resulting soil models (Castaldi et al., 2019b; Dvorakova et al., 2022; Vaudour et al., 2021). In most cases, the trade-off is described as the choice between applying low thresholds to optimize the selection of bare soil observations or using higher thresholds to maximize the SRC coverage. In this context, it has been shown that low NBR2 values (e.g., 0.08) can reduce interfering factors in the soil signal (soil moisture, vegetation, crop residues, etc.) and significantly improve SOC predictions, while also strongly limiting the number of available soil samples for the models. The findings in *manuscript 2* demonstrated, however, that this observed trade-off is not only a function of spectral quality and map extent but is strongly influenced by the systematic exclusion of specific soil properties during SRC generation (Broeg et al., 2026). It was observed that correlations between bare soil indices and soil reflectance can alter the distributions of the included soil properties, especially when implementing low NBR2 or NDVI thresholds. While this process can potentially improve the observed model accuracy, e.g., by excluding samples with high clay contents that interfere with the SOC prediction (Stenberg, 2010), it also strongly increases the bias in the resulting soil maps. These findings highlight that any observed trade-offs involving the SRC coverage must be carefully

examined to determine whether they result from improvements in the soil signal or the selective exclusion of specific soil conditions. In *manuscript 2*, it was demonstrated that one way of achieving this is to compare the distribution of the soil properties before and after SRC generation, ensuring that no bias has been introduced by the soil masks.

Key implications of RQ3:

- Trade-offs between SRC extent and quality can exist, but are only valid if they do not arise from the systematic exclusion of certain soil properties from the soil maps.
- For reliable and representative SOC mapping, SRC parameters must be optimized not only for spectral quality but also to ensure comprehensive coverage of soil variability.

3.2 Integration of soil reflectance composites for conventional digital soil mapping

In *manuscripts 1*, *2*, and *3*, it was evaluated how remote sensing-based SRC can enhance the performance of static DSM. Across all studies, the results demonstrate that the SRC bands have high predictive power and are strongly correlated to the observed soil properties, supporting their suitability to improve the accuracy, transferability, and explainability of DSM. While SRCs have been successfully used to improve the results of static DSM before (Stumpf et al., 2024; Žížala et al., 2022), these findings provide further insights about (1) model variances across different soil properties; (2) dependencies between the landscape conditions and soil reflectance; and (3) comparison between the SRC and other covariates.

RQ4: What correlations can be observed between the soil reflectance and different topsoil properties?

Despite the limited spectral resolution of multispectral satellite data, the observed relationships between the SRC and soil properties align with findings from laboratory soil spectroscopy, increasing the robustness of the results (Ben-Dor et al., 2008). As highlighted in *manuscript 2*, the spectral response of SOC is

primarily observed across the VIS–NIR spectrum, with carbon-rich soils showing lower reflectance values across all Sentinel-2 bands. The strongest correlations between the SRC and SOC were found in the red-edge region (705 nm), with R^2 values above 0.4, while significantly weaker relationships were observed in the SWIR bands (Broeg et al., 2026). These findings are in line with soil spectroscopy (Stenberg et al., 2010), confirming that the soil signal is retained in the vis-NIR signal at the satellite scale and can be exploited for large-scale SOC mapping.

In contrast, clay minerals generally show pronounced absorption features in the SWIR region around 2200 nm, which is captured by both the Sentinel-2 and Landsat sensors. Among the observed spectral bands and indices, the strongest correlations between the clay content and SRC were found in the NBR2 ($R^2 > 0.5$), which combines the SWIR1 and SWIR2 bands (Broeg et al., 2026). In general, these results are supported by other studies, confirming that the SWIR region is essential for the prediction of the clay content and further soil properties that are related to the soil texture and bulk density (Silvero et al., 2021; Wetterlind et al., 2025).

In comparison to other covariates used in conventional DSM (DEM derivatives, climate parameters, etc.), the predictive power of SRC can be classified as high; however, it strongly depends on the spectral response of the observed soil properties. While SOC and clay content are known to strongly affect the soil reflectance, other parameters like the silt content have a weaker spectral signal and are more difficult to detect (Sørensen and Dalsgaard, 2005). Overall, SRCs not only contribute to improving model performance but also offer spatial information on the physical relationships between soil properties and reflectance, enabling robust and transferable DSM applications across varying cropland conditions.

Key implications of RQ4:

- Satellite-derived soil reflectance offers high predictive power for the generation of large-scale soil maps.
- The observed correlations between the SRC bands and different soil properties are supported by the results of laboratory soil spectroscopy, increasing the robustness of the DSM models.

RQ5: How does the predictive power of the soil reflectance compare to other covariate groups?

The results of the manuscripts demonstrate that remote sensing data serves as an independent predictor in SOC models and can be used to produce accurate maps without requiring additional model features. As shown in *manuscript 3*, SRCs offer advantages in terms of prediction accuracy, spatial resolution, and model transferability when compared to other covariates used in conventional DSM (DEM derivatives, climate parameters, etc.). Instead of relying on indirect proxies to represent the soil formation process (i.e., the SCORPAN factors), the surface conditions can be analyzed directly to calibrate robust models based on the physical relationships between soil properties and reflectance (Ben-Dor et al., 2008). A main benefit of this methodology is the possibility to calibrate DSM models using soil samples from one region and generalize the observed correlation between the SRC and soil properties beyond the original research area. As shown in *manuscript 3*, only models trained using the soil reflectance maintained their predictive performance when applied to regions outside of the training data, whereas models based on conventional covariates show no signs of transferability (Broeg et al., 2023). Exploiting robust relationships between reflectance and corresponding soil properties, it is therefore possible to significantly increase the area of applicability beyond the training data, addressing a common limitation of DSM (Meyer and Pebesma, 2021).

While models based on SRC alone can yield strong results, it was also shown that the prediction accuracy can be further improved by integrating complementary covariate sources (Broeg et al., 2026, 2023). These include additional remote sensing-derived features such as spectral-temporal metrics (STM) or spectral indices, which can enrich the SRC without compromising model resolution. For large-scale models, the soil reflectance can be combined with further topographic covariates to provide spatial context on the landscape and environmental gradients within the research area. This can be especially beneficial in regions with heterogeneous soils, which may not be fully captured by the soil reflectance.

Key implications of RQ5:

- Based on the spectral relationships between the SRC and soil properties, robust DSM models can be calibrated that are transferable to regions out-

side the training data.

- While SRC-based SOC models offer high accuracy, the performance can be further improved by incorporating complementary covariates that capture the landscape heterogeneity.

RQ6: How are the relationships between the soil reflectance and training data affected by the landscape conditions?

In addition to the global relationships that have been observed for the prediction of SOC, the findings in *manuscript 1* show that the local soil conditions can significantly influence the correlations between the soil properties and reflectance. In general, these spatial dependencies highlight the need to address landscape-specific factors in SRC-based soil models and can be attributed to three main aspects: (1) The interactions across different soil properties, such as the correlation between SOC and clay content; (2) the presence of interfering factors, such as increased levels of soil moisture; and (3) the availability and spatial distribution of the soil samples used for model training.

As demonstrated in *manuscript 1*, SOC predictions across Germany are not only driven by the carbon signal in the SRC but also influenced by interactions with soil texture. In general, the clay content is positively related to carbon sequestration due to its stabilizing property and can, therefore, serve as a proxy for SOC predictions. The findings show that this co-dependency can alter the feature importance of the SRC bands depending on the dominant soil texture within the research area. For example, in clay-rich regions, the contribution of SWIR bands for the prediction of SOC is more pronounced than in other areas dominated by sandy soils (Broeg et al., 2024a).

Another factor that can limit the predictive power of SRC by reducing the sensitivity of the soil signal is the presence of soil moisture (Stenberg, 2010). Although lowering the thresholds for SRC generation can reduce this interference, some regions with high groundwater levels often contain soils that are saturated over long periods. As shown in *manuscripts 1 & 2*, these soil conditions can lead to model confusion and significantly increase the uncertainty for the prediction of SOC and clay. These results highlight the importance of integrating landscape information into SRC-based models, especially in large or heterogeneous research

areas (Broeg et al., 2024a). In general, the findings in *manuscript 1* have demonstrated that the calibration DSM model based on local soil samples can help identify overlapping spectral signals and improve the SOC prediction accuracy by accounting for the spatial variability of the soil conditions.

Key implications of RQ6:

- Dependencies between different soil properties can influence the feature importance of the SRC bands for soil mapping applications.
- Pedological conditions, such as the soil moisture or texture, can interfere with the prediction of soil properties and increase the model uncertainty, especially in regions with heterogeneous landscapes.

3.3 Development of enhanced digital soil mapping frameworks

While conventional DSM offers a robust framework for generating spatial information on cropland soils, several methodological limitations remain regarding the prediction of large-scale maps, including (1) the integration of information on the spatial position of the soil samples; (2) model adjustments to regional dependencies in the relationships between the SRC and soil properties; and (3) difficulties accounting for samples and extreme values that are underrepresented in the training data.

To address these issues, two novel frameworks were developed to adopt the existing DSM methods based on SRC as the main model input. In *manuscript 1*, it was tested whether the implementation of local ensemble models (LEM) can be used to integrate spatial context and address the regional soil variability, while *manuscript 2* explored a data balancing approach to improve the representativeness of the soil samples in the training data.

RQ7: To what extent can local ensemble models be used to improve the accuracy and explainability of large-scale soil maps?

The use of distance covariates, such as geographical coordinates, has become a common strategy for the integration of spatial information into DSM due to

its simplicity and positive influence on model accuracy (Behrens et al., 2018b). However, it has been argued that because of the lack of pedological relevance, these covariates can lead to artifacts in the resulting maps, limiting their usability for soil mapping (Meyer et al., 2019; Wadoux et al., 2020a). Instead, the results of *manuscript 1* demonstrate that an ensemble of locally calibrated models can effectively account for regional dependencies in the research area, without the adverse effects that have been attributed to the integration of distance covariates.

By breaking down a national-scale model into smaller, spatially defined sub-models, this approach improves the representation of diverse soil conditions by adjusting the relationship between SRC and soil properties to the local samples. In addition, the proposed framework reduces the impact of sample imbalance by accounting for soil variations that can be underrepresented in the training data when calibrating a global model (Broeg et al., 2024a).

The evaluation of the model performance shows that using this enhanced LEM approach, the spatial autocorrelation can be significantly reduced when compared to a conventional DSM model that has been calibrated based on distance covariates. Analyzing the model uncertainty and feature importance for the local sub-models reveals distinct patterns that can be explained by varying soil conditions within the research area. These findings demonstrate the value of adaptive modeling strategies based on SRC, such as the LEM, especially when applied to heterogeneous landscapes (Broeg et al., 2024a).

Key implications of RQ7:

- The proposed LEM framework was successfully used to integrate spatial information and account for regional dependencies in the correlations between the SRC and the soil properties.
- Regionalized analyses of the feature importance and model uncertainty can be implemented to provide further information about the soil heterogeneity within the research area.

RQ8: How does the accuracy and distribution of the predicted soil properties depend on the balancing of the training data?

In addition to the LEM, which relies on a two-step modeling process, a simplified data balancing framework was tested in *manuscript 2* to address the challenge of underrepresented soil conditions in the training data. It has been shown that for SOC prediction, DSM models tend to be biased toward the mean and produce significantly lower accuracies for extreme values, especially when the sample distributions are skewed (Feeney et al., 2022). As a result, the full range of the observed SOC variability is often not captured in the predictions, limiting the reliability of the resulting soil maps. To address this challenge, a data balancing framework was adopted, combining spatial augmentation and undersampling to increase the representation of the soil conditions within the research area without compromising on sample size (Broeg et al., 2026).

The findings show that the proposed approach significantly improved the range of the predicted SOCS and clay values. However, it strongly depended on identifying the optimal trade-off between the size and balance of the training data. Using strong balancing to adjust the distributions of the training data, the models showed tendencies to produce overpredictions when applied to the original sample distributions. Instead, the strongest model improvements were observed when implementing intermediate balancing strategies and were most significant in the Concordance Correlation Coefficient (CCC), while the RMSE remained similar to the results when using the unbalanced data. These findings suggest that when aiming to generate robust soil maps, the fit of the model distributions (i.e., the CCC) should be evaluated separately from the overall prediction uncertainty (i.e., the RMSE) to identify the optimal model configurations (Broeg et al., 2026).

While the observed trade-offs were strongly dependent on the imbalance of the soil properties, the results in *manuscript 2* also highlight that it remains challenging to reproduce the full range of the soil samples, even when implementing data balancing frameworks. This can be explained by the fact that for some properties, such as the SOC, the model uncertainty is significantly correlated to the standard deviation of the input samples, resulting in difficulties when predicting across broad environmental gradients (Feeney et al., 2022; Stenberg et al., 2010). Further research is necessary to overcome such challenges regarding the

SOC signal, which could be addressed by improvements in the SRC framework, additional spectral features, or the implementation of hyperspectral satellite data (Ward et al., 2020; Zhao et al., 2025).

Key implications of RQ8:

- Data balancing can be used to improve the range of the model output; however, it strongly depends on the specific soil properties and identifying the optimal trade-off between over- and underpredictions.
- To generate robust soil maps, the prediction range should be assessed separately from the uncertainty to ensure that the distribution of the model output reflects the reference samples.

3.4 Integration of bare soil time series for spatiotemporal soil mapping

The integration of remote sensing data into the DSM framework not only improves the prediction of static soil maps, as demonstrated in *manuscripts 1*, *2* & *3*, but also enables the monitoring of the temporal soil dynamics. The results of *manuscript 4* show that by using a spatiotemporal overlay to link soil samples with temporally corresponding satellite data, it is possible to monitor SOC changes in cropland soils over time. Although SRCs have been successfully applied to a wide array of soil mapping applications, their usage in dynamic SOC models remains limited, due to three main constraints (Croft et al., 2012; Heuvelink et al., 2021; Tian et al., 2024): (1) The low availability of repeated SOC samples that can be used for model training and validation; (2) the slow-changing nature of SOC and the requirement to analyze long-term time series to detect significant trends; and (3) the limited number of usable bare soil observations, especially in the pre-Sentinel-2 era before 2015. *Manuscript 4* evaluated the potential of deriving SOC dynamics using an SRC time series and soil monitoring data from 1986 to 2022. For the first time, it was possible not only to predict long-term SOC trends in cropland soils but also to validate the predictions using measured changes based on repeated soil samples (Broeg et al., 2024b).

RQ9: What are the temporal consistency and signal-to-noise ratio of long-term SOC predictions based on soil reflectance composites?

A major limitation regarding the implementation of dynamic DSM based on SRC is the limited availability of high-quality satellite observations in the early years of the observation period, especially before the launch of Sentinel-2 in 2015 and Landsat 7 in 1999. To address this issue, it is necessary to aggregate the reflectance data across multiple years to increase the number of bare soil observations and ensure sufficient coverage for SRC generation.

One strategy, which has been tested in *manuscript 4*, is to apply a temporal moving window approach to average the soil reflectance within defined time intervals. This approach can be useful when repeated soil samples are collected at irregular intervals; however, it also introduces a trade-off between the spectral quality of the SRC and the ability to detect changes in SOC. While broader intervals increase the number of available bare soil observations and can have a positive influence on prediction accuracy, they also reduce the temporal resolution of the resulting model. This trade-off is especially relevant in cropland soils, where SOC changes usually occur slowly and must be aligned with the objectives of the soil maps. In *manuscript 4*, a relatively long temporal window of ±4 years was used to address the low satellite data availability; however, the modeled time frame (36 years) was long enough to offset the reduced temporal resolution. The results demonstrate that model accuracy and prediction bias remained mostly consistent across the observation period, which is essential for detecting significant SOC trends (Broeg et al., 2024b).

To further assess how the temporal model uncertainty relates to the measured SOC trends over time, the signal-to-noise ratio was quantified for each sampling year. The findings show that while the overall prediction accuracy remained constant across the observation period, the measured SOC changes between individual sampling years remained relatively small and only exceeded the model uncertainty after a time frame of approximately 30 years (Broeg et al., 2024b). This highlights a key limitation for spatiotemporal SOC monitoring: in typical cropland soils, SOC changes occur slowly, and short-term trends are likely to remain within the range of the prediction uncertainty. As demonstrated in *manuscript 4*, the detection of significant SOC trends requires long-term ob-

servation periods and models that produce temporally consistent results.

Key implications of RQ9:

- Time series of SRC can be used to derive long-term SOC predictions, given that the moving window intervals are adjusted to both data availability and the temporal resolution of the modeling objectives.
- The signal-to-noise ratio can be used to derive the temporal consistency of the model accuracy and estimate the necessary time frame to detect SOC changes in cropland soils.

RQ10: Is it possible to validate predicted SOC trends and changes based on long-term soil monitoring data?

To evaluate the accuracy of the predicted SOC trends, two independent datasets were analyzed in *manuscript 4*, including cropland soils that were resampled once after 10 years and long-term monitoring sites that were revisited every five years from 1986 to 2022. Although the overall prediction accuracy of the SOC model was high, the validation based on the measured short-term SOC changes yielded non-significant results (Broeg et al., 2024b).

Based on these findings, several conclusions can be drawn that contribute to the observed limitations. For most sites, the measured SOC changes after 10 years were relatively small (mostly within ± 2 g kg^{-1}) and well below the observed model uncertainty (RMSE >4 g kg^{-1}). As discussed in the previous section, the resulting signal-to-noise ratio was therefore most likely too low to support the prediction of significant short-term SOC changes (Broeg et al., 2024b). While the use of a moving window for SRC generation may blur the temporal precision of the predictions, both the measured and predicted SOC changes shared similar distributions. This suggests that the temporal model resolution was not the main limiting factor and that the temporal variability captured by the SOC model was similar to the repeated soil samples. Additional factors that can further reduce the reliability of the validation are sampling uncertainties associated with the small-scale spatial heterogeneity and the non-linear character of SOC changes over time (Poeplau and Gregorich, 2023). Because of these aspects, it can be challenging to attribute measured differences between two sampling events to actual SOC trends rather than sampling noise, highlighting the need for soil monitoring data that has been resampled multiple times.

In contrast, the direct validation of long-term SOC trends from 1986 to 2022 yielded statistically significant results, although the accuracy of the predictions remained low. Further analyses revealed, however, that results are strongly influenced by a series of soil and management factors (Broeg et al., 2024b). In general, trends were predicted more accurately at sites with high initial SOC contents. These findings are consistent with the signal-to-noise ratio, showing that high absolute SOC changes increase the likelihood of detecting significant trends over time. For the remaining sites, the magnitude of the measured changes was often smaller than the prediction uncertainty, limiting the detection of accurate SOC trends. Sites that were managed under reduced tillage frequently showed positive SOC trends, which were also reflected in the predictions (Broeg et al., 2024b). In general, studies have shown that reduced soil disturbance can induce the redistribution of SOC in the topsoil (Haddaway et al., 2017), which appears to have influenced both the SRC signal and the model output.

Overall, the findings in *manuscript 4* indicate that the prediction of relative SOC trends is more reliable than the estimation of absolute change rates over time. This fact is often overlooked in discussions regarding the signal-to-noise ratio, arguing that the prediction uncertainty must be lower than the actual SOC changes to derive significant trends (Van der Voort et al., 2023). By categorizing the SOC trends into three main groups (i.e., increasing, decreasing, and stable), it was observed that only 5% of the sites showed predictions that were opposite to the measured trends, highlighting that a consistent SOC signal was present in the soil reflectance (Broeg et al., 2024b). These findings demonstrate the potential of remote sensing-based DSM to derive long-term SOC dynamics, especially when supported by sufficient soil monitoring data. However, they also underline the factors limiting the validation, which will be further discussed in the following chapter.

Key implications of RQ10:

- Validations based on the soil monitoring data revealed that while it was possible to derive significant predictions of long-term SOC trends, the accuracy was low and influenced by a series of soil and management factors.
- Overall, the prediction of relative SOC trends was more reliable than the detection of absolute changes, underlining the potential of remote sensing data to support large-scale soil monitoring.

3.5 Usage of bare soil time series to derive further agricultural activity data

In addition to directly quantifying topsoil properties using SRC, remote sensing can provide auxiliary information on surface conditions and agricultural management, which in turn may serve as proxies for characterizing SOC dynamics. This includes data on the timing and duration of soil cover, such as green vegetation, crop residues, or bare fallow, which are key variables for estimating carbon inputs and losses in process-based SOC models (Ciais et al., 2010). In grassland areas, where exposed soils are rare and the generation of SRC is often not possible, the assessment of surface conditions remains the only approach for assessing the SOC dynamics. The findings in *manuscripts 5 & 6* show that bare soil time series can be effectively utilized to derive soil cover and management indicators, therefore supporting process-based models and soil monitoring in regions where direct observations of soil properties are not possible.

RQ11: How can cropland and grassland management data be derived from satellite time series to support soil monitoring?

The results in *manuscript 5* have shown that using a spectral unmixing approach, time series on the three main soil cover fractions (green vegetation/dead vegetation/bare soil) can be derived to describe key cropland management practices at a large scale. It was demonstrated that the selection of spectral endmembers, which are usually defined based on spectral libraries (Kowalski et al., 2023), can be optimized by using an SRC to estimate the bare soil reflectance for each cropland pixel individually (Lobert et al., 2025a). Using this novel approach, soil-specific spectral unmixing models can be calibrated and adopted to different bare soil conditions, which are known to interfere with the detection of vegetation cover (Montandon and Small, 2008). The resulting time series not only provides year-round information on the presence of bare fallows and the surface conditions, but also hold the potential to monitor key carbon farming measures such as winter cover cropping activities (Lobert et al., 2025b).

Furthermore, *manuscript 6* tested to what extent bare soil time series can

be derived to generate information on cropland conversions and estimate the establishment date of the current grassland areas. While sod renewals are a common management procedure for grassland soils, cropland conversions are usually accompanied by significant changes in bare soil frequency over time. The findings illustrate that harmonized Landsat and Sentinel-2 bare soil time series can be used to monitor the age of grassland plots in Germany. Using a rule-based classification approach, it was possible to differentiate short-term management practices (tillage, sod renewal, etc.) from long-term cropland conversions with high accuracy. As the establishment age of permanent grassland is an important factor influencing the SOC dynamics, the resulting maps can be used to complement the SRC-based soil maps and support large-scale monitoring.

Key implications of RQ11:

- Analyses of the bare soil reflectance can be used to optimize the endmember definition for spectral unmixing approaches, enabling to generate information on cropland soil cover over time.
- Based on harmonized bare soil time series, the established age of permanent grassland can be estimated to support soil monitoring in regions where no SRC can be derived.

Chapter 4

Conclusion

This dissertation investigated the extent to which remote sensing data can be applied to monitor changes in cropland SOC and meet the increasing demand for spatially explicit soil data. Overall, the results demonstrate that multispectral SRC, derived from Sentinel-2 and Landsat imagery, offer high potential for generating large-scale SOC maps and spatiotemporal predictions. The included manuscripts show that SRC hold strong explanatory power for the prediction of cropland topsoil properties such as SOC and clay. The remote sensing-based soil signal was found to be consistent with results from soil spectroscopy, supporting the calibration of robust and transferable DSM models. Using high-quality reference samples from the *German Agricultural Soil Inventory*, it was further demonstrated that the selection of spectral indices and thresholding values significantly influences the spectral quality and bias of the resulting SRC. In regions with heterogeneous soil conditions, prediction accuracy can be improved by applying *local ensemble models* that account for regional variability in the soil reflectance signal. Beyond static mapping, a harmonized bare soil time series was used to evaluate the potential for predicting SOC dynamics from 1986 to 2022. Based on the validation of the predicted trends, the findings of this dissertation allow for reassessing the limitations and opportunities of EO-based soil monitoring.

Relative to the total carbon stocks, SOC changes are typically small and subject to high uncertainties due to the small-scale spatial variability and measurement error. The SOC dynamics, therefore, pose a significant challenge for the quantification of carbon sequestration and the implementation of policies such as the CRCF regulation: While the detection of significant SOC changes requires a high density of repeated soil samples, the costs associated with such intensive verification strongly limit the scalability and economic feasibility of the proposed result-based approach. To address this limitation, integrated monitoring frameworks are needed that reduce verification costs while maintaining robustness to reduce the risk of greenwashing. The use of remote sensing for SOC monitoring has been proposed as a cost-efficient alternative; however, current approaches are often viewed as insufficient due to the lack of validation based on soil data.

Throughout *manuscripts 1* to *4*, it was shown that multispectral bare soil composites provide high explanatory power for the prediction of SOC and can increase the accuracy, resolution, and transferability of the resulting DSM products. Despite these methodological improvements, however, the prediction uncertainty remains significantly higher when compared to direct soil sampling. This presents one of the major limitations associated with EO-based soil monitoring and the detection of short-term SOC stock changes, which can be challenging, even when entirely relying on direct soil measurements. The findings of this dissertation support this observation, showing that the accuracy of remote sensing-based SOC predictions was not sufficient to reliably detect measured changes over 10-year intervals. As a result, it can be concluded that EO data alone should not be used to replace direct sampling when estimating absolute SOC changes over short timeframes.

Instead, spatiotemporal EO-based models can complement monitoring systems by supporting the optimization of soil sampling strategies. The models can be used to evaluate whether the outcomes of carbon farming measures are consistent with expected trends, therefore reducing the number of required field samples. As shown in *manuscript 4*, spatiotemporal models can provide significant information on relative SOC trends, even if the absolute changes remain below the overall prediction uncertainty. The confusion rate between increasing and decreasing SOC trends was low, indicating that EO-based models can be effectively used to monitor long-term soil changes. This enables a targeted sampling strategy, in which direct soil samples are prioritized at sites where predicted SOC trends diverge from expected effects of climate-smart management. Such strategies offer the potential to significantly reduce the verification costs while maintaining the robustness of the reported carbon removals.

In addition to evaluating recent SOC changes, EO-based models can capture long-term legacy effects resulting from historical land-use or management changes. This can be either done by assessing changes in the soil reflectance and topsoil SOC directly (*manuscript 4*) or by deriving proxies on the cropland soil cover (*manuscript 5*) and management history at sites with limited bare soil observations (*manuscript 6*). It has been shown that these legacy dynamics can persist for decades and strongly influence the outcome of carbon farming efforts. Incorporating such ex-post information increases the interpretability of present-day trends and strengthens the attribution of SOC change to the reported climate

mitigation practices. For example, if SOC stocks have been declining over multiple decades due to historic cropland conversion, carbon farming may not result in carbon sequestration but could reduce ongoing SOC losses. In such cases, EO-based models can help to reconstruct baselines in regions that lack historical soil data and provide context for the evaluation of the current management practices.

Chapter 5

Limitations & Further Research

While the results presented in this dissertation support the potential of EO-based models for SOC monitoring, they also highlight the need for further research and validation to enable their operational use within integrated soil monitoring systems. In particular, several key challenges remain regarding the understanding of the prediction uncertainties, especially in the context of SOC stock changes; the assessment of site-specific factors such as management practices and soil conditions; and the determination of the minimum time frames required to detect significant SOC dynamics. These aspects are critical for designing soil monitoring frameworks that balance cost and robustness to meet the demands of the emerging policies, such as the CRCF regulation.

5.1 Missing information on SOC stock and bulk density changes

To improve the verification of carbon farming measures, it is essential to consider site-specific factors influencing the vertical distribution of SOC in the soil profile, as well as the uncertainty of the measured and predicted trends. Changes in cropland management not only contribute to changes in SOC content but can also affect the bulk density, which is key for estimating the total carbon stocks (Bai et al., 2019). As a result, observed changes in the topsoil SOC contents do not necessarily indicate carbon sequestration and can be caused by vertical displacement of carbon within the soil profile (Ogle et al., 2019). This is supported by the results in *manuscript 4*, showing that positive trends in SOC contents were more frequently predicted at sites under reduced tillage. Yet, without corresponding bulk density data and information on the vertical SOC distribution, it is currently not possible to verify whether the observed changes can be attributed to carbon sequestration.

These findings highlight that the success of future soil monitoring framework programs will highly depend on identifying the key drivers behind both predicted

and measured SOC trends. To achieve this, it will be crucial to significantly expand the soil database on bulk density and SOC stock changes, which are necessary to further improve the model accuracy, as well as the validation of the predicted soil dynamics.

5.2 Assessment of site-specific factors

The results of the spatiotemporal model indicate that the accuracy of predicted SOC trends varies significantly across different soil types and cropland management conditions. The results in *manuscript 4* demonstrated that sites with low initial SOC content show the highest prediction uncertainties, despite representing areas with high potential for future carbon sequestration. Neglecting such variability may introduce bias in the evaluation of carbon farming outcomes across heterogeneous cropland landscapes. A deeper understanding of these site-specific interactions is therefore essential to improve the reliability of the SOC predictions for soil monitoring applications.

This issue is especially relevant under the recently adopted CRCF regulation, as it is expected to induce significant changes in the cropland management practices. It remains unclear, however, how the implementation of measures such as reduced or no tillage influences the prediction of SOC dynamics based on soil reflectance data. Currently, no systematic information exists on how such management changes interact with the spectral soil signal or the model's ability to predict SOC trends. In this context, it will be increasingly important to complement and compare the results of SRC-based maps with process-based soil models, which do not depend on the availability of bare soil observations. As shown in *manuscripts 5 & 6*, remote sensing can play a central role in providing high-resolution information and soil cover and cropland management, which are key for the estimation of model features such as C-input and turnover time (Wiesmeier et al., 2016).

5.3 Improving the time frame to detect SOC changes

The total length of the observation period is a critical factor influencing the robustness of the estimated SOC changes, both when relying on direct measurements and using models based on EO data (Croft et al., 2012). It has been shown that even under experimental conditions, it can take a decade or even longer to detect significant SOC changes, highlighting the importance of long-term analyses for reliable soil monitoring (Smith, 2004). To increase the effectiveness of the CRCF regulation, however, there is a need to minimize the time lag between the implementation of climate mitigation measures and the detection of the resulting SOC sequestration. As shown in *manuscript 4*, the signal-to-noise ratio of the presented spatiotemporal model is currently too low to support the prediction of short-term changes, limiting its applicability.

To reduce the detection interval, additional research and soil data are needed to evaluate whether the increasing availability of satellite imagery can positively affect the temporal resolution and prediction uncertainty of the resulting SOC maps. Future studies should focus on improving the signal-to-noise ratio, which is essential to enable the detection of short-term SOC dynamics and to increase the accuracy of the predicted long-term trends. Both factors will be crucial for the success of future carbon farming regulations, underlining the necessity to develop new methodologies and maximize the predictive power of the satellite data used as model input.

5.4 Retrieval of soil reflectance based on deep learning

The results of this dissertation have demonstrated that the generation of SRC provides a reliable basis for large-scale soil mapping. However, it can be assumed that key information is lost when deriving band-wise mean values from multi-temporal bare soil observations. While the temporal aggregation was a necessary step to support the generation of large-scale soil maps by reducing the noise from vegetation, soil moisture, and other surface factors, these approaches may be increasingly outdated considering the recent advances in deep learning methods

(Miller et al., 2024). The current SRC-based workflow, which is widely adopted by DSM studies, relies on the extraction of EO-based covariates (e.g., spectral indices, reflectance composites) that are compatible with conventional machine learning algorithms such as *Random Forest* or *Support Vector Machines*. While this two-step approach reduces model complexity and increases the interpretability of the resulting maps, it utilizes only a limited portion of the soil information contained in the original EO time series.

An alternative approach would be to bypass the generation of soil composites entirely and directly integrate the full multitemporal time series into a single deep learning model (Tseng et al., 2023). Using such a one-step modelling framework, it would be possible to exploit the temporal dimension of bare soil observations, capturing features related to tillage timing, sowing dates, and cropland phenology (Tarasiou et al., 2023). In addition, soil-related variables such as water content, surface roughness, and residue cover can provide further valuable information, which are usually averaged out during SRC generation.

Despite these promises, this approach remains untested in the context of SOC mapping and presents several methodological challenges. Deep learning models generally require larger training datasets compared to traditional algorithms (Ng et al., 2020). It is therefore necessary to assess whether the current soil sample availability is sufficient to support such models or if an expansion of the existing soil monitoring network is required. Additionally, it should be explored if the higher demand can be met with other strategies, for example, by matching the sample layout to the spatial resolutions of the remote sensing data to increase the quality of the reference data. This is especially important for soil properties like SOC, which show a high spatial heterogeneity that must be adjusted to the satellite data (Poeplau et al., 2022). Future research should therefore focus on evaluating the trade-offs between model complexity, data availability, and prediction accuracy to advance EO-based SOC monitoring beyond the existing SRC approaches.

5.5 Hyperspectral satellite data

In addition to methodological advancements, the increasing availability and quality of hyperspectral satellite data offer substantial potential for improving EO-based soil mapping (Guanter et al., 2015). Although the current spatial coverage

of operational hyperspectral sensors such as EnMAP is insufficient to support large-scale, wall-to-wall soil predictions, this limitation is expected to improve significantly with the deployment of upcoming satellite missions (Rast et al., 2021). Recent studies have demonstrated that hyperspectral data can enhance the prediction of key soil properties, including SOC (Chabrillat et al., 2024). However, they also emphasize that compositing strategies for hyperspectral imagery are more complex than for multispectral data and require further methodological developments. Challenges include the preprocessing and data cleaning, data sparsity and cloud coverage, as well as seasonal effects and the influence of soil moisture and vegetation cover (Zhao et al., 2025). Given these limitations, it can be concluded that in the short term, the benefits of comprehensive multispectral archives such as Sentinel-2 or Landsat are likely to outweigh the advantages provided by the additional spectral detail of hyperspectral sensors. Therefore, a critical next step will be to determine under which conditions and for which soil properties, the increasing availability of hyperspectral data can outperform models based on multispectral composites. This assessment will be essential for guiding future investments in data processing pipelines and defining the role of hyperspectral data within the emerging soil monitoring systems.

Chapter 6

Relevance for Policy and Practice

The results of this dissertation highlight that, while remote sensing alone is not yet sufficient to quantify short-term SOC changes, EO-based models offer significant potential to enhance the scalability and cost-effectiveness of SOC monitoring frameworks. The following considerations are derived from the findings to support the effective implementation of emerging policies such as the CRCF regulation and the upcoming SML.

6.1 Hybrid monitoring systems with targeted field sampling

EO-based spatiotemporal models should be integrated into hybrid monitoring frameworks that combine large-scale soil mapping with repeated soil sampling to ensure the robustness of the reported carbon removals. Model outputs can be used to guide targeted field campaigns by identifying locations where the predicted SOC trends do not align with the expected outcomes of carbon farming measures. This approach can enhance the cost-effectiveness of soil monitoring by focusing sampling efforts on regions where they are most needed, without compromising the verification of SOC changes.

6.2 Expansion of soil monitoring networks

To improve the validation of predicted trends, long-term monitoring networks based on repeated soil samples must be extended to include data on bulk density and the vertical distribution of SOC. This information is essential for accurate estimations of SOC stock changes and the evaluation of site-specific factors such as management practices. Furthermore, it must be addressed that changes in topsoil SOC content do not necessarily imply net carbon sequestration, as they may result from vertical relocations within the soil profile. The incorporation of additional soil monitoring data will significantly reduce the risk of misinterpreting observed SOC trends and increase the reliability of the predictions by providing

additional validation.

6.3 Integration of long-term observation periods and historical land-use context

Given the slow dynamics of SOC, robust verification requires long-term observation periods (>10 years) to distinguish carbon sequestration from measurement uncertainties and seasonal fluctuations. Policies should therefore specify minimum time frames for both the measured and predicted SOC changes. Furthermore, the interpretation of recent SOC trends must be supported by information on legacy effects from historical land-use changes, which can affect the short-term dynamics. EO-based models can enable the estimation of baseline conditions, allowing for more accurate attribution of SOC changes to the implementation of recent carbon farming measures.

6.4 Standardized validation of EO-based soil models

To ensure comparability and transparency, standardized guidelines should be established for accuracy and uncertainty reporting in EO-based SOC models. This includes the use of external soil datasets with repeated measurements for validating spatiotemporal predictions of SOC changes, as well as documentation on the fitness-for-use, spatial uncertainty, and potential biases in static soil maps. Implementing such validation frameworks is essential to support the development of improved methods for the retrieval of soil information from satellite data.

Appendix

Chapter 7

Manuscript 1: Using local ensemble models and Landsat bare soil composites for large-scale soil organic carbon maps in cropland

Tom Broeg, Axel Don, Alexander Gocht, Thomas Scholten, Ruhollah Taghizadeh-Mehrjardi, Stefan Erasmi

Geoderma, 2024

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

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 been proving 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., 2021a). 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 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., 2018b), 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). 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., 2020a). 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 et al., 2010; Stenberg, 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., 2020b). 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 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) (Figure 7.1 a). 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, 2007a) (Figure 7.1 b). The North German Pain 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 Pain, organic soils are present in lowland areas and widely drained for agricultural purposes (Figure 7.1 b). 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 (Figure 7.1 b). Further west, the South German Scarplands are built by strata of the Triassic and Jurassic. Here, the soil conditions are very diverse (Figure 7.1 b). 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 (Figure 7.1 b). 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.

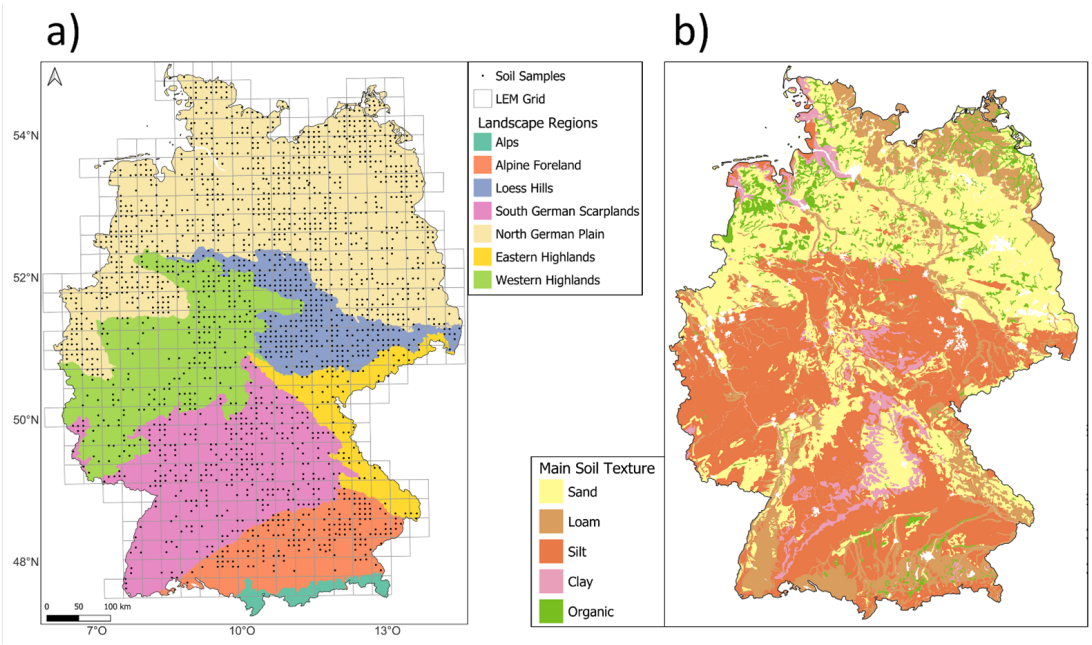

Figure 7.1: Overview of the research area: a) Major landscape regions of Germany. The tiling scheme of the LEM and the soil samples (BZE-LW) used to train the SOC models are shown in black; b) Main soil texture based on (BGR, 2007a). White areas are urban areas or mining areas.

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 0-10 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 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 (Figure 7.2 + Table 7.3). 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} (Figure 7.2). Extreme values on organic soils (> 87 g kg^{-1}) were unrepresented but included to cover the full range of cropland SOC values.

Table 7.1: Descriptive statistics of SOC contents (g kg^{-1}) for each landscape region (Figure 7.1). SD = standard deviation; IQR = interquartile range.

Landscape Region	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

Landscape Region	Samples	Min	Median	Mean	Max	SD	IQR
Total	1966	2.0	14.0	16.2	135.2	9.8	7.7

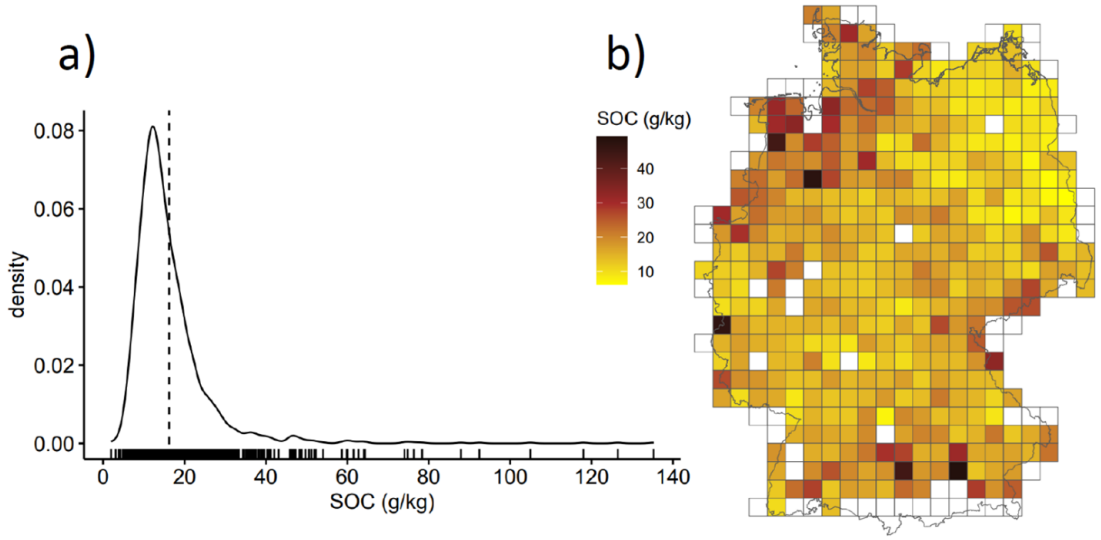

Figure 7.2: a) Density plot of the SOC content (g kg^{-1}) of 1966 soil samples across German croplands used for model training and validation (dotted line = mean value). Individual observations are shown above the x-axis; b) Mean SOC contents (g kg^{-1}) of the soil samples within the LEM tiles (white = no samples available).

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., 2016a). 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 Figure 7.3 a. 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).

$$NDVI = \frac{NIR - Red}{NIR + Red} \tag{1}$$

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

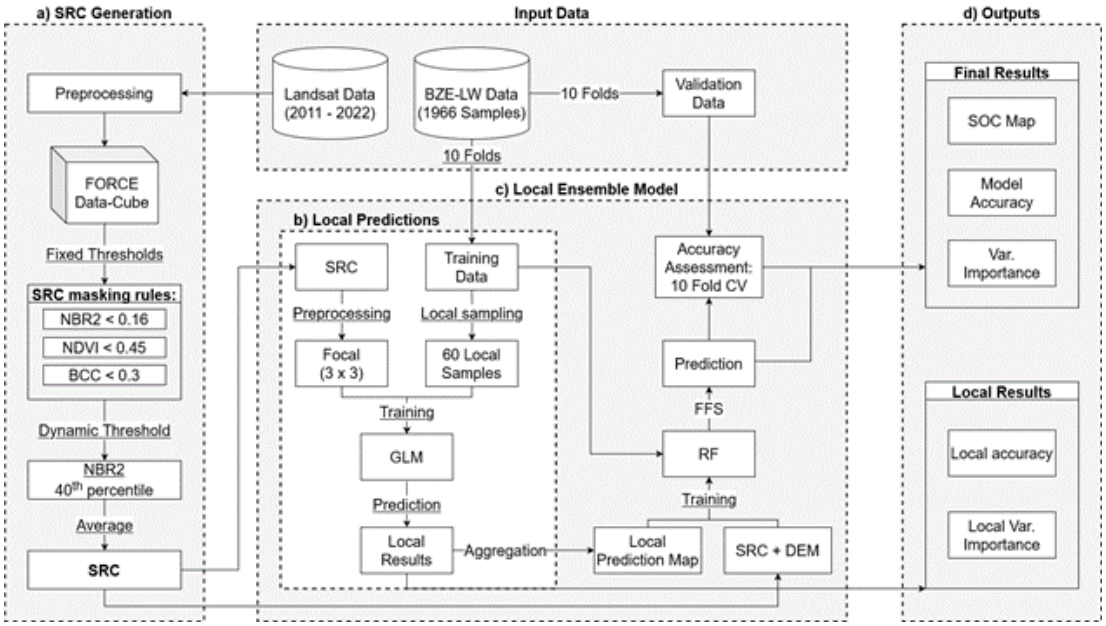

Figure 7.3: Flow chart of the Local Ensemble Model: a) Generation of the SRC based on Landsat data (section 2.4); b) Training and prediction of the local GLMs, based on the SRC and local training samples; c) Aggregation of the local maps and RF model for the final SOC prediction, based on all training samples; d) Outputs of the LEM: Final SOC map (RF) and the local model results (GLMs).

West Germany, however, we observed that higher NDVI thresholds are necessary to cover all cropland areas (Figure 7.14). 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., 2019a; Demattê et al., 2018; Dvorakova et al., 2022):

$$NBR2 = \frac{SWIR1 - SWIR2}{SWIR1 + SWIR2} \tag{2}$$

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 for 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., 2019a; 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):

$$BCC = \frac{Blue}{Red + Green + Blue} \tag{3}$$

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., 2019a; 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) (Figure 7.14). 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 (Figure 7.15). 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 (Figure 7.15). 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 Figure 7.4.

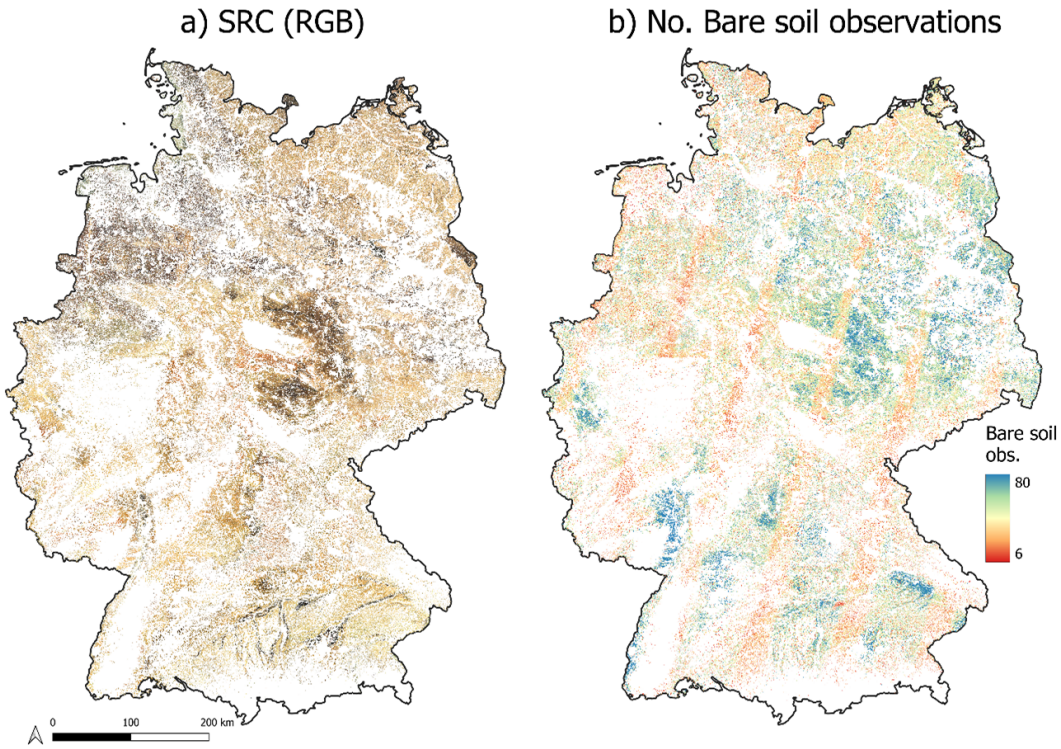

Figure 7.4: a) Final SRC based on Landsat data from 2011 to 2022, RGB = Red-Green-Blue composite; b) Number of bare soil observations for each pixel.

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.

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 7.2.

Table 7.2: 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

Variable	Abbreviation	Wavelength (μm)	Models
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

2.5.1 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. (2018b): 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 Figure 7.3. The LEM was conducted in three parts: First, multiple geographically distinct sub-models were trained and used for local SOC predictions (Figure 7.3 b). 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 (Figure 7.3 c). 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 (Figure 7.5). 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 (Figure 7.5). 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 (Figure 7.5). 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 Figure 7.5, 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 (Figure 7.3 b). 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 (Figure 7.5). 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. Like with the global model, additional distance covariates were included to assess the effect on the model performance (LEM + Distance).

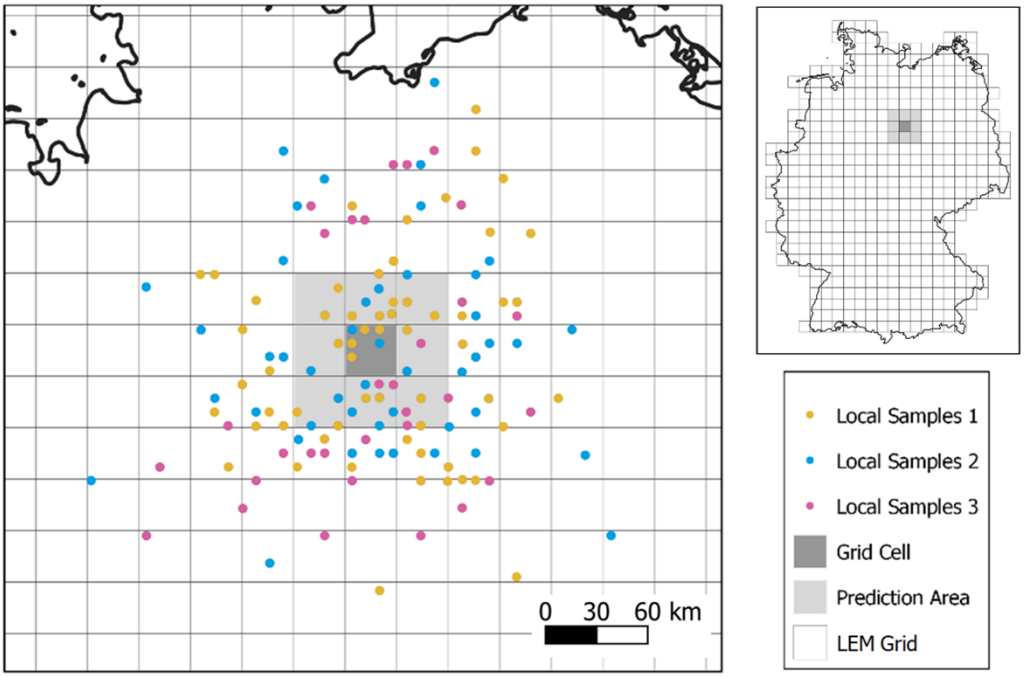

Figure 7.5: Sampling and prediction process for the local models: For each LEM grid tile (dark grey), 60 nearby samples were randomly selected. This process was repeated three times (colored dots) to obtain three local training sets. Each local training set was used to train a GLM and predict the SOC 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.

2.6 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 (Figure 7.3). 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 ($\pm 1\ \sigma$) 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).

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 Figure 7.6 a. 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 Figure 7.6 b. 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 (Figure 7.6 b). Variograms were calculated to test the degree of spatial autocorrelation in the cross-validation residuals (Figure 7.6 c). 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 (Figure 7.6 c). In this case, the addition of distance covariates only marginally affected the results.

The spatial distribution of the prediction residuals is shown in Figure 7.7. 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 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.

Table 7.3: 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	6.40	0.72	1.20	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

3.2 Final soil organic carbon predictions

The final SOC map of the LEM is presented in Figure 7.8. 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 Figure 7.9. 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 Figure 7.9 a+b. The LEM predicted higher extreme values and increased the local range of SOC contents (Figure 7.9 a). The integration of distance generally increased the predictions in some regions, leading to a smaller range of local SOC values in comparison to the LEM (Figure 7.9 b). Uncertainty maps of the SOC predictions are shown in Figure 7.10. The prediction intervals show a similar spatial pattern in all models and are strongly connected to the SOC content (Figure 7.8) and the

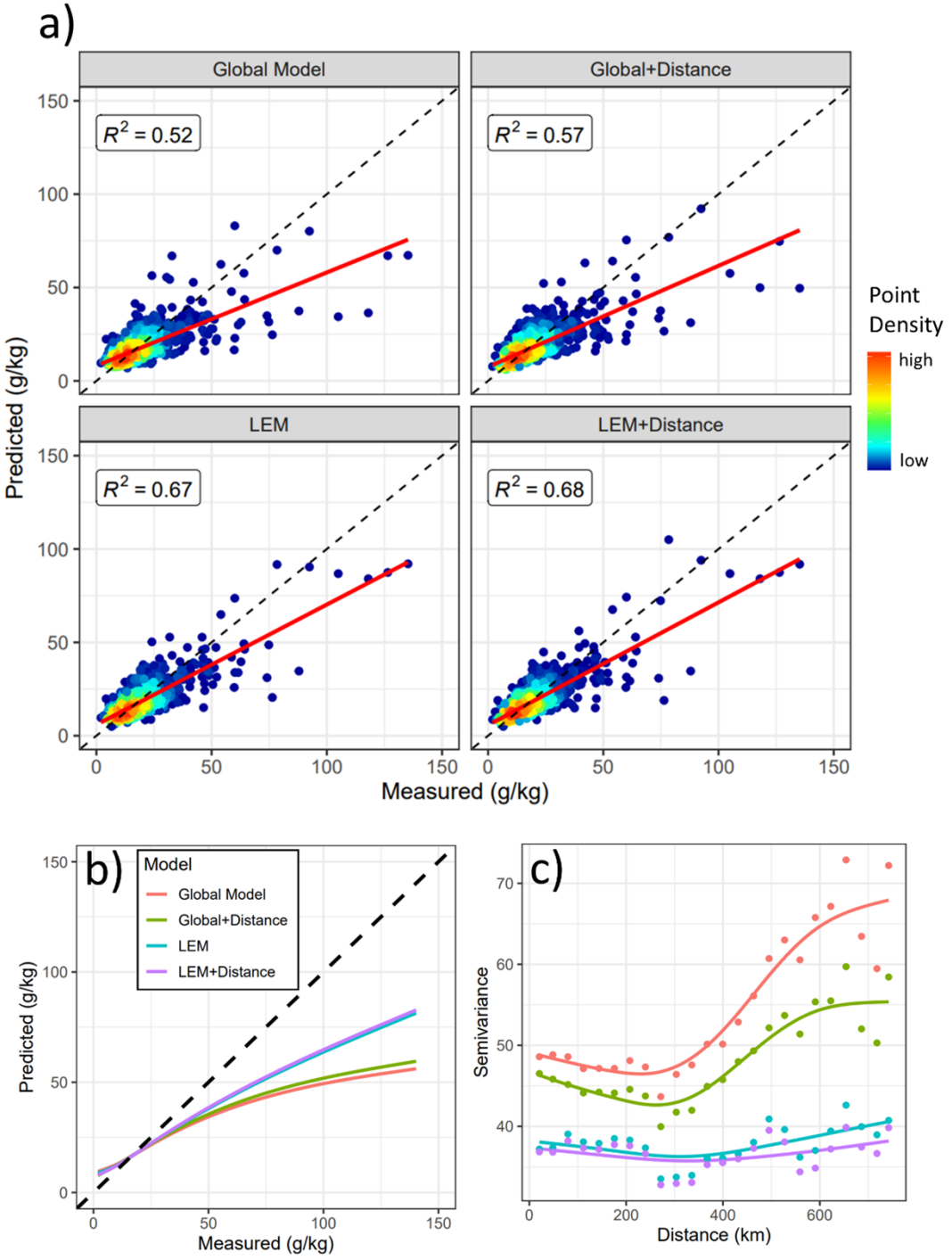

Figure 7.6: a) Scatter plots and R^2 values of the cross-validated predictions and the measured topsoil SOC contents (g kg^{-1}). Red = regression lines, black = 1:1 lines; b) Direct comparison of the local regression lines (LOESS), based on the predictions of each model. Black = 1:1 line; c) Variograms, based on the cross-validation residuals of each model.

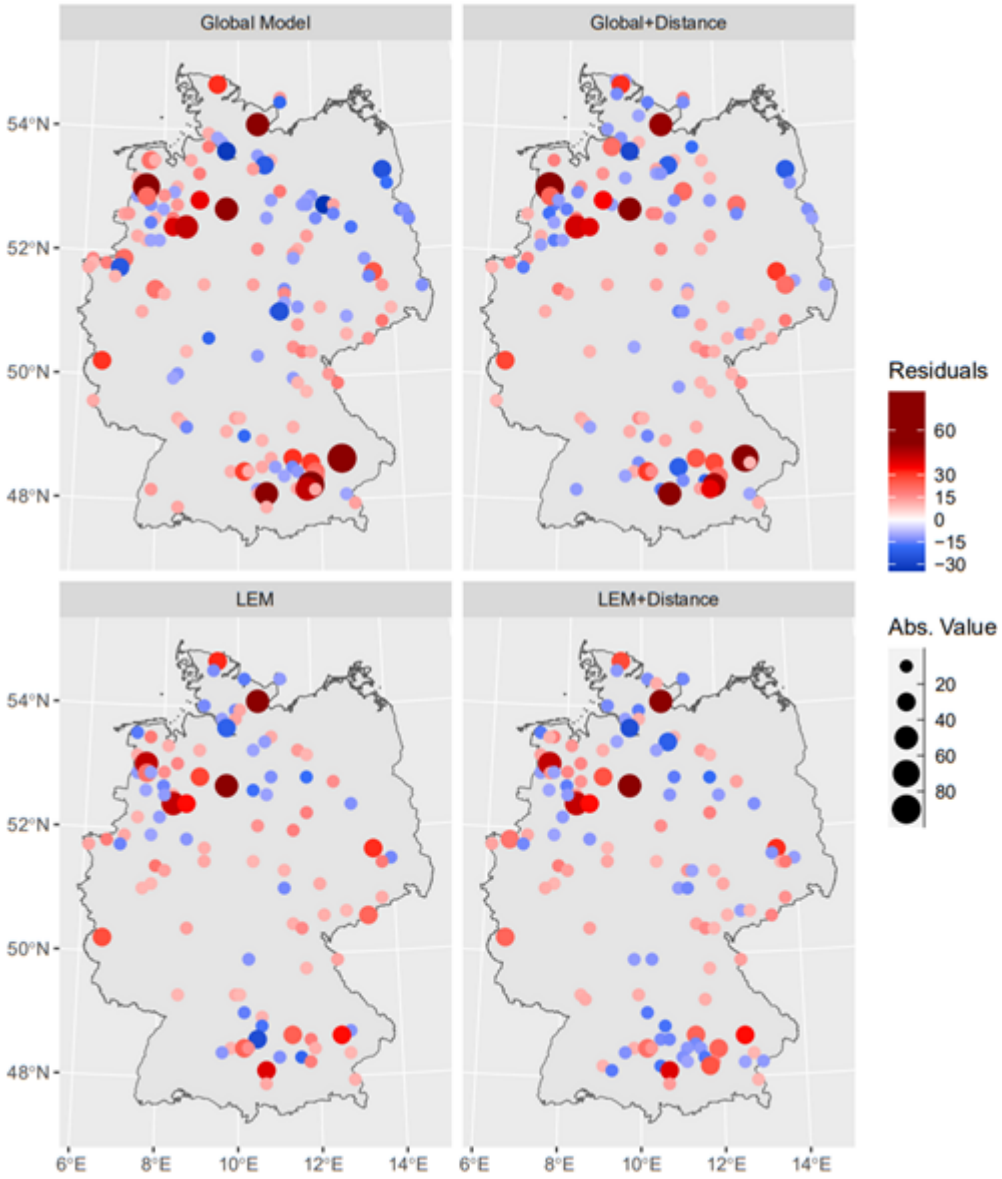

Figure 7.7: Spatial distribution of the cross-validation residuals of the topsoil SOC content (g kg^{-1}). Only residuals with absolute values above 10 g kg^{-1} are displayed. Red = underpredictions, Blue = overpredictions.

NBR2 values of the SRC (Figure 7.15). 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.

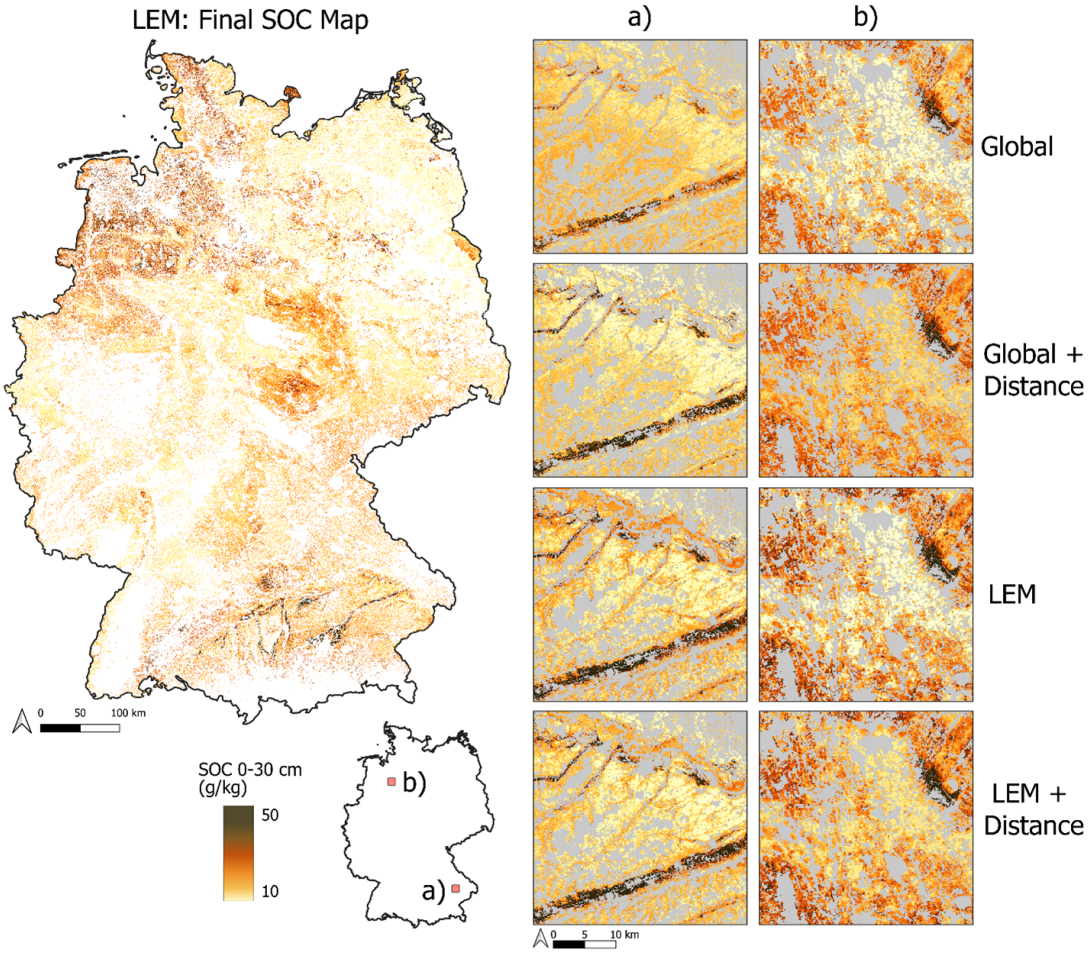

Figure 7.8: Left: Final SOC map of the LEM; Right: Direct comparison between examples (30x30 km tiles) for the SOC predictions of the LEM and the global models. The sections were enlarged to showcase small-scale differences between the maps.

3.3 Variable importance of the final soil organic carbon models

The relative variable importance of the final RF models is shown in Figure 7.11. The red bad was by far the most important variable for the global model (>

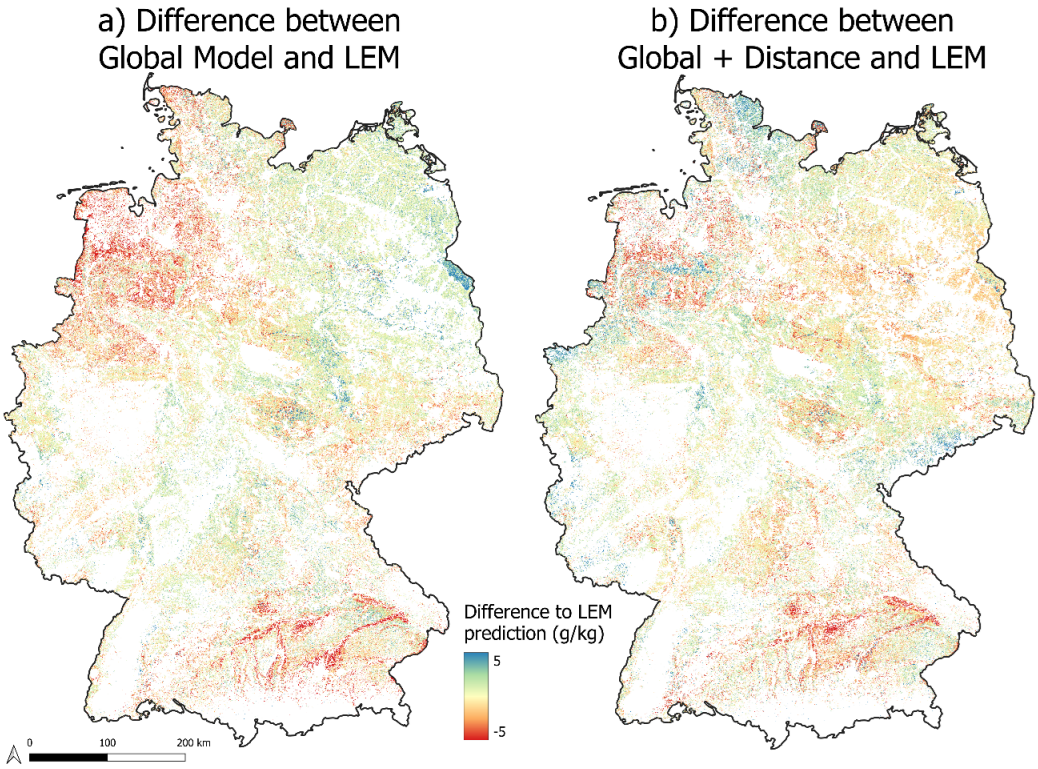

Figure 7.9: a) Difference (g kg^{-1}) between the SOC predictions of the LEM and the global model; b) Difference (g kg^{-1}) between the SOC predictions of the LEM and the global + distance model; Red = lower predictions of the global models compared to LEM; Blue = higher prediction of the global models compared to LEM; Yellow = no difference.

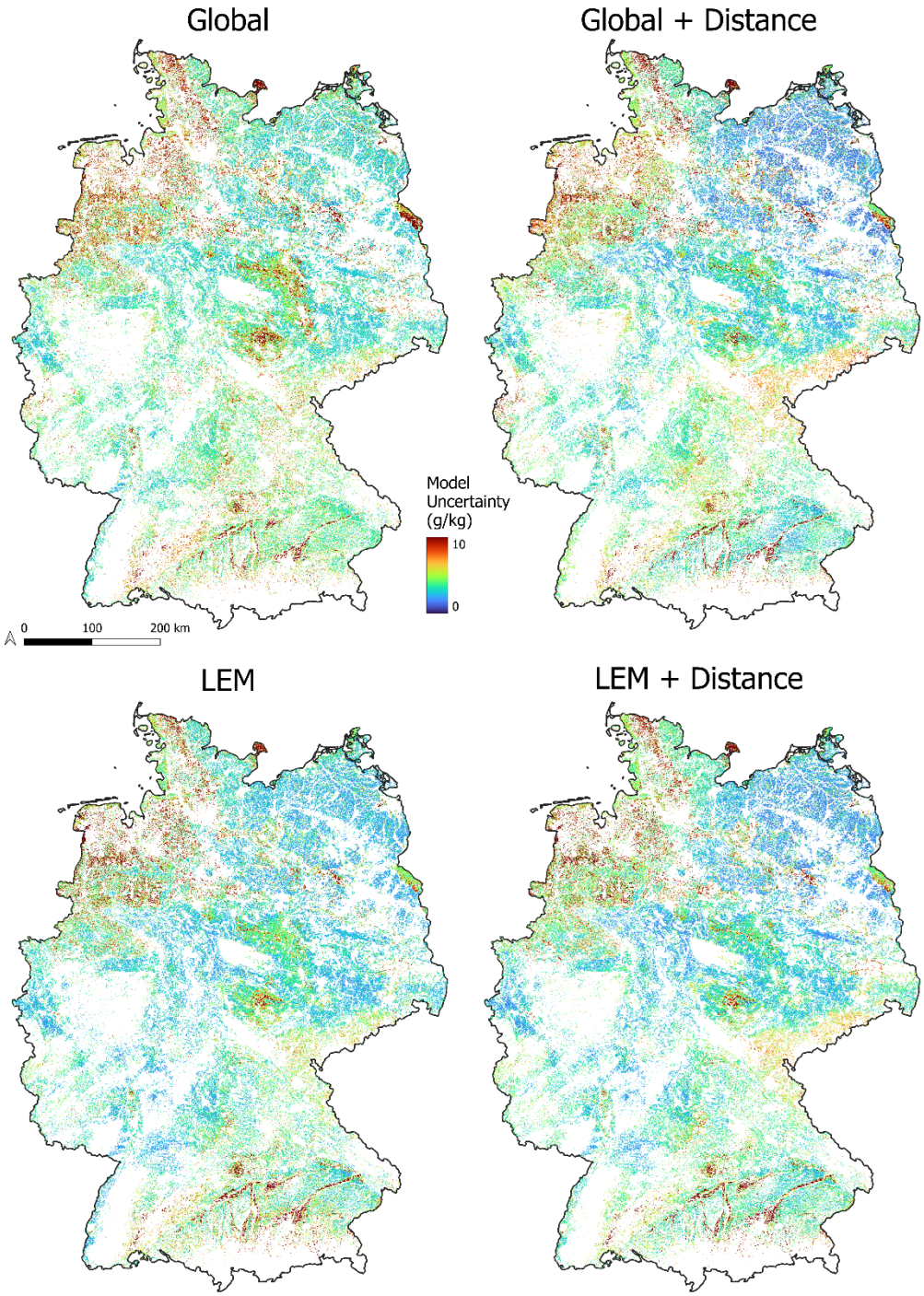

Figure 7.10: Uncertainty maps of the final RF predictions: Model uncertainty (g kg^{-1}) within one standard deviation of the prediction interval.

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 %.

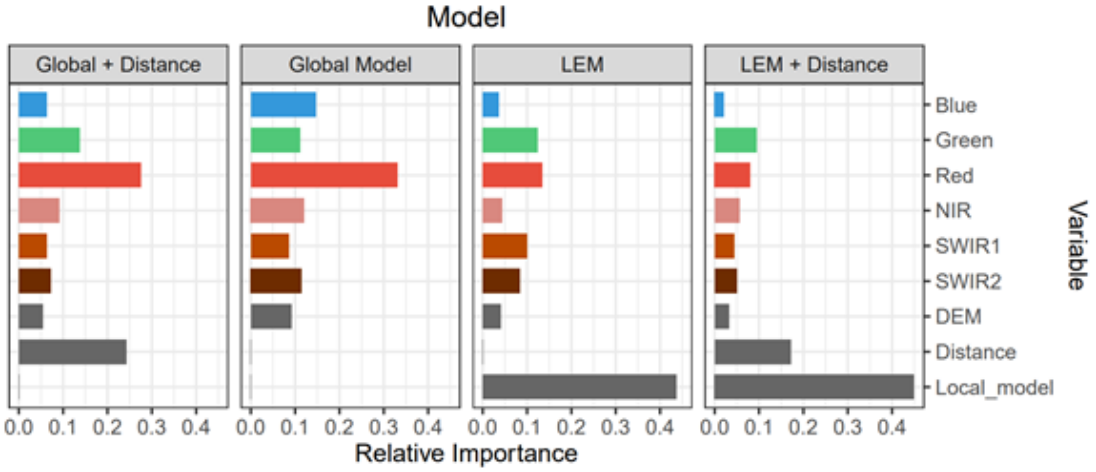

Figure 7.11: Relative variable importance of the final SOC models; Blue/Green/Red/NIR/SWIR1/SWIR2 = SRC bands; Distance = aggregated importance of the distance covariates; Local_model = aggregated predictions of the local GLMs; DEM = digital elevation model.

3.4 Local model accuracy

The accuracies of the local sub-models are shown in Figure 7.12. 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 Figure 7.12 c, illustrating a significant positive correlation between both measures.

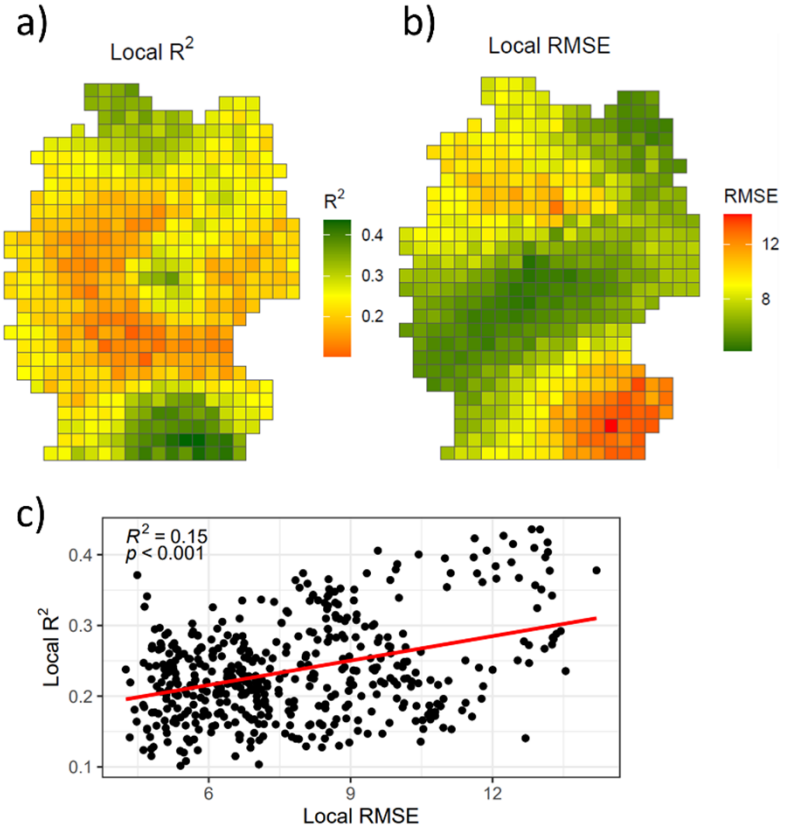

Figure 7.12: Model accuracies of the local GLMs. Mean values of all model repetitions were calculated for each tile: a) Local R^2; b) Local RMSE; c) Correlation between the local R^2 and local RMSE. Red = regression line.

3.5 Local variable importance

Based on the GLMs, the local variable importance of the SRC was calculated for each grid tile (Figure 7.13). 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 (Figure 7.13 c) with high values throughout Germany (Figure 7.13 a). 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 (Figure 7.13 a). 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 (Figure 7.13 b). 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 Figure 7.2, training samples above 40 g kg^{-1} were generally underrepresented in the global model, leading to poor prediction for high SOC values (Figure 7.6 b). 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} (Figure 7.6 b). As shown in Figure 7.9, the LEM predicted higher SOC values in south and northwest Germany and generally reduced the number of underpredictions (Figure 7.7). 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

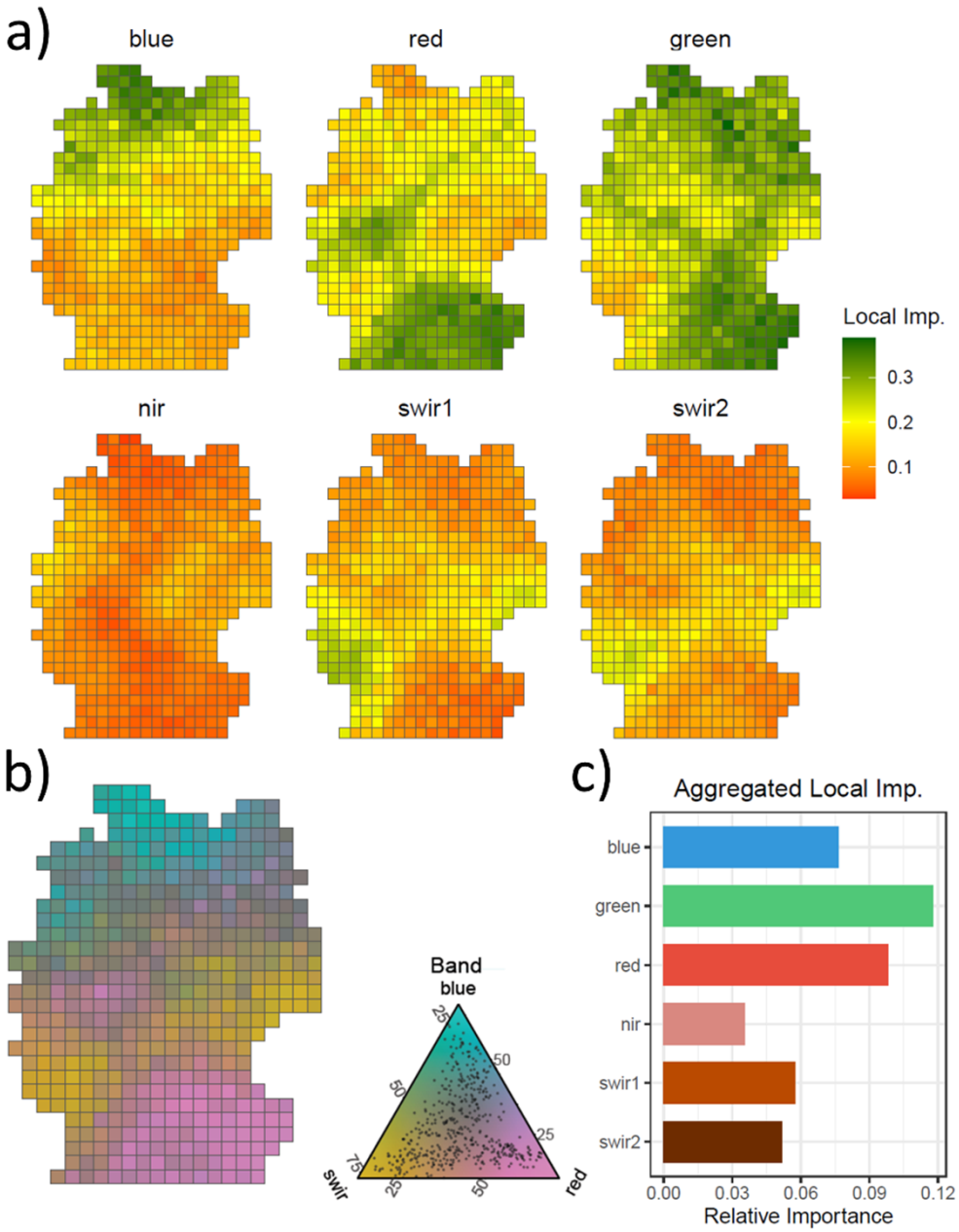

Figure 7.13: Results of the local variable importance, based on the GLMs: a) Relative local importance of each spectral band; b) Ternary plot of the relative local importance of the blue, red, and SWIR bands. SWIR1 and SWIR2 bands were aggregated for the illustration. c) Aggregated local importance of all grid tiles.

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 (Figure 7.4) (Castaldi et al., 2019a; Nocita et al., 2013; Stenberg et al., 2010). Consequently, the global model showed a strong tendency to overpredict the SOC concentrations in this region (Figure 7.7). 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 (Figure 7.9) leading to fewer overpredictions in the Oderbruch (Figure 7.7). These differences between the predictions of the LEM and the global can also be illustrated with the uncertainty maps. As shown in Figure 7.10, the LEM produced lower uncertainties in regions with high NBR2 values (Figure 7.15), 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 Figure 7.5, 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., 2020a).

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., 2020a). 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 (Figure 7.13). 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 (Figure 7.11). 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 the global model, the variable importance of the different spectral bands was more homogenous for the LEM (Figure 7.11 + Figure 7.13 c) and expressed clear spatial patterns across the research area (Figure 7.13 a+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 (Figure 7.12 c). 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 (Figure 7.12). 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 error in these regions could 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 Figure 7.10, underlying that the spatial uncertainty of the LEM shows a similar pattern to the RMSE values of the local GLMs.

Explainability

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 (Figure 7.13). For the sandy soil of the North German Plain (Figure 7.1), 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 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 SWIR spectrum (Demattê et al., 2018; Stenberg et al., 2010). Previous studies proved the capacity of multispectral satellite data to predict the clay content in cropland topsoils (Gholizadeh et al., 2018). 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 (Figure 7.1 + Figure 7.13). The results show that the clay minerals are affecting the spectral signal and therefore influence the

correlation between the SWIR bands and the SOC (Stenberg et al., 2010).

4.3 Influence of the distance covariates

Distance covariates have been commonly integrated into the DSM framework to improve model performance (Behrens et al., 2018b; 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 7.2; Figure 7.6). 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 Figure 7.11, they were important predictors and partially masked out the relevance of the spectral bands (Figure 7.11). A similar concern was raised by Wadoux et al. (2020a), highlighting that the integration of distance covariates can 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. Figure 7.8 b 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 (Figure 7.6 b). The distance covariates only marginally improved the results but showed high variable importance for the final SOC predictions (Figure 7.6 + Figure 7.11). These findings support the concern of Wadoux et al. (2020a), 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 (Figure 7.6 c). 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 (Figure 7.14). 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., 2019a; 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 Figure 7.15, 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 are mostly affected by high soil moisture contents, e.g. floodplains or lowlands. Here, the model showed higher uncertainties as the spectral signal of the SOC is influenced by the presence of soil water (Figure 7.10). 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 Figure 7.4, 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 (Figure 7.8 + Figure 7.10). However, multiple studies have highlighted a connection 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.

Acknowledgements

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

Appendix

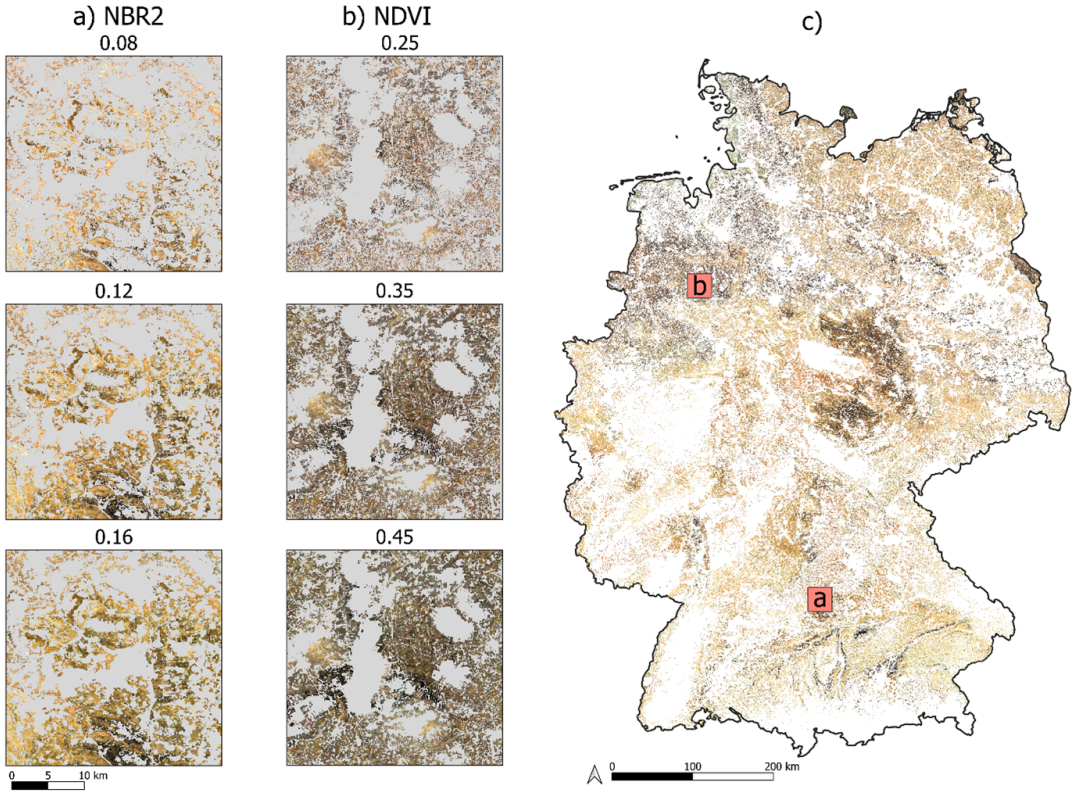

Figure 7.14: Reference tiles highlighting the visual selection of NBR2 and NDVI thresholds for the generation of the SRC: Low thresholds (NBR2 = 0.08; NDVI = 0.25) significantly reduce the extent of the SRC by excluding soils with higher soil moisture and SOC contents. Higher thresholds (NBR2 = 0.16; NDVI = 0.45) were selected to remove bias and to include the full range of cropland soils.

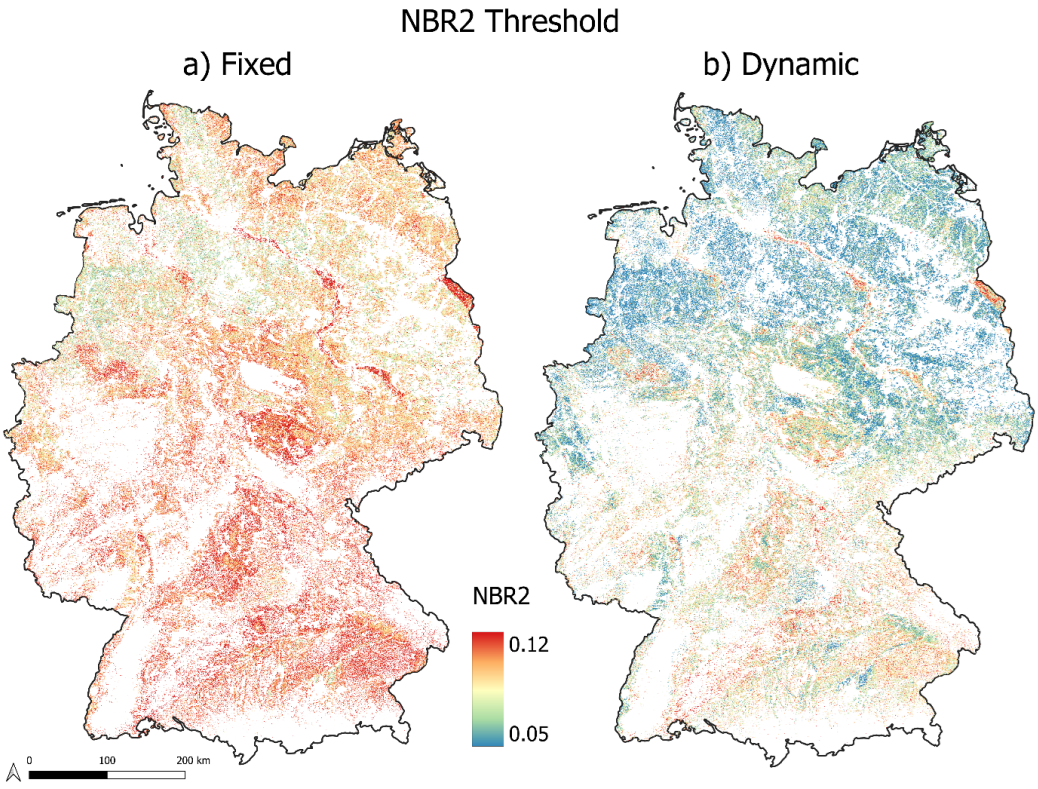

Figure 7.15: Mean NBR2 values of the SRC before (a) and after applying the dynamic NBR2 thresholds (b).

Chapter 8

Manuscript 2: Reducing bias in cropland soil organic carbon and clay predictions using Sentinel-2 composites and data balancing

Tom Broeg, Axel Don, Thomas Scholten, Stefan Erasmi
Remote Sensing of Environment, 2025

https://doi.org/10.1016/j.rse.2025.115109
Received 9 May 2025; Received in revised form 9 September 2025; Accepted 26 October 2025; Available online 6 November 2025

Abstract

Accurate maps of cropland soil organic carbon stocks (SOCS) and clay content are essential for climate-smart agriculture. Soil reflectance composites (SRC), derived from multispectral bare soil observations, offer a scalable approach to high-resolution soil mapping. While studies often focus on maximizing model performance, challenges remain regarding (1) the bias introduced by masking and excluding soil samples during SRC generation and (2) the accurate representation of the full range and distribution of soil properties in the resulting maps. Evaluating different SRC parameters, we found that commonly used indices such as the Normalized Burn Ratio 2 (NBR2) and the Normalized Difference Vegetation Index (NDVI) were significantly correlated with clay content and SOCS, respectively. These dependencies can lead to the systematic exclusion of high SOCS (>80 Mg ha^{-1}) and clay (>30 mass%) samples during SRC generation, introducing bias in the resulting maps. Models trained solely on SRC bands failed to capture the full range of the training data, limiting the applicability of the soil property maps. While the inclusion of additional remote sensing features, such as spectral-temporal metrics and indices, significantly improved the prediction accuracy, the representation of the imbalanced samples remained challenging. We demonstrated that a combined framework of spatial data augmentation and majority undersampling was effective in improving the range and concordance correlation coefficient (CCC) of the predictions (SOCS = 0.82; Clay = 0.9). Our findings emphasize the importance of (1) evaluating excluded samples to identify potential SRC-induced bias, and (2) optimizing model predictions reflecting the observed data range to improve the reliability and usability of the resulting soil maps.

1 Introduction

It is estimated that due to land use change and cropland intensification, most agricultural soils have lost up to 50 % of their original carbon stocks, significantly contributing to climate change (Lal, 2003). As the largest terrestrial C pool, soil organic carbon (SOC) is a central aspect of European policies like the Green Deal, targeting to achieve climate neutrality by 2050 (Montanarella and Panagos, 2021). The sequestration of atmospheric CO_2 in agricultural soils, promoted by

initiatives such as '4 per 1000', is widely recognized as a win-win strategy to both mitigate climate change and improve soil health (Minasny et al., 2017). Most of the SOC stock (SOCS) is concentrated in the topsoil (0–30 cm) and is essential for soil functions, including erosion control, nutrient cycling, water retention, and aggregate stabilization (Murphy, 2015). As a key component of the global carbon cycle, SOCS are dynamic and vary spatially and temporally due to agricultural management practices (e.g., fertilization, tillage, intensification) and site-specific biophysical factors such as climate and soil properties (Wiesmeier et al., 2019). Clay minerals play a crucial role in carbon retention due to their high sorption rates and ability to stabilize SOC in soil aggregates, making the clay content a key factor in predicting and interpreting SOCS (Prout et al., 2021; Rasmussen et al., 2018).

The demand for spatially explicit information on biogeochemical and physical soil properties has grown significantly in recent decades (Paustian et al., 2019). National soil sampling campaigns, such as the German Agricultural Soil Inventory (BZE-LW), can provide such soil data but are time-consuming and limited in sample size (Poeplau et al., 2020b). Based on soil measurements, it has been shown that the SOCS can vary significantly across regional, local, and plot scales, further increasing the need for high-resolution soil data (Don et al., 2007; Poeplau et al., 2022). To address this issue, various digital soil mapping (DSM) frameworks have been developed to predict soil properties at different levels (Chen et al., 2022). Earth observation (EO) can play a crucial role in improving the accuracy, resolution, and transferability of predicted maps (Broeg et al., 2023; Tziolas et al., 2021). Recent works highlight the potential of soil reflectance composites (SRC), derived from multispectral bare soil observations, to directly analyze cropland soil properties through spectral responses (Broeg et al., 2024a; Demattê et al., 2018; Diek et al., 2017; Rogge et al., 2018). With the growing number of studies, however, questions arise on what factors should be considered when optimizing and comparing different SRC frameworks for the prediction of reliable and unbiased soil maps (Vaudour et al., 2022).

Most soil reflectance composite (SRC) workflows use spectral indices to determine whether a pixel represents bare soil or is influenced by vegetation or other surface covers (Delaney et al., 2025). In general, the selection of optimal indices and thresholding values serves two key purposes. First, pixels and samples without valid bare soil observations are masked out and excluded from the soil

map. Second, the highest-quality bare soil observations are selected to enhance the soil signal and improve the model accuracy (Dvorakova et al., 2022). Studies have shown that methods of soil spectroscopy can be applied to optimize and refine the SRC parameter settings (Dvorakova et al., 2020; Heiden et al., 2022). Demattê et al. (2018) demonstrated that under laboratory conditions, the Normalized Difference Vegetation Index (NDVI) increases incrementally when adding photosynthetic vegetation, and recommended a threshold of 0.25 to select bare soil pixels. Other studies based on spectral soil libraries indicate, however, that the NDVI is not only influenced by the presence of vegetation but also by the soil itself (Montandon and Small, 2008). As a result, NDVI thresholds of 0.35 and higher have been recommended to account for a broader range of bare soil conditions (Broeg et al., 2024a; Safanelli et al., 2020). While many studies focus on maximizing the spectral quality of the soil signal, less is known about the interactions between bare soil indices and soil properties and the potential influence on the exclusion of specific soil conditions during SRC generation.

In addition to laboratory analyses, efforts have been made to evaluate and compare SRC parameters based on the performance of the final soil models. It has been suggested that predictive power can serve as a proxy for spectral quality, as it reflects the selection of the purest bare soil observations for the SRC (Dvorakova et al., 2021). Multiple studies have highlighted the significant influence of the Normalized Burn Ratio 2 (NBR2), derived from shortwave infrared (SWIR) bands, on the SOC prediction accuracy (Gomez et al., 2022; Vaudour et al., 2021). Low NBR2 values are associated with optimal seedbed conditions, characterized by dry, homogenized soils that are free from vegetation and residues (Dvorakova et al., 2022). Further research has shown, however, that NBR2 can also be affected by the soil clay content, complicating its interpretation (Shabou et al., 2015). While the most accurate SOC models are typically achieved using very low NBR2 thresholds (<0.08), values up to 0.15 are often required to capture a broader range of bare soil conditions (Demattê et al., 2018; Safanelli et al., 2020). In many cases, this trade-off has been resolved by prioritizing prediction accuracy over spatial coverage, significantly reducing the number of soil samples available for modeling (Castaldi et al., 2019a; Demattê et al., 2018; Dvorakova et al., 2022; Silvero et al., 2021; Vaudour et al., 2021). Considering these interdependencies, questions remain about whether the selection of bare soil indices and thresholds affects the exclusion of certain soil parameters (e.g., the clay content)

from the soil models, influencing the bias and applicability of the final maps.

Another key challenge in DSM is the accurate representation of soil properties with skewed distributions, such as the SOC, which often includes outliers and extreme values that are underrepresented in the training data. Studies have shown that in this case, models tend to underpredict high values and fail to reproduce the SOC range of the samples (De Brogniez et al., 2015; Rawlins et al., 2009). Feeney et al. (2022) argue that the availability and quality of environmental covariates are major factors limiting model performance and the accurate representation of extreme values. Zepp et al. (2021a) showed that the calculation of additional spectral indices, based on the original SRC bands, improved model accuracy but did not affect the range of predicted SOC values. Other studies reported that including further EO-based covariates such as spectral temporal metrics (STM) (Stumpf et al., 2024) or digital elevation models (DEM) (Žížala et al., 2022) can improve the overall model performance, though at an increased risk of overfitting (Meyer et al., 2019). It has been shown that in classification tasks, such as the predictions of soil types, DSM models tend to minimize uncertainty by favoring majority classes (Rau et al., 2024). While issues regarding class imbalance have been successfully addressed using over- and undersampling (Sharififar et al., 2019), it remains unclear whether similar methods can improve the predictions of continuous variables like SOCS and clay content.

In this study, we present a novel framework to evaluate and compare different Sentinel-2 SRC for producing large-scale soil property maps. Based on high-quality SOCS and clay content samples from the German Agricultural Soil Inventory, we assessed the predictive power of each SRC using band-wise regressions and support vector machines (SVM). Rather than only considering soil samples that are covered by the SRC, we quantified potential mapping bias by comparing the SOCS and clay values between samples that have been included and excluded/masked out from the soil maps. To evaluate the reliability of the resulting predictions, we further analyzed the extent to which the models reproduced the full range and distribution of measured soil properties. In summary, the following three research questions were addressed:

- How does the predictive power of the SRC for mapping SOCS and clay depend on the choice of bare soil indices and thresholding values?

- How do interactions between the bare soil indices and soil properties influence the exclusion of soil samples and the potential introduction of bias during SRC generation?
- To what extent is it possible to further improve the distribution and range of SOCS and clay predictions using additional EO features and data balancing methods?

2 Material and methods

2.1 Research Area

Germany covers a land area of 357,592 km^2, with roughly 36% dedicated to cropland (Destatis, 2022). The soil conditions are diverse and can be divided into seven major landscape regions that follow a general North-South gradient: the Northern Plain, Loess Hills, Eastern and Western Highlands, Southern Scarplands, Alpine Foreland, and the Alps (Figure 8.1). The North German Plain was formed by glacial processes and is mostly flat with sandy soils. In the northwestern part of the plain, organic soils are common in lowlands and are often drained for agricultural purposes. The Loess Hills in Central East Germany are covered by aeolian sediments, resulting in the presence of fertile loamy soils. The Eastern and Western Highlands feature multiple low mountain ranges below 1,500 m a.s.l. with diverse soils depending on various bedrocks. In the South German Scarplands, changing sediments from the Devonian and Triassic induced the formation of diverse textures, including clay-rich soils. The Alpine Foreland is influenced by glaciofluvial deposits from the Tertiary Molasse Basin and primarily consists of loamy soils. Organic soils are found in the pre-Alpine lowlands that are fed by mountain rivers. Towards the Alps in the south, the altitude increases, and cropland areas decrease and are substituted by grassland.

2.2 Soil Samples

The soil samples in this study were collected as part of the first German Agricultural Soil Inventory (BZE-LW), conducted from 2011 to 2018 (Poeplau et al., 2020a). A total of 2,254 cropland soils were sampled based on an 8 × 8 km regular grid across Germany (Figure 8.2). At each site, a one-meter soil profile was excavated and a composite sample (around 1 kg) was collected from five fixed

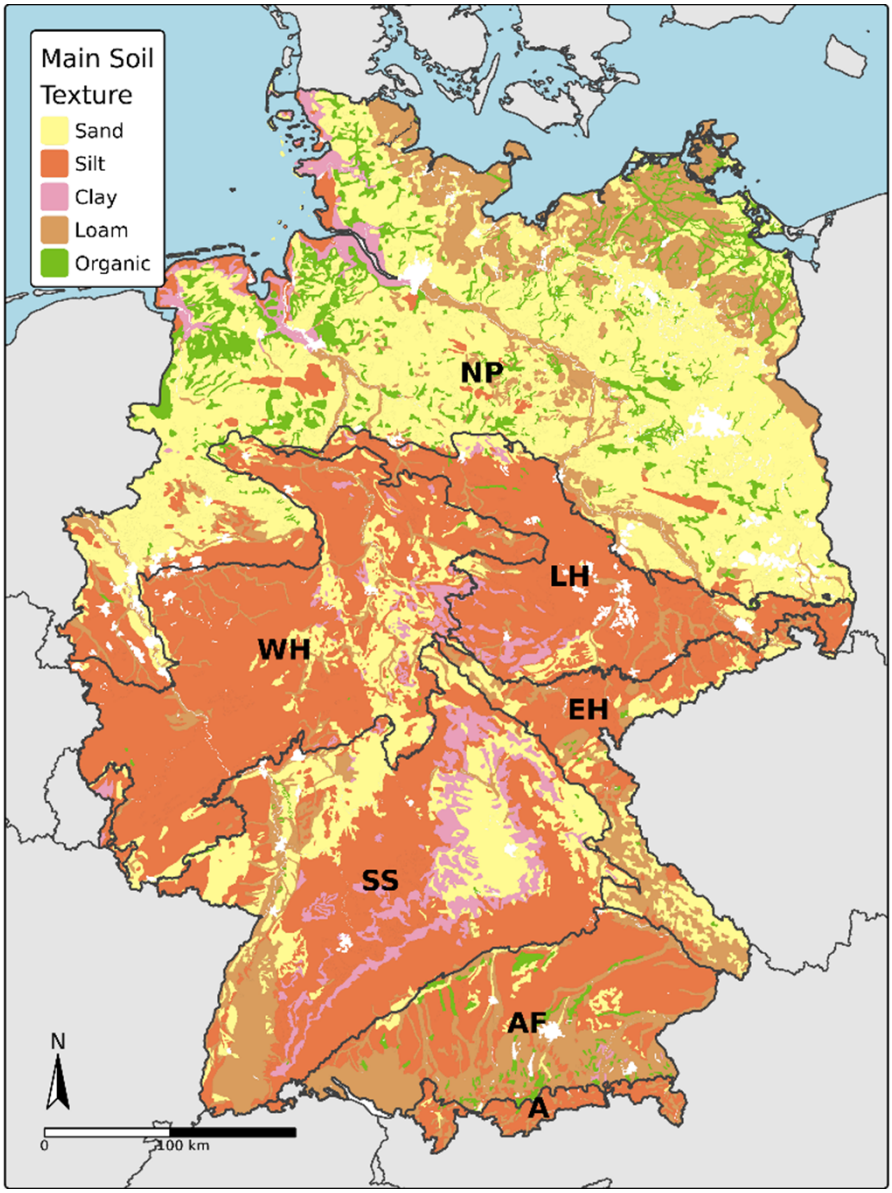

Figure 8.1: Overview of the research area. Main soil texture (BGR, 2007a) and major landscape regions in Germany (Thünen-Institut, 2022). NP = Northern Plain; LH = Loess Hills; WH = Western Highlands; EH = Eastern Highlands; SS = Southern Scraplands; AF = Alpine Foreland; A =Alps. Map lines delineate study areas and do not necessarily depict accepted national boundaries.

depth increments for chemical and textural analysis: 0–10 cm, 10–30 cm, 30–50 cm, 50–70 cm, and 70–100 cm. Additionally, two to ten undisturbed samples were taken to determine the bulk density and rock fragment fraction. The SOCS were estimated according to methods described by Poeplau et al. (2017). First, the composite samples were oven-dried at 40 °C and sieved to 2 mm to determine the SOC content of the fine soil using dry combustion. Next, the fine soil stocks (FSS in Mg ha^{-1}) were estimated for each depth increment (i) using the undisturbed samples

$$FSS_i = \frac{\text{mass}_{\text{fine soil}}}{\text{volume}_{\text{sample}}} \times \text{depth}_i \tag{1}$$

and the SOCS (Mg ha^{-1}) were calculated using

$$SOCS_i = SOC_{\text{fine soil}} \times FSS_i \tag{2}$$

To determine the total topsoil SOCS of each cropland site in 0–30 cm, the measurements of the depth increments 0–10 and 10–20 cm were combined. The clay content (mass%) of the fine soil was determined for each depth increment using sieving and sedimentation techniques, based on the dried composite samples. The total clay content of the topsoil (0–30 cm) was estimated using a weighted average of the depth increments 0–10 and 10–20 cm. An overview of the measured SOCS and clay distributions is given in Figure 8.2.

2.3 Sentinel-2 Data Preparation

The remote sensing data used in this study was processed and organized using the Framework for Operational Radiometric Correction for Environmental Monitoring (FORCE) (Frantz, 2019). All available Sentinel-2A/B scenes from February to November (2015–2024) were collected and organized into a data cube using the FORCE level-1 module. The scenes from December and January were excluded to minimize interference from snow cover and water logging, which can lead to a distortion of the soil signal. To prepare an analysis-ready data cube, the data was preprocessed with the FORCE level-2 module, including cloud masking, radiometric corrections, co-registration, and resolution merging (Frantz, 2019).

Cloud and cloud shadow detections were carried out using a modified version

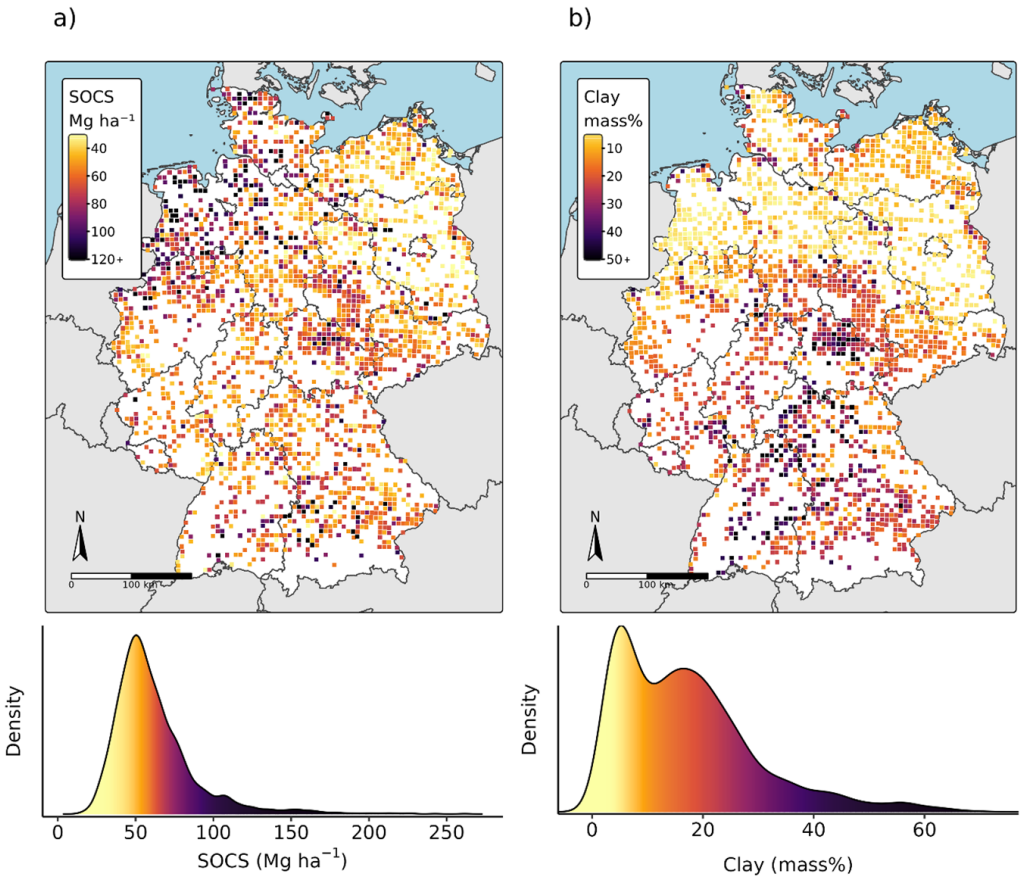

Figure 8.2: Overview of the cropland sampling locations and the distributions of the measured SOCS (a) and clay content (b) in Germany.

of the Fmask algorithm (Frantz et al., 2015; Zhu and Woodcock, 2012). Based on parallax effects between the Sentinel-2 near-infrared bands, a cloud displacement index was implemented to distinguish clouds from land features (Frantz et al., 2018). A 300 m buffer was applied around clouds to remove hazy transition zones, and pixels identified as opaque clouds were marked out to prevent spectral distortions unrelated to the land surface. All Sentinel-2 scenes with an estimated cloud cover below 70 % were considered for the data cube.

Radiometric corrections were used to convert the top-of-atmosphere reflectance through radiative transfer modeling, as described in Frantz et al. (2016b). Adjustments were used to account for atmospheric conditions like aerosol optical depth and water vapor, using the 1 arc-second Copernicus Digital Elevation Model for accurate path length corrections. Additionally, corrections for adjacency effects, bidirectional reflectance distribution function (BRDF), and topography were applied. An enhanced C-correction method was used to reduce illumination artifacts in sloped areas (Buchner et al., 2020).

To address geolocation errors in the pre-Global Reference Image (GRI) Sentinel-2 data, the scenes were co-registered with Landsat base images to improve geometric accuracy (Rufin et al., 2021). A data fusion algorithm (ImproPhe) was applied to increase the spatial resolution of the 20-meter Sentinel-2 bands (5, 6, 7, 8a, 11, 12) to a common resolution of 10 meters (Frantz et al., 2016b). The processed data was projected to the ETRS89-LAEA coordinate system (EPSG:3035) and used for the generation of the SRC in the following steps.

2.4 Soil Reflectance Composites

2.4.1 Bare Soil Indices and Thresholds

Using the preprocessed Sentinel-2 data cube, band-wise spectral averages of all available bare soil observations were calculated and used as primary model features for soil mapping. Four SRC variants were generated based on different combinations of bare soil indices (Table 8.1) and thresholding values (Table 8.2). Due to various effects (fallow land, land-use change, mixed pixels, etc.), it is expected that the SRC will be unable to cover all 2,234 soil samples. To ensure comparability of the results, the indices and thresholds were defined to cover a

similar proportion of around 90% of the total soil samples (Table 8.2).

The NDVI and NBR2 are the most used bare soil indices with recommended thresholding values typically ranging from 0.25 to 0.35 (NDVI) and 0.05 to 0.16 (NBR2) (Delaney et al., 2025). Three combinations were tested to evaluate how lowering both thresholds (SRC1: Low NDVI, NBR2) or only one of them (SRC2: Low NDVI; SRC3: Low NBR2) affects the resulting SRC (Table 8.2). Thresholds with NBR2 values below 0.1 were avoided, as they would significantly reduce the spatial coverage of the SRC. An additional SRC was tested using a modified vegetation index (PV+IR2), proposed by Heiden et al. (2022) (Table 8.1). The corresponding threshold values (SRC4: Low PV+IR2) were selected to maintain a comparable number of training samples (~2000) across the four SRC variants (Table 8.2).

Table 8.1: Overview of the bare soil indices considered in this study (band names refer to Sentinel-2).

Index Name	Abb.	Equation	Source
Normalized Difference Vegetation Index	NDVI	$\frac{B8_{soil}-B4_{soil}}{B8_{soil}+B4_{soil}}$	Tucker (1979)
Normalized Burned Ratio 2	NBR2	$\frac{B11_{soil}-B12_{soil}}{B11_{soil}+B12_{soil}}$	Van Deventer et al. (1997)
Combined NDVI and SWIR2 index	PV+IR2	$\frac{B8_{soil}-B4_{soil}}{B8_{soil}+B4_{soil}} + \frac{B8_{soil}-B12_{soil}}{B8_{soil}+B12_{soil}}$	Heiden et al. (2022)

Table 8.2: Overview of the four SRC variants, derived in this study.

SRC No.	Name	Threshold 1	Threshold 2	Included Samples (% from total)
1	Low NBR2, NDVI	NBR2 < 0.1	NDVI < 0.25	2003 (90%)
2	Low NDVI	NBR2 < 0.16	NDVI < 0.25	2025 (91%)
3	Low NBR2	NBR2 < 0.1	NDVI < 0.35	2028 (91%)
4	Low PV+IR2	NBR2 < 0.16	PV+IR2 < 0.24	2038 (91%)

2.4.2 Evaluation and Comparison of the SRC

The evaluation of the resulting SRC was conducted in multiple steps to assess (1) the predictive power of individual spectral bands for modeling cropland SOCS and clay content in Germany; and (2) the potential introduction of model bias due to

the selective masking of specific soil samples. In the first step, linear regressions and R^2 values were derived for the ten SRC bands and the respective SOCS and clay samples. Next, support vector machines (SVM) (Cortes and Vapnik, 1995) were used to model both soil properties using all SRC bands as input features. For the SVM, a radial kernel was applied with default settings (epsilon = 0.1; cost = 1) without additional parameter tuning.

Since the SRC were designed to cover roughly the same proportion of total soil measurements, we were able to compare the potential bias introduced by selective sample masking. To quantify the results, we analyzed how closely the probability distributions of included ($\approx 90\%$) and excluded ($\approx 10\%$) soil samples aligned for each SRC using the Wasserstein distance (W_p) (Vaserstein, 1969). The W_p evaluates the costs of converting one distribution into another one, providing an estimation of the model bias that is introduced by excluding certain soil properties. To simplify the comparison of the SOCS and clay, the W_p values were normalized by dividing them by the standard deviation of the corresponding soil property. To further examine relationships between the bare soil indices and soil samples included in the four SRC, scatter plots with linear regressions were plotted for NDVI, NBR2, and PV+IR2, and the corresponding SOCS and clay measurements.

2.5 Additional Covariates and Feature Importance

To assess whether the model performance could be further improved, we tested the inclusion of additional EO-based features alongside the ten original Sentinel-2 bands (Table 8.3). In the first step, spectral indices were derived from the SRC bands to enhance the spectral information of soil models. To test the influence of various spectral indices on the predictions, we used the standardized catalog provided by Montero et al. (2023), referencing more than 200 entries. Based on the catalog, a forward features selection (Meyer et al., 2019) was conducted to reduce the number of features and to identify the three most important indices for each model (Table 8.3). Next, band-wise spectral temporal metrics (STM) were generated by deriving the mean reflectance of the full Sentinel-2 time series without applying SRC thresholds. To reduce the number of features and avoid overfitting, only the bands with the highest correlations to SOCS (RED-EDGE1) and clay (SWIR2) were considered for the corresponding models (see

section 2.4.2). In the last step, a 10-meter digital elevation model (DEM) was incorporated as the only non-optical EO covariate (BKG, 2024). All model features share a common resolution of 10×10 meters (Table 8.3).

To assess the influence of the additional covariates and original SRC bands on model results, the feature importance was analyzed according to the permutation method described by Fisher et al. (2019). The permutation feature importance works by randomly shuffling the values of specific features while keeping the remaining features unchanged. Based on the modified datasets, predictions are carried out using a pre-trained model, and the feature importance is estimated as the difference in model performance before and after shuffling.

Table 8.3: Overview of the Sentinel-2 SRC bands and additional features used for the SOCS and clay models.

Name	Abb.	Equation	Model	Reference
SRC Bands				
Blue (492 nm)	BLUE	$B2_{\text{soil}}$	SOCS,Clay	-
Green (560 nm)	GREEN	$B3_{\text{soil}}$	SOCS,Clay	-
Red (665 nm)	RED	$B4_{\text{soil}}$	SOCS,Clay	-
Red edge (704 nm)	RE1	$B5_{\text{soil}}$	SOCS,Clay	-
Red edge (741 nm)	RE2	$B6_{\text{soil}}$	SOCS,Clay	-
Red edge (783 nm)	RE3	$B7_{\text{soil}}$	SOCS,Clay	-
Broad NIR (833 nm)	BNIR	$B8_{\text{soil}}$	SOCS,Clay	-
NIR (865 nm)	NIR	$B8a_{\text{soil}}$	SOCS,Clay	-
SWIR1 (1614 nm)	SWIR1	$B11_{\text{soil}}$	SOCS,Clay	-
SWIR2 (2202 nm)	SWIR2	$B12_{\text{soil}}$	SOCS,Clay	-
SRC Indices				
Normalized Difference Yellowness Index	NDYI	$\frac{B3_{\text{soil}}-B2_{\text{soil}}}{B3_{\text{soil}}+B2_{\text{soil}}}$	SOCS	Sulik and Long (2016)
Modified Bare Soil Index	MBI	$\frac{B11_{\text{soil}}-B12_{\text{soil}}-B8_{\text{soil}}}{B11_{\text{soil}}-B12_{\text{soil}}-B8_{\text{soil}}+0.5}$	SOCS	Nguyen et al. (2021)
Excess Red Index	ExR	$1.3 \cdot B4_{\text{soil}} - B3_{\text{soil}}$	SOCS	Meyer et al. (1999)
Normalized Burn Ratio 2	NBR2	$\frac{B11_{\text{soil}}-B12_{\text{soil}}}{B11_{\text{soil}}+B12_{\text{soil}}}$	Clay	Van Deventer et al. (1997)
Simple Ratio (555/750 nm)	SR555	$\frac{B6_{\text{soil}}}{B3_{\text{soil}}}$	Clay	Gitelson and Merzlyak (1994)

Name	Abb.	Equation	Model	Reference
Blue Normalized Difference Vegetation Index	BNDVI	$\frac{B8_{\text{soil}}-B2_{\text{soil}}}{B8_{\text{soil}}+B2_{\text{soil}}}$	Clay	Wang et al. (2007)
STM				
Vegetation red edge (704 nm)	STM_RE1	$B5_{\text{all}}$	SOCS	-
SWIR2 (2202 nm)	STM_SWIR2	$B12_{\text{all}}$	Clay	-
DEM				
10 Meter Elevation Model	DEM	–	SOCS,Clay	BKG (2024)

2.6 Training Data Balancing

We developed a framework to evaluate whether data balancing affects model performance and improves the alignment between the range of predicted and observed soil properties. Like the 'Synthetic Minority Over-sampling Technique' (SMOTE) (Chawla et al., 2002), our proposed method balances the training data by combining minority over-sampling with majority under-sampling, following a three-step process (Figure 8.3). First, the total number of training samples is extended by extracting the input features from a 3 by 3 pixel grid, instead of only using one pixel from the sample coordinates. This step is comparable to the data augmentation in SMOTE but utilizes the spectral information of the adjacent pixels instead of generating synthetic training data. Next, the balancing is conducted by assigning sampling weights to each data point based on an inverted histogram (Figure 8.3) to ensure that underrepresented soil properties are more likely to be included in the training data. Nine random subsamples with different sizes between 10 and 90% of the total training samples are drawn to reflect a trade-off between strong balancing with a small sample size and weak balancing with a large sample size (Figure 8.3). To identify the optimal balancing strategy for the prediction of SOCS and clay, each of the nine training sets (10-90%) was used to conduct a separate SVM (see section 2.4.2). In the final step, the results were evaluated in terms of model accuracy and prediction fit, using the original, imbalanced sampling data (Figure 8.3).

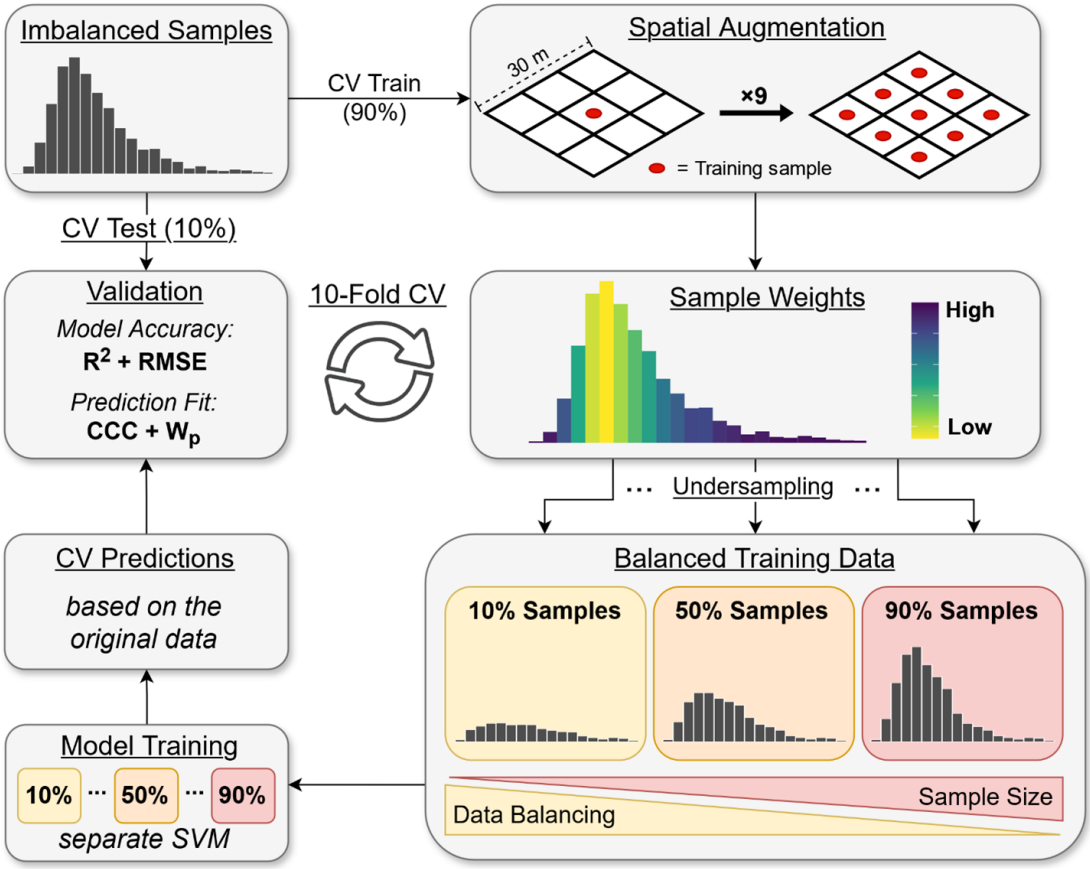

Figure 8.3: Flowchart of the proposed data balancing framework, tested for the prediction of SOCS and clay. The yellow and red colors reflect the trade-off between data balancing and sample size. RMSE = root mean square error; CCC = Lin's concordance correlation coefficient; W_p = Wasserstein distance.

2.7 Accuracy Assessment

The accuracy assessment for the prediction of SOCS and clay was conducted based on a 10-fold cross-validation. As shown in Figure 8.3, the validation folds were always excluded from the training samples before data augmentation, ensuring that no sites were used for both training and validation. Based on the ten cross-validation folds, the linear regression (R^2) and root mean square error (RMSE) were calculated to quantify the overall accuracy of the predictions. In addition, Lin's concordance correlation coefficient (CCC) was derived to provide further details about the agreement between the measured and predicted values (Lin, 1989). In contrast to the R^2, the CCC also takes into account the position of the regression line, assigning higher values to predictions that are closer to the optimal 1:1 line. In the final step, the W_p was derived to quantify the distance between the distributions of measurements and predictions (see section 2.4.2). Low W_p values indicate that the model was able to accurately reproduce the distribution of the measured SOC Sand clay samples, while higher W_p values hint at the presence of over- or underrepresentations in the predictions.

2.8 Final Predictions and Uncertainty Maps

Final maps were produced for SOCS and clay content using the best-performing SRC (see section 2.4.2), the optimized data balancing (see section 2.6), and the covariates described in Table 8.3. For both soil properties, SVM were trained using a radial kernel and the following hyperparameters: epsilon = 0.1; cost = 1. Spatial predictions were carried out using the R packages mlr3 (Lang et al., 2019), mlr3spatial (Becker and Schratz, 2024), and terra (Hijmans, 2025). In addition to the final predictions, uncertainty maps were generated based on the cross-validation models. For each fold, a separate map was generated using the methods described above, and the pixel-wise uncertainty was derived using the radius of the 95 % prediction intervals (i.e., standard deviation $\times$ 1.96).

3 Results

3.1 Predictive Power of the SRC

In general, strong relationships were found between all four generated SRC and the observed soil properties (Figure 8.4 a). For SOCS, the strongest correlations occur in the RED-NIR spectrum ($R^2 > 0.4$), while moderate relationships have been found across all bands. In contrast, the clay content shows a much more selective dependence, which is limited to the SWIR spectrum ($R^2 > 0.3$). This observed trend was consistent across all SRC, however, the magnitude of the relationships varies significantly. SRC1 and SRC2, which were both generated using low NDVI thresholds (0.25), show a noticeably lower correlation to SOCS in comparison to the other SRC (Figure 8.4). Similarly, SRC1 and SRC3 were generated using low NBR2 thresholds (0.1) and showed slightly weaker correlations to clay in the SWIR2. The SRC4, based on the PV+IR2 (<0.24), consistently showed the highest correlations with SOCS across all bands, while SRC4 and SRC2 produced similar R^2 values with clay in the SWIR2 region (Figure 8.4 a). This overall tendency is supported by the results of the SVM, which were trained using all SRC bands (Figure 8.4 a). In general, SRC4 achieved the highest model accuracy for SOCS (CCC = 0.74), followed by SRC3. In contrast, SRC1 and SRC2 show a significantly lower performance with CCC values around 0.65. For the prediction of clay, the observed patterns are slightly different. While SRC4 still produced the most accurate model (CCC = 0.82), SRC2 showed a slightly higher accuracy compared to the remaining SRC.

3.2 Dependencies between Soil Properties and Bare Soil Indices

Across the generated SRC, strong positive correlations were observed between NBR2 and clay content (Figure 8.5 a). The highest relations were found for SRC4 ($R^2 = 0.54$) and SRC2 ($R^2 = 0.47$) which both utilize high NBR2 thresholds (0.16), as illustrated by the dashed lines in Figure 8.5 a. In contrast, the low NBR2 thresholds (<0.1) used for the generation of SRC1 and SRC3 significantly reduced the range of NBR2 values, strongly decreasing the observed correlations ($R^2 = 0.44$ and 0.41 respectively). A similar, but slightly weaker relationship was found between NDVI and SOCS (Figure 8.5 d). In this case, the strongest corre-

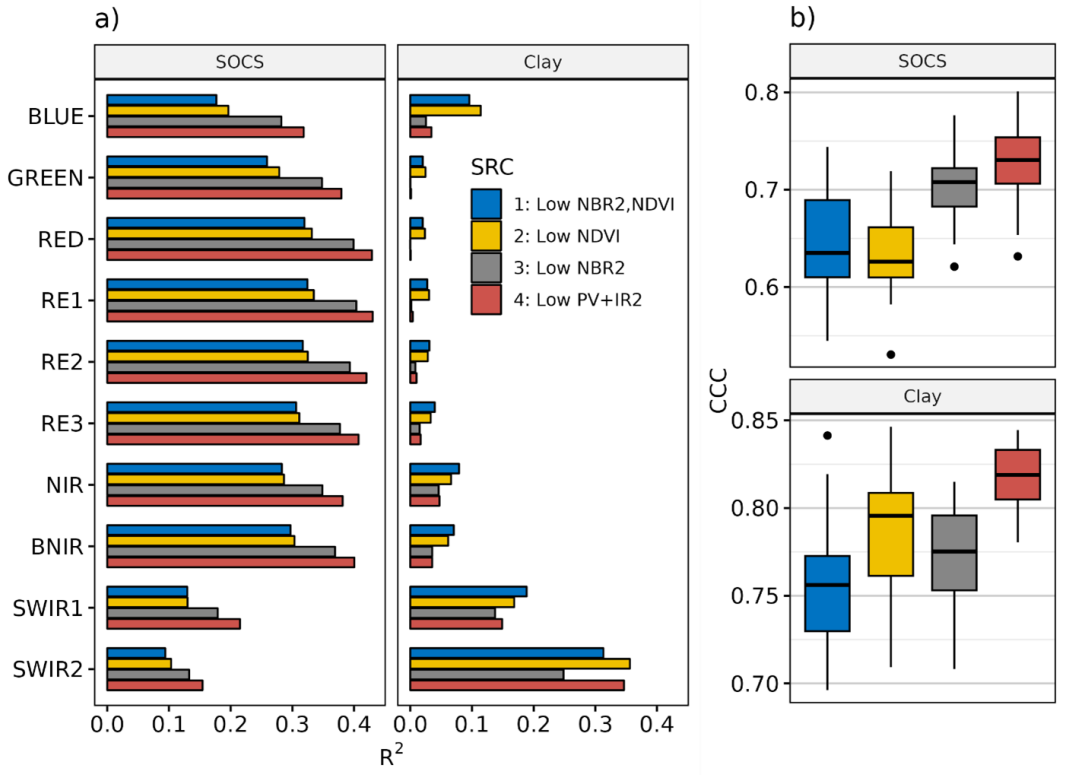

Figure 8.4: a) Band-wise regressions (R^2) between the four generated Sentinel-2 SRC and the corresponding SOCS and clay samples. b) Model accuracy (CCC) of each SRC for the prediction of SOCS and clay, based on the ten Sentinel-2 bands.

lation was observed for SRC4 ($R^2 = 0.29$), followed by SRC3 ($R^2 = 0.23$). Again, the SRC utilizing low NDVI thresholds (SRC1 and SRC2) generally showed a restricted range and a lower correlation of around 0.14. In addition to these strong correlations, a local dependency was found between NDVI and clay contents below 20%, which is most pronounced in SRC3 and SRC4 (Figure 8.5 c). Weak relationships ($R^2 < 0.14$) were found between clay and PV+IR2, but are limited to SRC2 and SRC4 (Figure 8.5 e). Across all four observed SRC, no significant correlations were observed between NBR2, PV+IR2, and SOCS (Figure 8.5 b+f).

To identify potential spatial dependencies and patterns in the results, the three observed bare soil indices were calculated for the whole research area using the results from SRC4 (Figure 8.5). For the NDVI, a strong spatial clustering is visible, with the highest values being present in Northwest Germany. The NBR2 shows a clear gradient across the research areas, with low values in the north that are increasing towards the south. In comparison, the PV+IR2 generally shows a much more homogeneous distribution across the research area, although slight spatial variability is visible, especially in floodplains with elevated soil moisture levels.

3.3 Comparison of Included and Excluded Samples

In addition to quantifying the predictive power of each SRC, we evaluated the distributions of the included and excluded samples to identify potential bias (Figure 8.7). For SOCS, the highest agreements between the values above and below the SRC thresholds were found in SRC4 ($W_p = 0.3$) and SRC3 ($W_p = 0.44$). In contrast, the SRC1 and 2 showed much lower agreements, with W_p values above 0.8. In these cases, high SOCS values were excluded more frequently, altering the distributions when compared to the soil samples.

Significant differences were also observed in the distributions of the clay content (Figure 8.7). Again, the highest agreement between included and excluded samples was found for SRC4 ($W_p = 0.17$), suggesting the lowest bias in the representation of the clay content. In contrast, the distributions show that SRC3 disproportionately excluded samples with high clay values, which significantly increased the W_p to 0.9. Similarly, SRC2 led to the exclusion of samples with low clay values ($W_p = 0.45$), while SRC1 showed a combination of both exclusion patterns ($W_p = 0.52$).

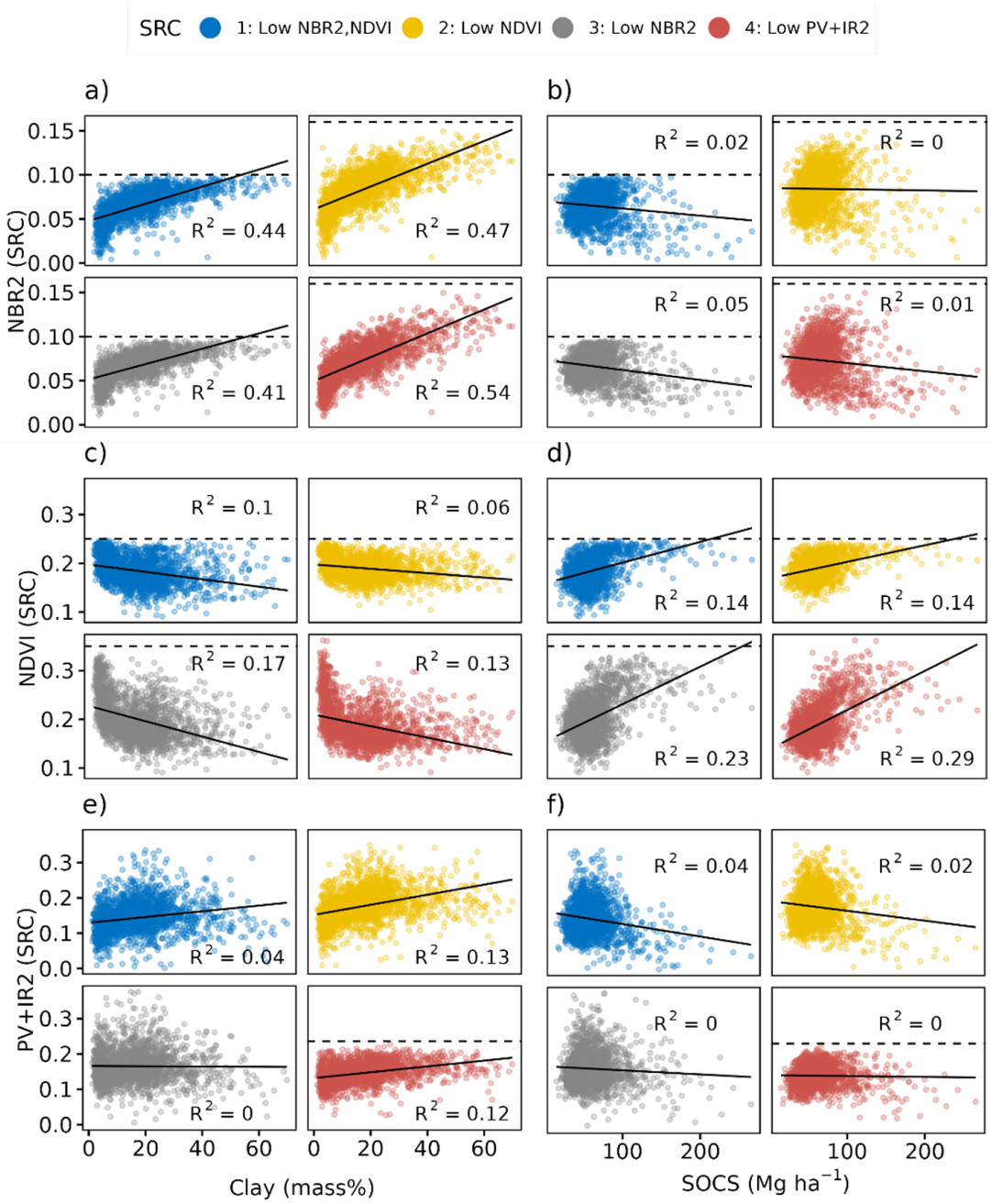

Figure 8.5: Regressions (R^2) between the bare soil indices (NBR2, NDVI, and PV+IR2) derived from the SRC bands, and the observed soil properties (Clay and SOCS). Black lines = linear regressions; Dashed lines indicate SRC thresholds.

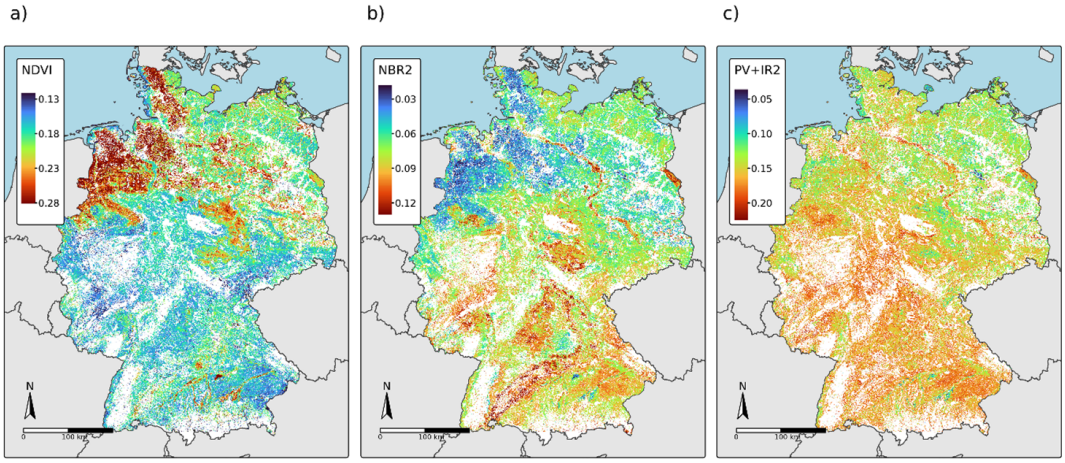

Figure 8.6: Maps of the bare soil NDVI (a), NBR2 (b), and PV+IR2 (c) in Germany, based on the results from SRC4.

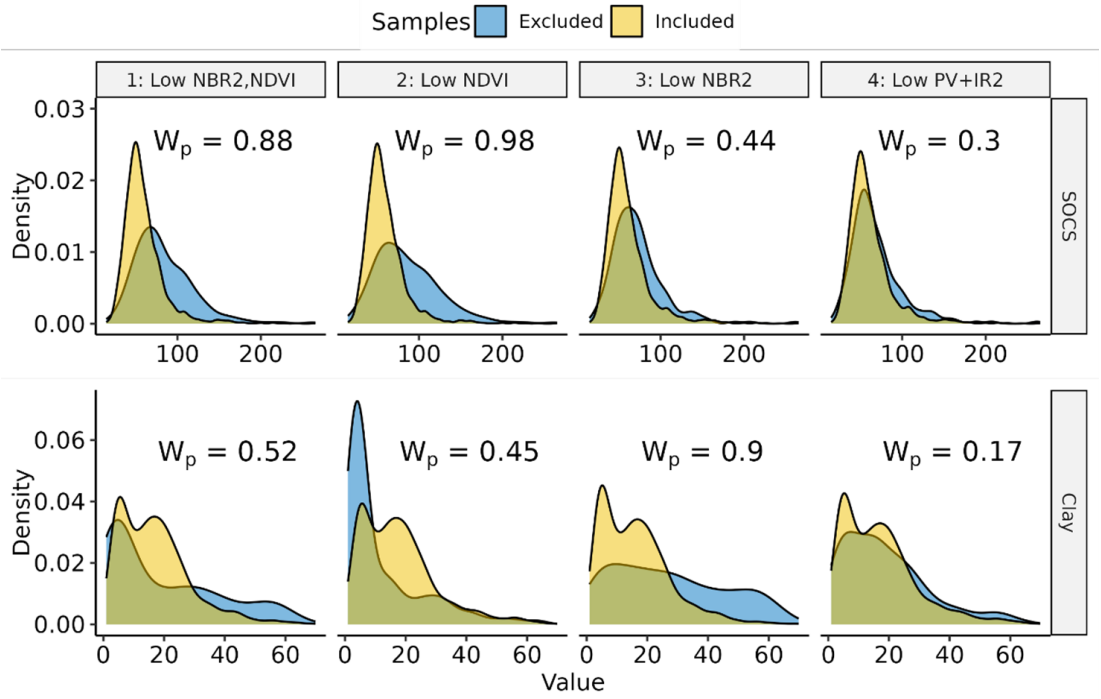

Figure 8.7: Density plots of the SOCS/clay samples that were included (yellow) and excluded (blue) by the corresponding SRC. For each SRC, the proportion of soil samples below the thresholds is around 90% (see Table 3).

3.4 Model Performance and Influence of Additional Features

Based on the fact that SRC4 (PV+IR2 < 0.24 & NBR2 < 0.16) showed the highest predictive power (Figure 8.4 b) and lowest W_p values/bias (Figure 8.7), it was used for further analyses of the model performance. Comparing the results of the baseline models, trained on the SRC bands only, the prediction of clay (Figure 8.8 a) achieved slightly higher accuracy than SOCS (Figure 8.9 a) (R^2 = 0.72 and 0.61, respectively). As illustrated by the regression lines, the SOCS model showed a slightly stronger tendency to underpredict high values, which is reflected by the CCC values (SOCS = 0.74; clay = 0.82). In general, the observed SOCS distribution of the baseline model was significantly narrower compared to the measurements, resulting in a relatively high W_p of 0.19 (Figure 8.8 a). Similarly, extreme values were underrepresented in predictions of the clay model, shifting the distribution toward the mean and increasing the W_p (0.15) (Figure 8.9 a).

Including additional EO-based features significantly increased the prediction accuracy for both SOCS (CCC = 0.78) and clay (CCC = 0.88) (Figure 8.8 b & Figure 8.9 b). In both cases, these enhancements were also visible in the distributions of predicted values. In the clay model, the W_p strongly improved from 0.15 to 0.09, aligning the predictions more closely with the measured value range. While the addition of covariates also slightly reduced the W_p of the SOCS predictions (0.19 to 0.16), the models' limitation to reproduce the full range distribution of the measurements remained.

The analysis of the permutation feature importance revealed that the SRC bands and indices were by far the most influential covariates for the prediction of clay, with a relative RMSE loss of around 150% (Figure 8.10). While the SRC features (bands + indices) were also the most influential covariates in the SOCS model, the observed magnitude of the permutation performance was much smaller, with values below 50%. Unlike the SRC, the inclusion of STM showed moderate importance for the prediction of SOCS, while it had no impact on the clay model. In both cases, a similar level of importance was assigned to the DEM, though its influence was significantly weaker compared to features based on the SRC.

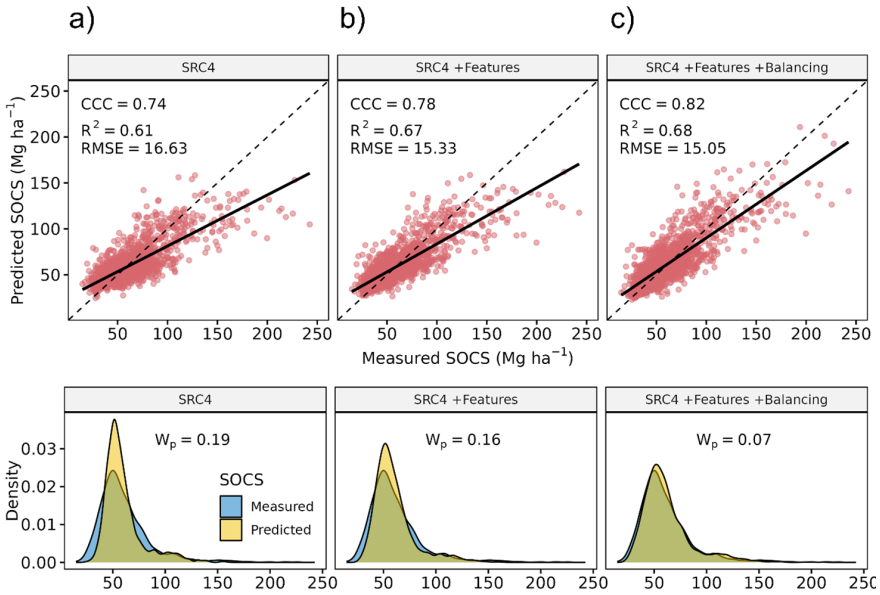

Figure 8.8: (Top) Comparison of the SOCS model performance using the SRC bands only (a), the SRC bands in combination with additional features (b), and the SRC bands with features and optimal data balancing (c). Black lines = linear regressions; dashed lines = 1:1 line. (Bottom) Corresponding density plots and W_p values of the measured (blue) and predicted (yellow) SOCS values.

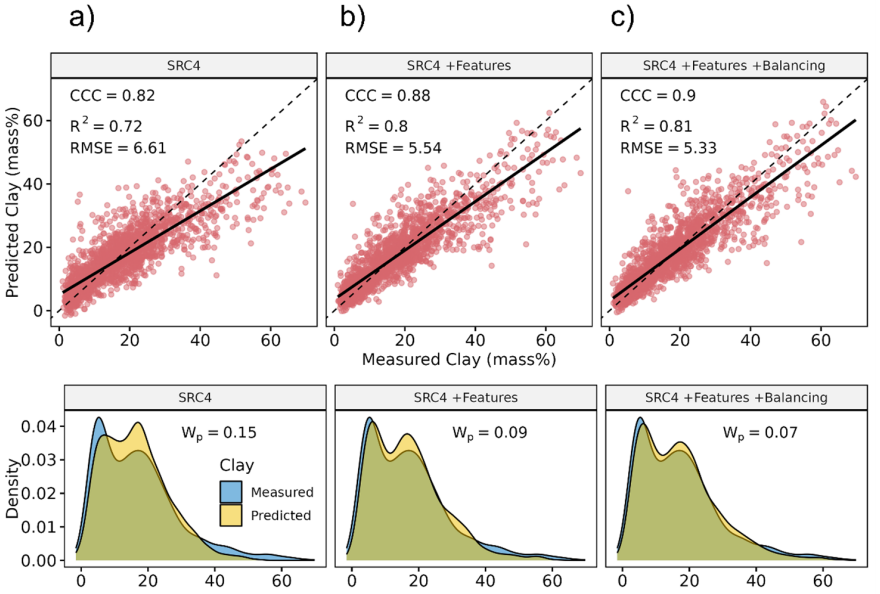

Figure 8.9: (Top) Comparison of the clay model performance using the SRC bands only (a), the SRC bands in combination with additional features (b), and the SRC bands with features and optimal data balancing (c). Black lines = linear regressions; dashed lines = 1:1 line. (Bottom) Corresponding density plots and W_p values of the measured (blue) and predicted (yellow) clay values.

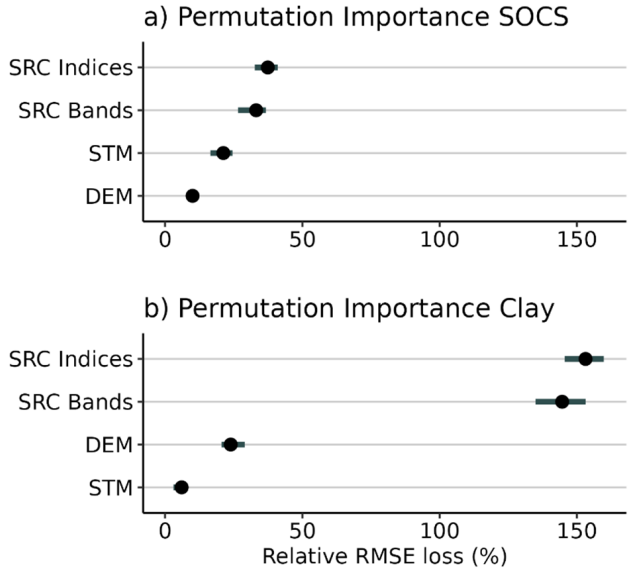

Figure 8.10: Relative feature importance (permutation loss) of the covariate groups (Table 4) for the prediction of SOCS (a) and clay (b). STM = Spectral temporal metrics; DEM = Digital elevation model.

3.5 Effects of the Data Balancing on the Model Performance

We expanded the training dataset using a spatial augmentation approach (Figure 8.3) to evaluate the influence of data balancing on the model performance. For the prediction of SOCS, a clear trade-off emerged between training data size and balancing (Figure 8.11 a). Models based on strong balancing (i.e., using only 10% of the samples) showed significantly higher RMSE values, indicating a tendency to overpredict SOCS. With increasing sample size, the overall accuracy improved, reaching an optimum when using intermediate balancing and around 30% of the training samples. While the models based on the weak balancing (i.e., 90%) produced similar R^2 and RMSE values, the performance of the CCC and W_p was significantly lower. A similar trend was observed for the clay model, though the optimal trade-off between data size and balancing was higher at around 60% (Figure 8.11 b).

The final predictions based on the optimal data balancing for SOCS (30%) and clay (60%) are shown in Figure 8.8 c & Figure 8.9 b. In both models, the data augmentation and balancing increased the CCC from 0.78 to 0.82 (SOCS) and 0.88 to 0.9 (clay), shifting the regressions closer to the optimal 1:1 line.

In addition to the CCC, minor improvements were also observed in the RMSE and R^2 values. While the data balancing also improved the distributions of the predictions in both cases, the reduction of the W_p was more significant for the SOCS (W_p = 0.16 to 0.07) than for clay (W_p = 0.09 to 0.07).

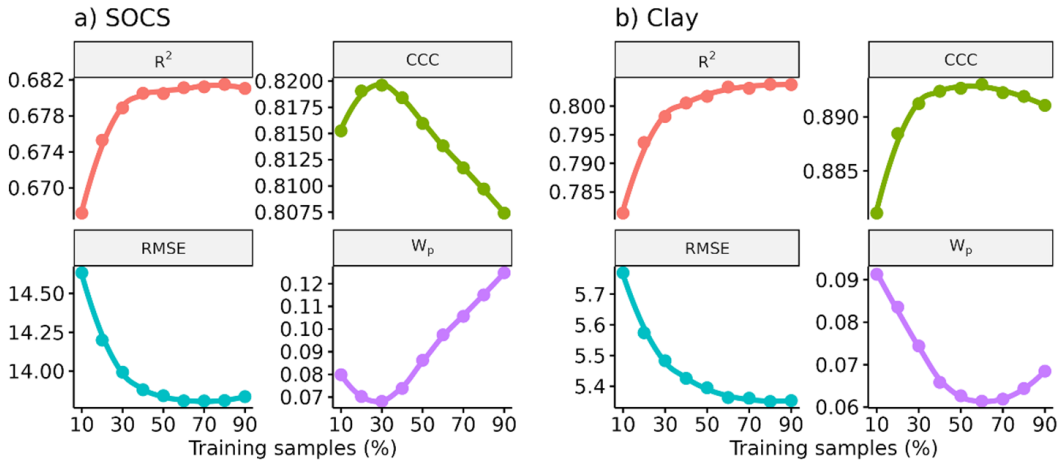

Figure 8.11: Influence of data balancing (i.e., percent of training samples included) on the model accuracies (R^2, RMSE, CCC, W_p) for the prediction of SOCS (a) and clay (b). Small sample size = strong balancing; Large sample size = weak balancing (see Figure 8.3).

3.6 Final Predictions and Uncertainty Maps

Based on the final models, predictions were carried out to map cropland SOCS (Figure 8.12) and clay (Figure 8.13) in Germany. The Northern Plain can be divided into two regions, exhibiting soils with low SOCS and clay in the east, while the western parts predominantly show soils with high SOCS and low clay. Regions with high SOCS and clay are mostly found in the Loess Hills and the Southern Scarplands. The Western Highlands are characterized by soils with low SOCS and medium to high clay, while regions with medium and high SOCS and clay are mostly present in the Eastern Highlands and Alpine Foreland.

To provide further information on the predictions of SOCS and clay, uncertainty maps were generated based on cross-validation models (Figure 8.14). For the former, the highest uncertainties were present in Northwest and South Germany, corresponding to the regions with the highest SOCS predictions (Figure 8.14 a). In addition, increased uncertainties were also found in parts of Central and Southern Germany, which are characterized by high clay contents

(Figure 8.13). In contrast, the uncertainty map for clay (Figure 8.14 b) is similar to the overall prediction map and is not influenced by the presence of SOCS. In this case, increased uncertainties are connected to floodplains and similar regions with elevated levels of soil moisture.

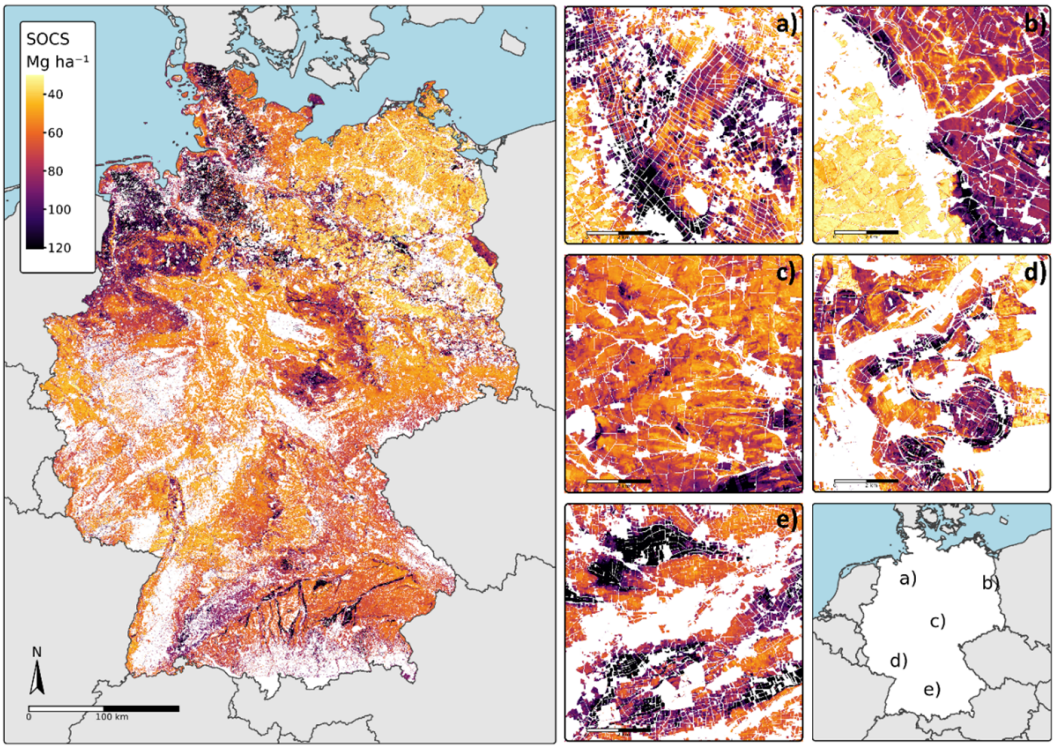

Figure 8.12: Final prediction map of the cropland SOCS (0 – 30 cm) in Germany.

4 Discussion

4.1 Influence of Bare Soil Indices on the SRC

Our results indicate that the NDVI and NBR2, which are often used to generate SRC (Delaney et al., 2025), are strongly related to the bare soil reflectance and the presence of SOCS and clay. The relationship between NBR2 and clay content is significantly more pronounced ($R^2 = 0.54$) compared to the underlying SRC Bands (Figure 8.4 & Figure 8.5). These findings align with previous EO-based studies and soil spectroscopy, showing that SWIR bands are particularly sensitive to the presence of clay minerals (Chabrillat et al., 2019; Shabou et al., 2015; Wetterlind et al., 2025). Similarly, the observed correlation between NDVI and

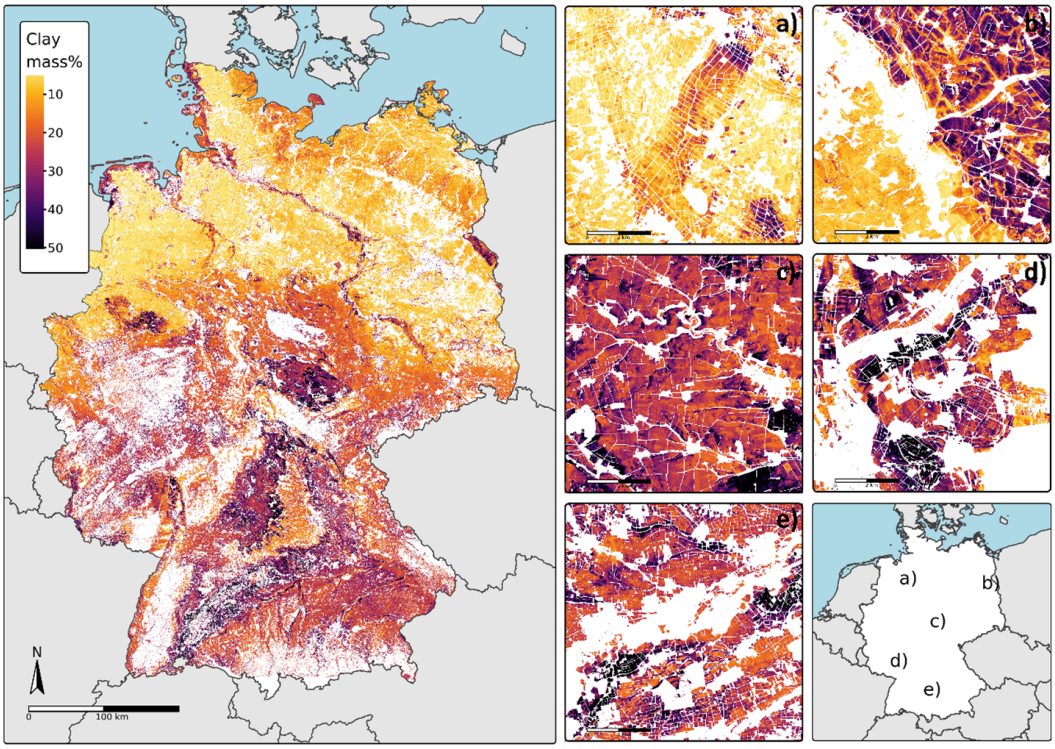

Figure 8.13: Final prediction map of the cropland clay content (0 – 30 cm) in Germany.

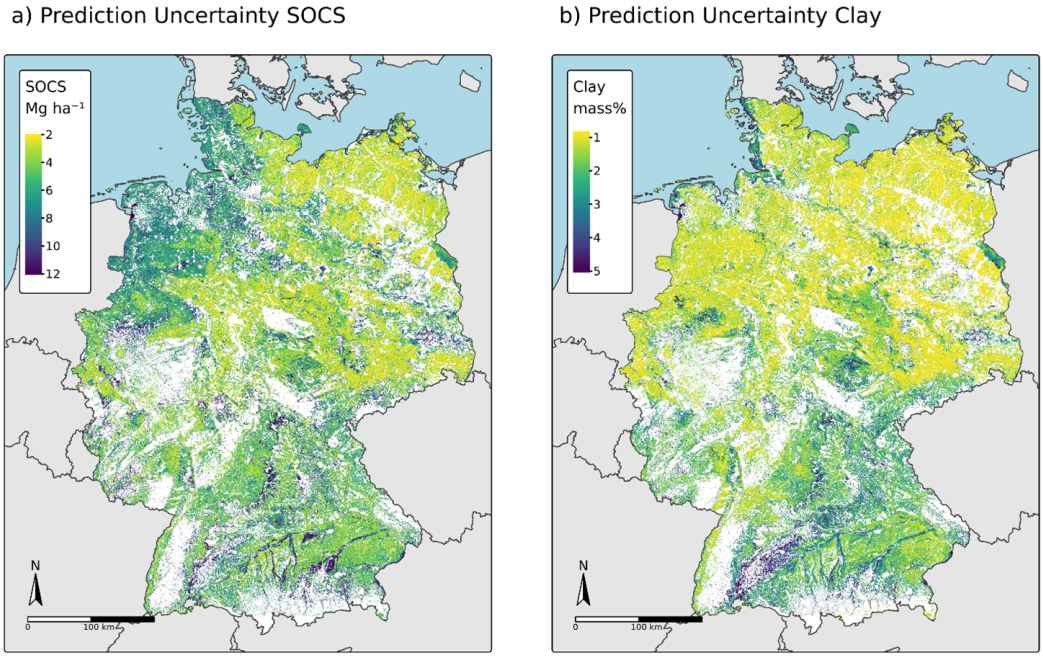

Figure 8.14: Prediction uncertainty for SOCS (a) and clay (b), based on the cross-validation models.

SOC ($R^2 = 0.29$) can be explained by studies showing that "dark" (i.e., carbon-rich) soil substrates produce higher NDVI values than "bright" soils, given the same presence of vegetation (Huete, 1988).

Directly comparing different parameter combinations, we were able to demonstrate that low NBR2 (0.1) and NDVI (0.25) thresholds significantly affect the distributions of the soil samples that are included and excluded by the SRC. As illustrated in Figure 8.7, low NBR2 thresholds strongly favor the selection of samples with lower clay content (SRC1, SRC3), while a similar effect was observed for low NDVI thresholds and high SOCS values (SRC1, SRC2). The corresponding W_p values underline that the implementation of such thresholds to generate SRC introduces systematic bias by excluding certain soil parameters from the resulting maps (Figure 8.7). These dependencies are also reflected in the spatial distribution of NDVI and NBR2 (Figure 8.6) showing that both indices are clustered in soil regions with high SOCS or clay values.

Given these clear dependencies, it is surprising that many studies still rely on low NDVI or NBR2 thresholds when generating SRC for soil mapping applications (Castaldi et al., 2019a; Demattê et al., 2025; Silvero et al., 2021). One reason is that these dependencies are overlooked when focusing on individual soil properties to optimize the SRC and resulting maps. No direct correlation was found between NBR2 and SOCS (Figure 8.5) and it would be much more difficult to detect the underlying bias when only comparing SOCS distributions of the included and excluded samples (e.g. in SRC3 in Figure 8.7). In addition, a biased selection of soil samples for training can also enhance the apparent model performance in some cases, making it less obvious and harder to detect. This is supported by several studies, reporting that the SOC prediction accuracy increased when using very low NBR2 thresholds to generate SRC, while also significantly reducing the number of training samples (Castaldi et al., 2019a; Dvorakova et al., 2022; Vaudour et al., 2021). Our findings suggest, however, that this effect is strongly influenced by the systematic exclusion of soil samples with high clay contents, which are known to interfere with the prediction of SOC (Stenberg, 2010). This effect is supported by Figure 8.14 a, highlighting that the highest prediction uncertainties for SOCS are located in regions with clayey soils.

To minimize the bias in the final soil maps, it is recommended to compare the distributions of multiple soil properties, even the ones that are not included

in the models (Figure 8.7). In general, the scatter plots in Figure 8.5 show the high NDVI or NBR2 thresholds exclude fewer samples and are less likely to introduce bias for soil modeling. If applying low thresholds to reduce the influence of vegetation cover instead, it is strongly suggested to rely on bare soil indices that are not affected by the soil reflectance itself. Heiden et al. (2022) found that while the NBR2 provided the highest SRC quality, the PV+IR2 was better suited to differentiate bare soil from vegetated soil and provided the best SRC masking. This is supported by our findings, showing that the PV+IR2 is affected less by the observed soil properties, preserving the original distributions of the SOCS and clay samples to a high degree (Figure 8.7).

4.2 Influence of the SRC on the Prediction of SOCS and Clay

Despite being recommended in the literature (Demattê et al., 2020), our findings show that SRC1 based on low NDVI (<0.25) and NBR2 (<0.1) thresholds yielded the lowest overall model performance to predict SOCS and clay (Figure 8.4). Instead, significantly higher accuracies were achieved using SRC4 (PV+IR2 $<$ 0.24 & NBR2 $<$ 0.16), highlighting the importance of the SRC settings on the model results. Analyzing the dependencies between the SRC bands and soil properties, we were able to attribute these findings to two key factors. As noted by several studies, the NDVI lacks sufficient spectral information to effectively differentiate bare soil from vegetation cover, especially in the presence of crop residues (Castaldi et al., 2023; Demattê et al., 2018). It has been shown that in spectrally mixed regions (bare soil, vegetation, residues), the NDVI produced the lowest performance out of the observed bare soil indices and should, therefore, not be used to generate SRC for soil mapping (Heiden et al., 2022). Consequently, our findings underline that the SRC based on low NDVI thresholds (SRC1 and SRC2) exhibited significantly weaker correlations with SOC across all bands, suggesting a lower spectral quality.

In contrast, it has been suggested that NBR2 is highly sensitive toward the presence of non-photosynthetic vegetation and effective in detecting optimal bare soil observations (Dvorakova et al., 2022). Our results indicate, however, that the positive influence of low NBR2 thresholds on the prediction of SOC does not apply under all circumstances. When using the same NBR2 threshold (0.1)

across different SRC, we found that increasing the NDVI threshold from 0.25 (SRC1) to 0.35 (SRC3) significantly improved the accuracy of the SOCS model (Figure 8.4). This improvement can be explained by the systematic exclusion of high-SOCS samples in SRC1 (>80 Mg h^{-1}), restricting the range of SOCS values and weakening the overall model regression (Figure 8.5 & Figure 8.7). As noted in the previous section, it is questionable if the positive influence of low NBR2 thresholds on SOCS prediction truly results from an improved selection of bare soil observations or is caused by the systematic exclusion of clayey soils. This is supported by the fact that the improved performance of SRC3 (NBR2 <0.1 & NDVI <0.35) did not extend to the clay model and is strongly influenced by the removal of high-clay samples above 30 mass% (Figure 8.4 & Figure 8.7). In general, our findings show that SRC4 yielded the highest accuracy for both models, which is supported by the results of Heiden et al. (2022) (Figure 8.4). Considering the high dependence between the bare soil indices and soil properties (Figure 8.7), however, we were able to demonstrate that a direct comparison of different SRC is only valid when also taking into account the samples that have been excluded by the thresholds.

4.3 Model Performance and Feature Importance

The baseline models, trained using the ten SRC bands only, produced good accuracies for the prediction of SOCS and clay, which are comparable to other large-scale soil property maps (Chen et al., 2022). Similar to other studies based on SRC (Silvero et al., 2021; Stumpf et al., 2024), the model performance was slightly better for clay than for SOCS, which can be explained by several factors. First, it is well-documented that clay minerals influence the reflectance in the SWIR region (Chabrillat et al., 2019; Wetterlind et al., 2025), which is supported by the strong correlation between NBR2 (i.e, the normalized SWIR-ratio) and clay (Figure 8.5). Even though SOC affects a much broader spectrum (Figure 8.4), the correlation is slightly lower, reducing the predictive power of the SRC compared to clay. A simple explanation for this tendency is the fact that clay usually makes up a much larger portion of the total soil mass, increasing its influence on soil reflectance as compared to SOC. The inclusion of further EO-based covariates significantly improved the performance of both models, however, for different reasons (Figure 8.8 & Figure 8.9). While both models profited from additional indices based on SRC, the improvement was much more distinct for

the prediction of clay (Figure 8.9). This is in line with the strong influence of clay on the soil reflectance and is supported by the high magnitude of the SRC bands and indices in the permutation feature importance (Figure 8.10). In contrast, our findings show that the inclusion of STM only improved the SOCS model but did not affect the prediction of clay (Figure 8.10). As illustrated in Figure 8.5, the "darkness" of the soil can influence vegetation indices, allowing the SOCS signal to persist in STM to some degree even when derived from the complete Sentinel-2 time series (Huete, 1988). The same does not apply to the spectral signal of clay in the SWIR range, which is strongly affected by vegetation cover and and lost when deriving STM. Additional research is necessary to uncover what kind of additional information is provided by the STM to improve the SOCS prediction.

The addition of a 10-meter DEM improved both models to a similar degree, which is supported by the results of other studies (Stumpf et al., 2024; Urbina-Salazar et al., 2023). It has been shown that soil properties like SOC and clay are influenced by geomorphology and that landscape features such as elevation can play in key role in DSM (Behrens et al., 2018a). Given the strong north-south gradient in elevation and spatial clustering of soils in the research area (Figure 8.1), it is possible that the DEM also serves as a more general spatial proxy describing the positions of the sampling locations within the research area (Hengl et al., 2018; Møller et al., 2020).

4.4 Influence of Sample Balancing and Additional Covariates

Despite achieving good results, both baseline models showed tendencies to underpredict high SOCS and high clay values (Figure 8.8 & Figure 8.9). Similar findings have been reported for the prediction of SOC based on SRC (Zepp et al., 2021a) and are often attributed to imbalanced training data and the underrepresentation of soil samples with extreme values (Feeney et al., 2022). Comparing the measured and predicted SOCS and clay in Figure 8.8 & Figure 8.9, it is visible that the baseline models often predicted values close to the mean and were unable to reproduce the full range of the input data. As highlighted by Feeney et al. (2022), this tendency can be attributed to two main factors: (1) the ability of the spatial covariates to capture the full range of soil parameters, and (2) the distribution and representation of the training samples. Our results demonstrate

that both of these factors can be improved by providing further EO-based input features (Table 8.3) and implementing data balancing frameworks (Figure 8.3), but depend on the corresponding soil parameters. While the addition of STM, DEM, and spectral indices significantly increased the overall prediction accuracy in both cases, the improvement of the W_p was more significant for clay (Figure 8.8 & Figure 8.9). As shown in Figure 8.2, the measured SOCS is highly imbalanced with values that are concentrated around 50±20 Mg ha^{-1} and few outliers above 100 Mg ha^{-1}. Our findings demonstrate that in this case, data balancing can be used to further improve the prediction fit (i.e., CCC) and the model's ability to reproduce the range and distribution of the soil samples (i.e., W_p). As highlighted in Figure 8.11, however, the positive influence of our proposed method (Figure 8.3) strongly depends on identifying the optimal trade-off between sample size and data balancing. In general, it can be concluded that soil properties with outliers and highly skewed distributions also require a stronger balancing (SOCS = 30%) compared to properties with more even distributions (clay = 60%) (Figure 8.11). In both cases, the model improvements are most significant in the W_p and CCC, increasing the range and representation of the soil samples in the final prediction maps.

5 Conclusion

In this study, we used Sentinel-2 bare soil maps and soil monitoring data to generate national maps of cropland SOCS and clay with very high accuracy (CCC = 0.82 and 0.90, respectively). The analysis aimed to understand the impact of different bare soil indices and threshold values on the resulting SRC and soil models. Additionally, we explored whether data balancing and incorporating additional EO-based features could further improve prediction accuracy, especially when using imbalanced data sets. Our main findings demonstrate that:

- While varying SRC parameters can lead to similar model outcomes, interactions between the bare soil indices and soil parameters can significantly affect the prediction accuracy and bias of the final maps.
- Strong correlations were found between the NBR2 and clay (R^2 = 0.54) and NDVI and SOCS (R^2 = 0.29), resulting in the systematic exclusion of specific soil values when applying low thresholds to generate SRC.
- Instead of relying on NBR2 or NDVI, prioritizing bare indices that are less

correlated with soil properties (e.g., PV+IR2) improves model accuracy while maintaining the distributions of the soil samples.

- While the baseline models tended to underpredict high values and failed to capture the full range of the soil samples, the addition of further EO-based features (i.e., spectral indices, STM, DEM) significantly improved prediction accuracy (R^2, RMSE) and the alignment between the measurements and predictions (CCC, W_p).
- Data augmentation and balancing were successfully used to further improve the CCC and W_p and increase the range of the predictions, but require identifying the optimal trade-off between the size and balance of the training data.
- Based on our findings, we recommend extending the accuracy assessment of SRC-based soil property maps by (1) taking into account the excluded samples to identify potential model bias, and (2) comparing the range and distribution of the predictions and soil samples to increase the applicability and reliability of the final soil maps.

Acknowledgements

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

Chapter 9

Manuscript 3: Transferability of covariates to predict soil organic carbon in cropland soils

Tom Broeg, Michael Blaschek, Steffen Seitz, Ruhollah Taghizadeh-Mehrjardi, Simone Zepp, Thomas Scholten
Remote Sensing, 2023

https://doi.org/10.3390/rs15040876

Received: 19 December 2022; Revised: 20 January 2023; Accepted: 31 January 2023; Published: 4 February 2023

Abstract

Precise knowledge about the soil organic carbon (SOC) content in cropland soils is one requirement to design and execute effective climate and food policies. In digital soil mapping (DSM), machine learning algorithms are used to predict soil properties from covariates derived from traditional soil mapping, digital elevation models, land use, and Earth observation (EO). However, such DSM models are trained for a specific dataset and region and have so far only allowed limited general statements to be made that would enable the models to be transferred to different regions. In this study, we test the transferability of SOC models for cropland soils using five different covariate groups: multispectral soil reflectance composites (satellite), soil legacy data (soil), digital elevation model derivatives (terrain), climate parameters (climate), and combined models (combined). The transferability was analyzed using data from two federal states in southern Germany: Bavaria and Baden-Wuerttemberg. First, baseline models were trained for each state with combined models performing best in both cases ($R^2 = 0.68/0.48$). Next, the models were transferred and tested with soil samples from the other state whose data were not used during model calibration. Only satellite and combined models were transferable, but accuracy declined in both cases. In the final step, models were trained with samples from both states (mixed-data models) and applied to each state separately. This process significantly improved the accuracies of satellite, terrain, and combined models, while it showed no effect on climate models and decreased the models based on soil covariates. The experiment underlines the importance of EO for the transfer and extrapolation of DSM models.

1 Introduction

Soil organic carbon (SOC) is not only a large sink for atmospheric CO_2 but also one of the most important factors of soil quality and soil fertility (Lal, 2016). Since the beginning of human land-use activities, most intensely used soils lost a large fraction of their natural SOC stock (Lal, 2003). This is especially true for cropland soils which depend on good management to keep a stable carbon equilibrium (Janzen, 2006). In environmental policy, this deficit is commonly portrayed as an opportunity for synergetic policymaking; to use the ability of

soils to store organic carbon to sequester CO_2 and enhance soil quality and food security (Smith et al., 2010). To improve our understanding of cropland SOC stocks and support soil management, precise maps with a high spatial resolution are needed (van Wesemael et al., 2021).

The framework of digital soil mapping (DSM) is based on the idea that soil properties are represented by environmental covariates and can be mapped with soil spatial prediction functions (SSPF) (Mcbratney et al., 2003; Taghizadeh-Mehrjardi et al., 2022b). The basic DSM framework has been updated to account for modern machine learning (ML) algorithms (Hengl et al., 2018) and was successfully used to predict SOC over large areas (Fathizad et al., 2022; Taghizadeh-Mehrjardi et al., 2020). Along with terrain attributes, Earth observation (EO) data are important assets of DSM and multispectral images of bare soil are often used to predict SOC (Behrens and Scholten, 2006b; Emadi et al., 2020; Rentschler et al., 2022). Most times of the year, cropland is densely covered with vegetation and the soils of different areas are rarely exposed at the same time. To solve this problem, multi-temporal EO data are analyzed to detect bare soil areas over multiple years (Demattê et al., 2020; Diek et al., 2017; Rogge et al., 2018; Safanelli et al., 2020). Spectral indices such as the Normalized Difference Vegetation Index (NDVI) , the Bare Soil Index (BSI) (Diek et al., 2017), or the Normalized Burn Ratio 2 (NBR2) (Vaudour et al., 2021) are calculated to differentiate between covered and exposed soil and the reflectance of each pixel is averaged over a defined period. The resulting soil reflectance composite maps can be used as covariates to model SOC and other soil properties in cropland soils (Demattê et al., 2020; Möller et al., 2022; Vaudour et al., 2021; Zepp et al., 2021a).

Ground truth is necessary to train and validate DSM models but soil samples can be scarce in some regions (Maleki et al., 2020), limited in field operability (Stumpf et al., 2016), and analysis is expensive and time-consuming (Gehl and Rice, 2007). Efforts have been made to reduce the number of necessary ground measurements by extrapolating or transferring DSM models from donor areas that have a higher availability of soil data (de Arruda et al., 2016; Machado et al., 2019; Malone et al., 2016; McKay et al., 2010; Neyestani et al., 2021; Taghizadeh-Mehrjardi et al., 2022a; Wolski et al., 2017; Zhao et al., 2020). In most cases, models of soil types or classes were transferred with mixed overall accuracies between 46% (Neyestani et al., 2021) and 83% (Machado et al., 2019). One

study was able to transfer a SOC model with a relative overall accuracy of 70% (Zhao et al., 2020). The models were either based on expert knowledge (McKay et al., 2010), regressions (Malone et al., 2016), ML algorithms such as artificial neural networks (ANN) (de Arruda et al., 2016), or random forest (RF) (Machado et al., 2019). The choice of method is important as some models, such as RF, are unable to extrapolate values outside the range of the training data (Hengl, 2021). However, a direct comparison between different model types showed that tree-based algorithms such as RF had the highest accuracies in extended modeling areas (Neyestani et al., 2021). Since RF is one of the most common methods for DSM (Hengl et al., 2018), this study aims to define preconditions to successfully transfer RF models to new areas.

Covariates from various sources such as remote sensing (Neyestani et al., 2021), ground-based detection (Malone et al., 2016), digital elevation models (DEM) (Machado et al., 2019), and soil legacy data (Zhao et al., 2020) have been used to transfer machine DSM models. It is difficult to compare covariates between different studies as ML algorithms are sensitive to the context of the training dataset which often reflects a specific research area (Taghizadeh-Mehrjardi et al., 2020; Zeraatpisheh et al., 2022). Furthermore, covariates reflect different data sources, and some covariates groups likely express a higher degree of transferability while others should only be used in a local context (Zhao et al., 2020).

In this study, the transferability of SOC models based on multispectral soil reflectance composites (satellite), soil legacy data (soil), DEM derivatives (terrain), climate parameters (climate), as well as combined models (combined) were tested. This was done by comparing models in the two southern German federal states of Bavaria and Baden-Wuerttemberg, with a combined area of about 100,000 km^2. The degree of transferability was analyzed in three steps: (i) First, a baseline was set by testing the accuracy of each covariate group for the prediction of SOC in both states. (ii) These models were then transferred and tested with the soil samples of the respective other state to analyze how the performance changes if the models are applied to a new area. (iii) In the last step, mixed data models with different proportions of training data from both states were built to evaluate the effect of additional samples from outside the validation area. This was repeated for each covariate group and both states.

2 Materials and methods

2.1 Study Area

The study area includes the federal states of Baden-Wuerttemberg (BW) and Bavaria (BY), in southern Germany (Figure 9.1 A,B). Summary statistics of both states are provided in Table 9.1. The climate is temperate and humid with mean temperatures of about 9 °C and precipitation above 800 mm. Southern Germany is a diverse landscape and both states share most geologic formations and parent materials of pedogenesis (BGR, 2021). Most of the area of BW and the northern parts of BY is located in the Southwest German Scarplands which are characterized by fast-changing strata of the Triassic and Jurassic. BW is characterized by its distinct geological features, namely the Rhine Valley and the Black Forest in the west followed by the Swabian Alps, ranging towards the Bavarian border in the east (Figure 9.1 A). The southeast of BW and Southern BY is covered by the North Alpine Foreland Basin which is filled with sediments of the Alps and was further formed by the glacial activities of the Pleistocene. Luvisols and cambisols are the most frequent soil types while groundwater-influenced and carbon-rich gleysols and fluvisols are connected to streams originating from the Alps (BGR, 2020). Cropland is the predominant land use type throughout the research area, excluding the mountainous areas. The models in this study were limited to cropland without permanent vegetation and regularly exposed soils (Rogge et al., 2018) (Figure 9.1 C).

Table 9.1: Comparison of both federal states.

	Bavaria	Baden-Wuerttemberg
Area (km²)	71,000	36,000
Area of cropland soils (km²)	34,000	11,000
Proportion of cropland to total area (%)	48	31
Mean temperature (°C)	8.7	10
Mean precipitation (mm)	836	818
Predominant soil type	Cambisol	Luvisol

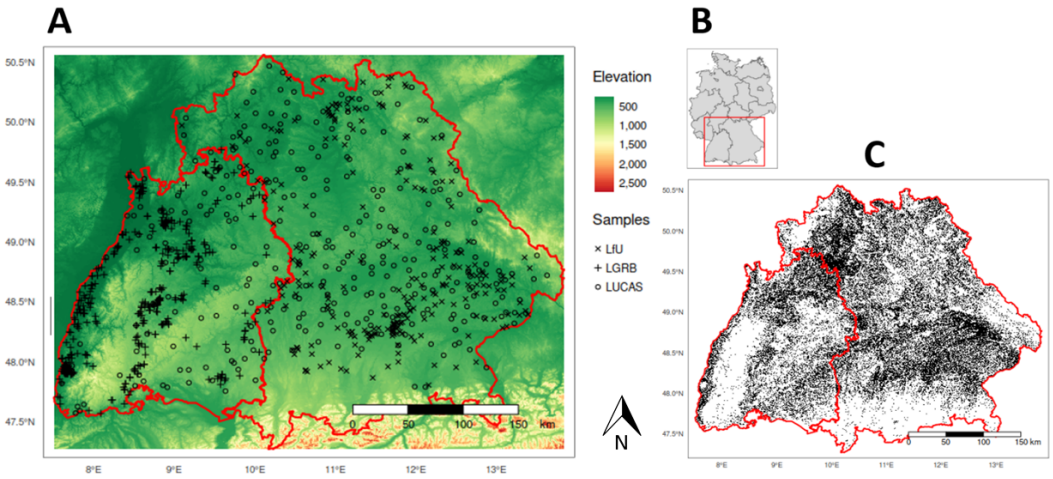

Figure 9.1: Study area: (A) Sampling locations in Bavaria (east) and Baden-Wuerttemberg (west). (LUCAS = The Land Use and Cover Area Frame Survey, LfU = Bavarian Environment Agency, LGRB = State Authority for Geology, Resources and Mining; (B) Location within Germany; (C) Mask of the exposed soils (black) in the soil reflectance composite.

2.2 Procedures

The DSM framework is based on the idea that unknown soil properties can be represented by environmental factors (covariates). The covariate layers should be selected to meet all soil forming factors (e.g., climate and topography) and to represent the complex processes behind the soil attribute of interest. Then, a quantitative relationship between measured soil samples and relevant covariates is used to predict values over large areas and to create soil maps. In the current study, we followed the DSM approach for spatial prediction of SOC in the two states of Germany. The procedure included (i) acquiring the SOC data, (ii) calculating covariates, (iii) model training, and (iv) SOC prediction. The four main components of DSM are discussed in the following sections.

2.3 Soil Samples

The SOC content of cropland topsoil (0–30 cm) was used to train and evaluate the DSM models. A total of 950 soil samples (475 per state) from three different sources were included (Table 9.2; Figure 9.1 A). The Land Use and Cover Area

Frame Survey (LUCAS) was started by the Statistical Office of the EU in 2006 and sampled the soils of over 20.000 individual locations on EU ground (Jones et al., 2020). Two sampling campaigns were carried out in 2009 and 2015 on a subset of a 2 × 2 km regular grid. The LUCAS 2015 dataset includes 837 samples of cropland soils in Germany. If a sampling location was visited in 2009 and 2015, only the more recent measurements were considered. Mixed samples were taken from the topsoil of each location and the SOC was determined using dry combustion. In addition to the LUCAS dataset, measurements from soil legacy databases of the Bavarian Environment Agency (LfU) and the State Authority for Geology, Resources and Mining (LGRB) were included.

Table 9.2: Summary statistics of SOC measurements for each data source and both states

State	Source	Samples	SOC (%) Min	SOC (%) Max	SOC (%) Mean	SOC (%) SD	SOC (%) IQR
BY	LUCAS	227	0.6	14.81	2.1	1.68	1.15
	LfU	248	0.54	15.6	3.13	2.8	1.91
	Total	475	0.54	15.6	2.63	2.38	1.54
BW	LUCAS	91	0.79	5.78	1.78	0.87	0.99
	LGRB	384	0.45	8.31	1.48	0.93	0.88
	Total	475	0.45	8.31	1.74	0.92	0.90

These point data are allocated to different sampling campaigns (mapping activities, long-term field experiments, etc.) and therefore do not follow a consistent sampling design (Figure 9.1 A). Only samples taken within the analysis period of the exposed soil composite (Section 2.4.1) were considered (1984–2014). The databases contain descriptive information about the soil profiles, which were raised for general soil surveys, as well as further analytic information about the soil properties. As in the LUCAS surveys, the SOC was also analyzed using dry combustion. Only topsoil samples from cropland profiles were considered. This was verified using the mask of the exposed soil composite (Figure 9.1 C). Duplicate measurements from the same sampling locations were removed. All LUCAS samples within the soil reflectance composite were used for the models. As the number of samples in BW was limited, all available samples from the soil database were added to the training data (Table 9.2). The general availability of soil samples was higher in BY. This is true for samples of the LUCAS campaigns and the legacy data. The number of soil samples in BY was adjusted to the BW

dataset with a random selection of samples from the legacy data (Table 9.2) so that both states can be trained with the same sample size (475). This step was included to increase the comparability of the results and to simplify the sample exchange in mixed-data models (Section 2.8).

2.4 Covariates

The covariates were selected to meet the criteria of the SCORPAN factors (Mcbratney et al., 2003). An overview of the covariate groups and layers used in the models is provided in Table 9.3. The layers were resampled to a common grid cell size of 30 m (1'') using a cubic spline interpolation in GDAL (GDAL/OGR contributors, 2022). It has been shown that distance-based covariates can be included in RF models to account for the spatial autocorrelation of the soil samples (Behrens et al., 2018b; Hengl et al., 2018; Møller et al., 2020). This was done using seven Euclidean distance fields (EDF), as proposed by Behrens et al. (2018b): one for X and Y coordinates, each corner and the center of the research area.

2.4.1 Earth Observation

The Soil Composite Mapping Processor (SCMaP) was developed to automatically analyze multispectral data over a defined period and detect exposed soils to create per-pixel composites for area-wide mapping (Rogge et al., 2018). A detailed description of the method and processing steps of the SCMaP workflow is provided in the original papers (Rogge et al., 2018; Zepp et al., 2021b). For the bare soil reflectance composite of Germany, all available Landsat 4, 5, and 7 scenes from 1984–2014 were processed. The period of 30 years was analyzed for several reasons. First, it allows to include of a higher percentage of cropland areas, as some soils are rarely exposed. The larger area also increases the number of available soil samples that can be used for modeling. Secondly, a longer period also increases the number of bare soil observations and therefore reduces seasonal effects like moisture or crop residues. A modified vegetation index (PV) was used to distinguish between exposed soils (i.e., croplands) and other land cover classes, like urban areas, grassland, or forests:

$$PV = \frac{NIR - RED}{NIR + RED} + \frac{NIR - BLUE}{NIR + BLUE} \quad (1)$$

Thresholds of maximum (PV_{max}) and minimum (PV_{min}) index composites were derived to create an exposed soil mask (Zepp et al., 2021b) (Figure 9.1 C). The reflectance of the detected exposed soil pixels was averaged per band to create the soil reflectance composites. The used composite map consists of six spectral bands from 0.45 to 2.35 µm with a spatial resolution of 30 m. A description of the bands is provided in Table 9.3. In the following, the six Landsat bands are referred to as scmap.1–6. The generated exposed soil mask is provided in Figure 9.1 C.

2.4.2 Terrain Attributes

All terrain-based covariates were extracted from the ALOS World 3D-30m DEM, provided by the Earth Observation Research Center of the Japan Aerospace Exploration Agency (JAXA EORC, 2021). The System for Automated Geoscientific Analyses (SAGA-GIS) (Conrad et al., 2015) was used to calculate five terrain-related covariates: topographic wetness index (TWI); valley depth (vdepth); multiresolution index of valley bottom flatness (VBF); and negative and positive topographic openness (openn and openp), as proposed by Hengl and MacMillan (2019). The contextual spatial modeling (CSM) framework was applied to enhance the representation of different scales in the modeling process (Behrens et al., 2018a). To do so, the DEM was up-scaled six times (90, 180, 360, 720, and 1440 m) and the terrain-based covariates were recalculated at each step (e.g., dem_twi_90-1440). It has been shown that the explanatory power of the terrain covariates can be improved with this process, as some large-scale patterns are only visible if the resolution of the DEM is reduced by aggregation.

2.4.3 Climate Data

Climate data were downloaded from the Open-Data-Server of the German Meteorological Service (DWD) with a spatial resolution of 1 × 1 km (Deutscher Wetterdienst, 2022). The rasters are based on interpolated mean values of weather and soil measurements at climate stations in Germany. Detailed descriptions of the methods are provided online at the Open-Data-Server (Deutscher Wetterdienst, 2022). Climate parameters are calculated in 30-year intervals and are only updated every 10 years. The period of 1981–2010 was selected, as it is as close as possible to the calculation of the soil reflectance composite (1984–2014).

Most climate parameters are available with 12 rasters as monthly averages. To reduce redundancy, a spatial principal component analysis (PCA) was performed for each climate parameter with more than one raster, using the GSIF package (Hengl, 2020). The most important components of each climate parameter with a combined explained variance of over 85% were kept.

Table 9.3: Overview of all covariates. Sources: 1 = Rogge et al. (2018); 2 = BGR (2020); 3 = BGR (2007b), 4 = BGR (2001); 5 = JAXA EORC (2021); 6 = Conrad et al. (2015); 7 = Deutscher Wetterdienst (2022)

Group	Covariate	Original Resolution	Layers (n)	Abbreviation	Source
Satellite	SCMaP Band 1: blue (0.45–0.52 µm)	30 m	1	scmap.1	1
	SCMaP Band 2: green (0.52–0.60 µm)	30 m	1	scmap.2	1
	SCMaP Band 3: red (0.63–0.69 µm)	30 m	1	scmap.3	1
	SCMaP Band 4: NIR (0.77–0.90 µm)	30 m	1	scmap.4	1
	SCMaP Band 5: SWIR1 (1.55–1.75 µm)	30 m	1	scmap.5	1
	SCMaP Band 6: SWIR2 (2.09–2.35 µm)	30 m	1	scmap.6	1
Soil	Soil type	1:200,000	1	soil_type	2
	Soil texture	1:200,000	1	soil_texture	2
	Sand content	1:200,000	1	soil_texture_sand	2
	Silt content	1:200,000	1	soil_texture_silt	2
	Clay content	1:200,000	1	soil_texture_clay	2
	Parent material	1:5,000,000	1	soil_geology	3
	Geomorphographic class	1:1,000,000	1	soil_geomorphology	4
Terrain	Digital elevation model	30 m	1	dem_30	5
	Topographic wetness index	90–1440 m	5	dem_twi_90-1440	6
	Valley depth	90–1440 m	5	dem_vdepth_90-1440	6
	Multiresolution index of valley bottom flatness	90–1440 m	5	dem_vbf_90-1440	6
	Negative topographic openness	90–1440 m	5	dem_openn_90-1440	6
	Positive topographic openness	90–1440 m	5	dem_openp_90-1440	6
Climate	Multi-year means of air temperature (2 m)	1000 m	2	DWD_temp	7
	Multi-year means of precipitation	1000 m	2	DWD_prec	7
	Multi-year soil temperature at 5 cm depth in bare soil	1000 m	2	DWD_soil_temp	7
	Multi-year grids of soil moisture under grass and sandy loam	1000 m	1	DWD_soil_moist	7
	Multi-year mean of the number of frost days	1000 m	1	DWD_frost_days	7
	Multi-year mean of the number of hot days	1000 m	1	DWD_hot_days	7
	Multi-annual mean onset/ending of vegetation	1000 m	1	DWD_vegetation	7
	Multi-year mean of the annual climatic water balance	1000 m	1	DWD_water_balance	7
	Multi-year mean of the monthly drought index	1000 m	2	DWD_drought	7
	Multi-year mean of sunshine duration	1000 m	2	DWD_sunshine	7

Table 9.4: Soil types and mean SOC content at the sampling locations. Main soil types were extracted from the German Soil Map (BÜK200) (BGR, 2020) and converted to the units of the World Reference Data Base.

Soil Type		BY		BW	
German Soil Sys.	World Ref. Base	Samples	Mean SOC (%)	Samples	Mean SOC (%)
Vega	Fluvisols	26	2.69	66	1.67
Braunerde	Cambisols	217	1.64	70	2.10
Pelosol	Vertisols	26	1.96	31	2.72
Gley	Gleysols	53	3.44	22	2.66
Anmoorgley	Gleysols	37	8.76	1	3.57
Parabraunerde	Luvisols	35	1.77	201	1.37
Rendzina	Leptosols	14	3.14	24	2.23
Pararendzina	Regosols	44	2.77	52	1.64
Pseudogley	Planosols	15	2.02	1	1.16
Kolluvisol	Colluvic	8	1.34	7	1.47

2.4.4 Legacy Soil Maps

Legacy maps and databases of soils and geological properties are provided by the German Federal Agency of Geoscience and Resources (BGR). The most recent soil map covering Germany is based on a 1:200,000 scale (BÜK200) and is composed of 55 individual tiles that have been harmonized into a common format (BGR, 2020). The vector format map is based on expert knowledge and the German soil mapping standards KA4/KA5 (Ad-hoc-Arbeitsgruppe Boden, 2006). The polygons are connected to a database containing details about soil types and sub-types as well as information about the properties of the individual soil horizons. The main soil types of the German Soil System were extracted and transformed into a raster layer. It was verified that both states share the same soil classes, as this is necessary for the models to be transferable to the other state. For illustration, the soil types were converted into the system of the World Reference Base (Table 9.4). Carbon-rich bog soils were excluded from the modeling area with the BÜK200 to avoid biases arising from outliers with very high SOC contents. Information on topsoil texture was extracted from the first mineral horizon. Separate layers were created for the corresponding sand, silt, and clay shares using mean values of the soil texture classes (Ad-hoc-Arbeitsgruppe Boden, 2006). The Geological Map of Germany (1:5,000,000) (BGR, 2007b) and the Map of Geomorphographic Regions in Germany (1:1,000,000) (BGR, 2001) were used to extract information about geological properties.

2.5 Similarity Analysis

The spatial variation of the natural conditions was analyzed to evaluate if both states are similar enough to successfully transfer the DSM models. Multivariate Environmental Similarity Surfaces (MESS) were created using Maxent (Elith et al., 2011). The MESS is calculated by comparing the covariate values at the training locations with the values of the research area. The resulting map can be used to evaluate the area of applicability of the models and to distinguish regions where covariates are outside the range of the training data. Two uncertainty maps were created for the whole research area, one with the training data of each state. Based on the feature selection, the most important climate, terrain, and soil covariates were used to calculate the similarity of the natural conditions. This approach is similar to the study by Meyer and Pebesma (2021) but does not assign weights according to the feature importance. Additionally, density plots were created to compare the values of the most important spectral bands at the sampling locations of both states. This process was repeated for other important numeric covariates from the climate and terrain group.

2.6 Random Forest

RF models were used as spatial prediction models. The algorithm creates a large array of decision trees that are trained with a random subsample of the input data and use random features at each decision node (Breiman, 2001). This approach is widely used because RF models can easily deal with a high number of covariate layers and achieve good prediction results with a minimal amount of hyperparameter tuning (Breiman, 2001; Hengl et al., 2018). The models were trained and evaluated with the R package h2o (version 3.38.0.1) (LeDell et al., 2021). To improve comparably, all RF models were trained with the same default set of parameters (ntree = 200, mtries = number of predictors divided by 3). Categorical variables were preprocessed with one-hot encoding. A variable importance ranking was conducted with the h2o package (LeDell et al., 2021) to identify the most important covariates for the combined models in both states. The importance of every feature is estimated by calculating the attributed reduction of error (i.e., the variance reduction of the response value) at each decision node. To reduce the risk of overfitting, the training of combined models was performed in two steps. First, the variable importance was calculated for all co-

variates. Subsequently, covariates with relative importance above 1% in either state were selected and used in the final combined models in both states.

2.7 Accuracy Assessment

The prediction accuracies of RF models were assessed based on the results of a 10-fold cross-validation. For each of the five covariate groups, the process was repeated six times. R^2, root-mean-square error (RMSE), mean absolute error (MAE), Lin's concordance correlation coefficient (CCC), as well as the out-of-bag error (OOB) were calculated from the results of each cross-validation run. The differences between the groups were tested using an ANOVA and Tukey's honest significance test (HSD). Plots of the results were created with the package ggpubr (Kassambara, 2020) and the significance levels were illustrated with stars: `* p < 0.5; ** p < 0.01; *** p < 0.001; **** p < 0.0001`. To check if the models under- or overestimated the SOC values, a locally estimated scatterplot smoothing (LOESS) function was applied to the cross-validation results. The mean R^2 of the cross-validation results was calculated to compare the results of the mixed-data models and the control models. This was done for each of the five covariate groups and both states. Differences between both groups were assessed with standard error (SE) of the means.

2.8 Transferability of Different Covariates

To test the transferability of different covariates, five groups were created based on the primary data sources: (i) remote sensing data (satellite), (ii) terrain attributes (terrain), (iii) climate data (climate), (iv) soil legacy maps (soil), and (v) a combination of the most important covariates from all groups (combined). An overview of the workflow is shown in Figure 9.2. Soil samples from both states and environmental covariates from different sources were harmonized into a common raster format. In the first step, models for BY and BW were trained and assessed separately. Each of the five covariate groups was used to predict the SOC content of the respective state. The results were used as a baseline to assess the prediction accuracy of each covariate group before the models were transferred. The transferability was evaluated in two steps. First, the pre-trained models of the previous step were tested with the soil samples of the respective other states (Figure 9.2 A). It is not uncommon that models are extended to regions with

low or no availability of soil data. To simulate this condition, soil samples from both states were used to train the models (mixed-data models) while only soil samples from one state were used to test the model accuracies. For each state and covariate group, nine proportions were tested (10–90%) (Figure 9.2 B). The percentages correspond to the number of training samples from the same state as the test data. Control models were set up with the same training samples but without the additional data from the other state (Figure 9.2 C). The training data of both states were divided into ten equal parts using stratified random sampling. The random strata were then assigned to both groups manually to ensure that the samples are identical and that no bias is introduced with the training data. Differences in the accuracy of both groups can thus only be explained by the additional training data from outside the validation area.

2.9 Prediction of the SOC Maps

A final map of the SOC in both states was predicted using the RF models with the highest accuracies and the ranger package (Wright and Ziegler, 2017). SOC values were predicted for BY and BW separately and subsequently combined into a final raster. To illustrate the influence of the satellite and soil covariates on the prediction results, additional maps were generated using the individual covariate groups. The results are presented in magnified sites to showcase small-scale variations. In the last step, the satellite models of both states were transferred and used to predict the SOC of the respective other state. To examine the effectiveness of model transferability, two sites were selected in each state: one in an area with high similarity (i.e., the soil forming factors explaining variation in SOC are quite similar in the two states) and one in an area with low similarity (i.e., the soil forming factors explaining variation in SOC are different in the two states). The results were then compared to the SOC predictions of the non-transferred satellite models.

3 Results

3.1 Performance of the Baseline Models

The cross-validation results for each covariate group and the combined models are provided in Table 9.5. A median RMSE between 1.37 and 1.6% of SOC was

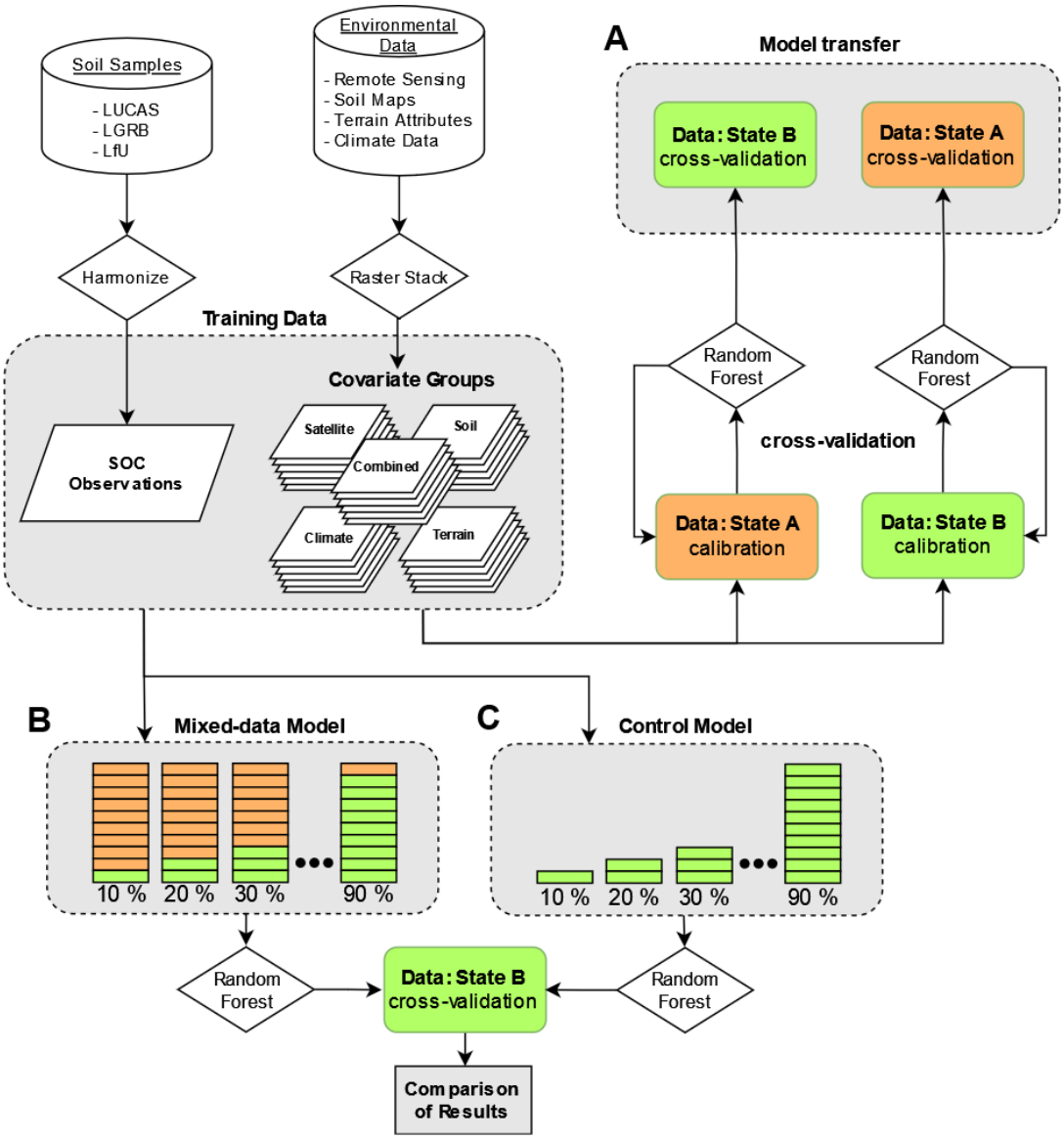

Figure 9.2: Flowchart of the SOC prediction and model transfer. (A) For each covariate group, models were trained with soil samples from one state and tested with samples from the other one. (B) Mixed-data models were trained with samples from both states to evaluate the effect of additional data. Nine models (10–90%) with different proportions were trained for all covariate groups and both states. The results were compared to control models built without the additional data (C).

Table 9.5: Cross-validation results (R^2, RMSE, CCC, MAE, and OOB) of the baseline and transferred models in both states.

		Baseline Models					Transferred Models			
State	Group	R^2	RMSE	CCC	MAE	OOB	R^2	RMSE	CCC	MAE
BY	combined	0.68	1.42	0.81	0.74	1.93	0.34	1.68	0.47	0.57
	satellite	0.53	1.53	0.71	0.88	2.69	0.42	1.63	0.50	0.58
	soil	0.61	1.48	0.77	0.79	1.97	0.00	1.91	0.00	0.76
	terrain	0.44	1.58	0.63	0.96	3.18	0.00	2.01	0.00	0.87
	climate	0.42	1.60	0.61	0.92	3.31	0.00	1.99	0.00	1.20
BW	combined	0.48	1.37	0.63	0.44	0.40	0.36	1.43	0.44	0.91
	satellite	0.30	1.44	0.48	0.52	0.53	0.25	1.50	0.41	0.87
	soil	0.31	1.43	0.50	0.49	0.51	0.00	1.64	0.00	1.15
	terrain	0.35	1.41	0.54	0.51	0.50	0.00	1.60	0.00	1.11
	climate	0.31	1.43	0.50	0.54	0.52	0.00	1.62	0.00	1.22

calculated. Combined models show significantly higher accuracies than all other covariate groups and produced the best results in both cases (R^2: BY = 0.68, BW = 0.48; RMSE: BY = 1.42%, BW = 1.37%). The results of the other covariate groups were different in both states (see Figure 9.3). Satellite (R^2 = 0.53) and especially soil models (R^2 = 0.61) produced good results in BY. In contrast to that, terrain models produced the second-best results in BW (R^2 = 0.35) while satellite models showed the overall lowest accuracy (R^2 = 0.30). In general, R^2 and RMSE were lower in BW.

Scatter plots of predicted and measured SOC values are illustrated in Figure 9.4. The models in BW showed a clear tendency to overestimate low and underestimate high values. Some models in BY had the same issue but a general underestimation of middle to high values seems to be apparent in most models. Soil and combined models produced the best results for high SOC values in BY and therefore also the overall highest R^2 values.

The results of the variable importance ranking of both states are shown in Figure 9.5. Spectral bands of the exposed soil composite were highly important for the models in both states. The SCMaP band 1–4 contributed 55% to the combined models of BY and 26% of BW. The categorical covariates for soil type, geology, and geomorphology also showed a high importance in both cases. In general, fewer covariates were classified with importance higher than 1% in BY. This is especially true for terrain covariates which were more important for models in BW.

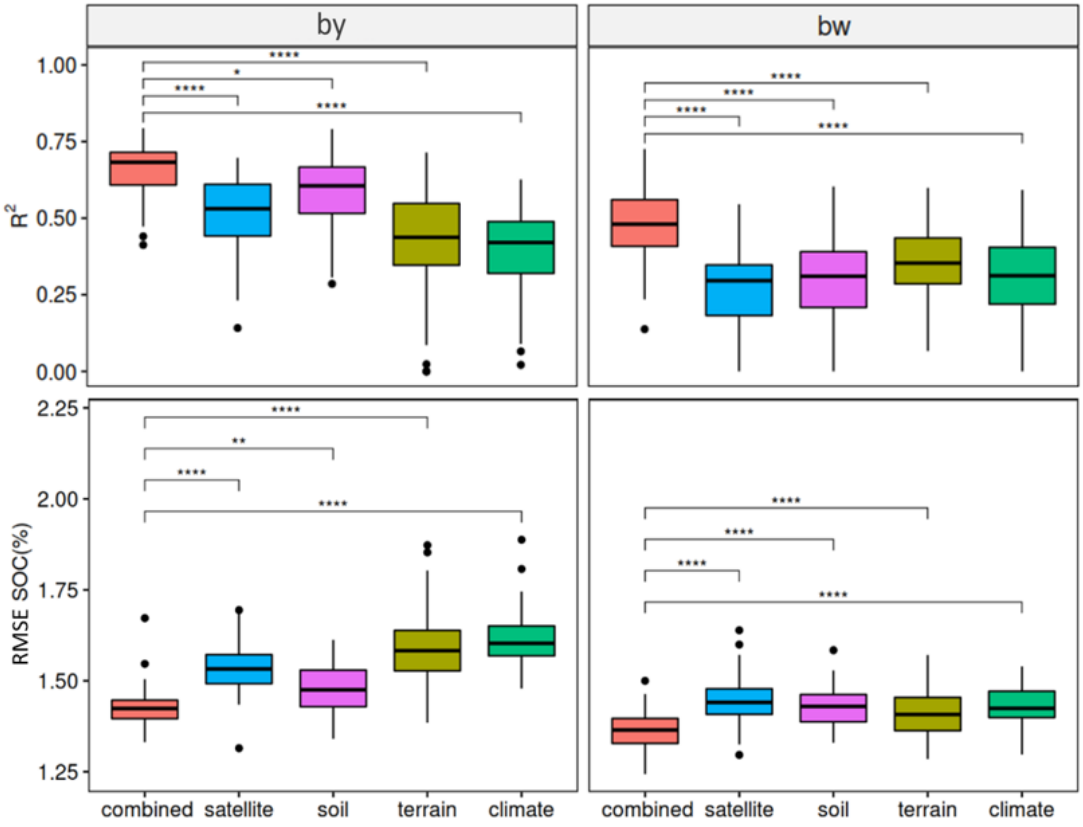

Figure 9.3: Boxplots (R^2 and RMSE) of the baseline models. Significant differences between combined models and the other covariate groups are indicated with stars (* $p < 0.5$; ** $p < 0.01$; **** $p < 0.0001$).

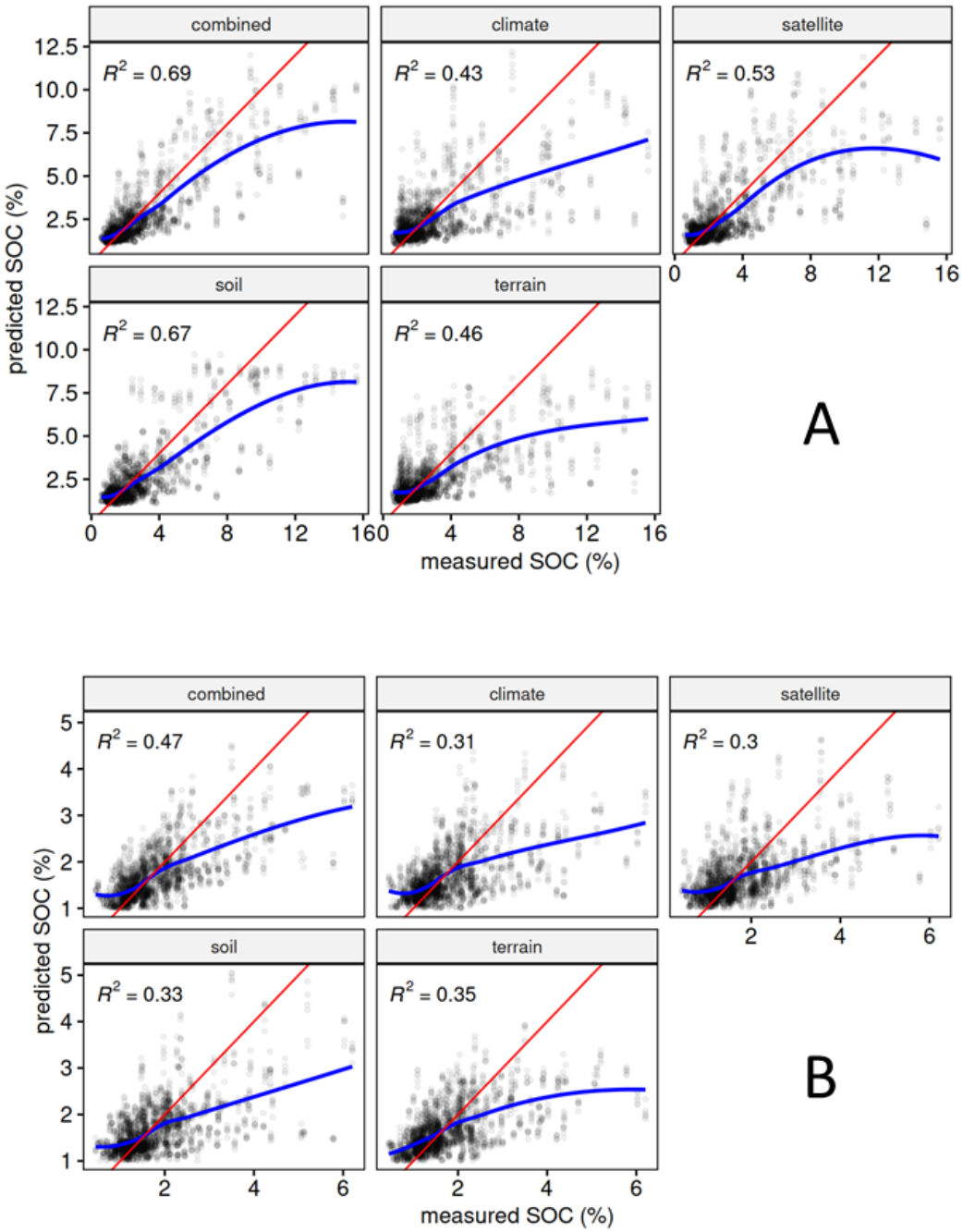

Figure 9.4: Scatter plot and LOESS-function (blue) of the predicted and measured SOC values in BY (A) and BW (B). R^2 values were calculated based on all cross-validation results (6 times 10). 1:1 lines are displayed in red.

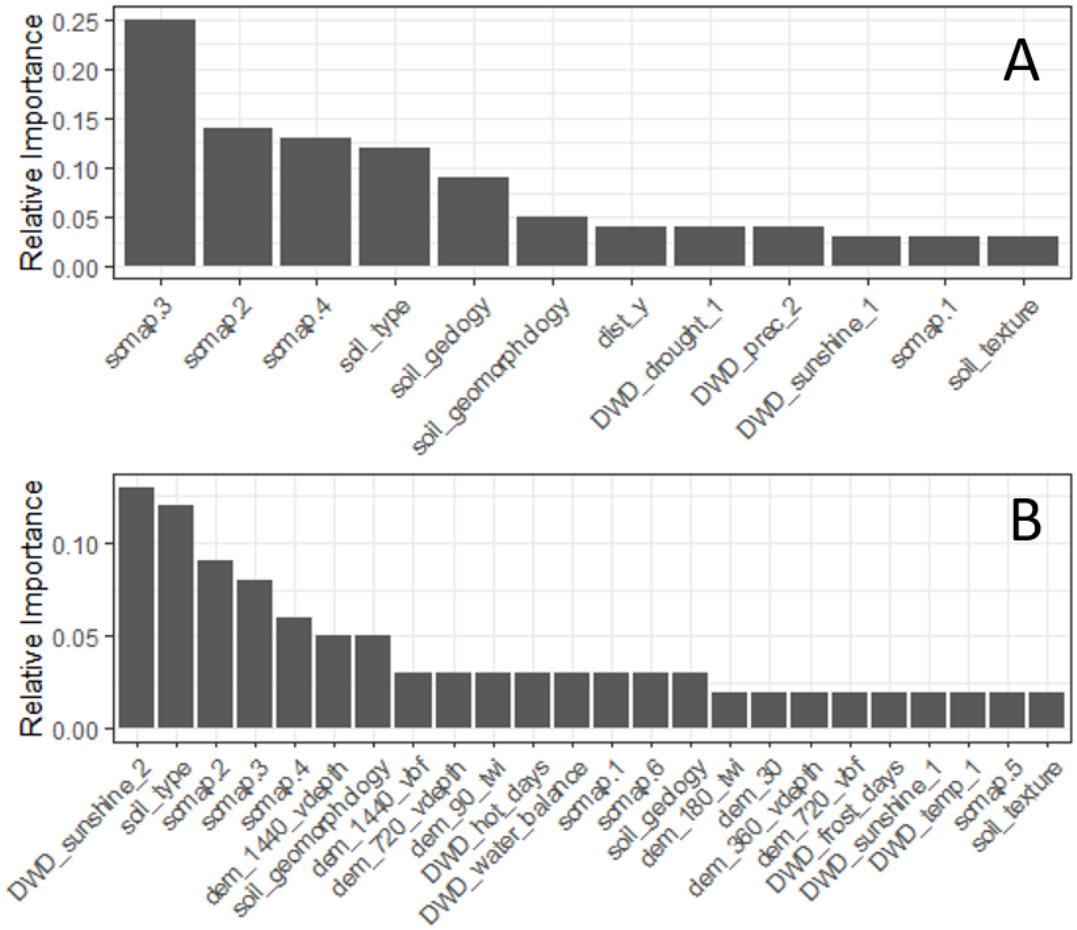

Figure 9.5: Relative variable importance of covariates in the combined models. Only covariates with relative importance above 1% are shown.

3.2 Performance of the Transferred Models

The results of the transferred models are presented in Table 9.5. Considering the comparatively low model performances for the soil, terrain, and climate group, it is clear that only combined and satellite models were successfully extended to the other state. Satellite models produced the best results in BY (RMSE = 1.63%) while the combined model showed better results in BW (RMSE = 1.43%). Differences between combined and satellite models before and after the model transfer are displayed in Figure 9.6. As expected, the R^2 declined in all cases but in different magnitudes. Generally, the decline was more significant for the models in BY and combined models in both cases.

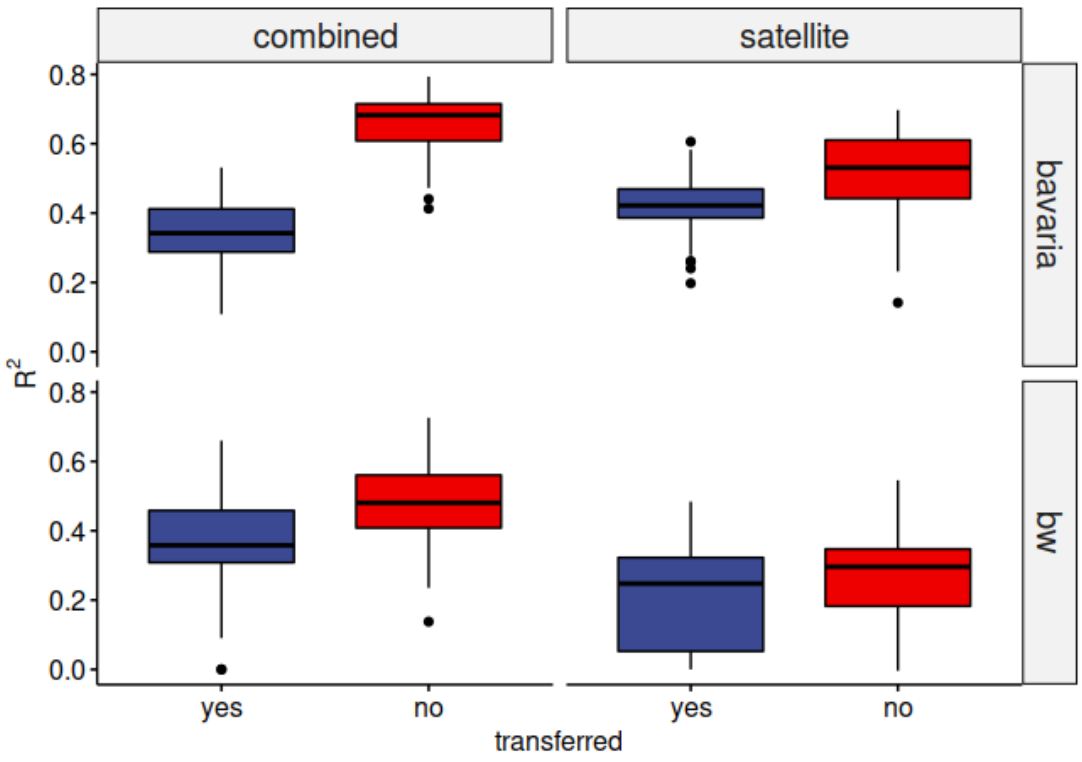

Figure 9.6: Comparison (R^2) of combined and satellite models before (red) and after (blue) the models were transferred and tested with the samples of the other state. Black dots indicate outliers.

3.3 Performance of the Mixed Data Models

The results of the mixed-data models are shown in Figure 9.7. R^2 values are plotted against the results of the control models (blue line) to evaluate the effect

of the additional data. The X-axis corresponds to the percentage of training data from the test area. It can be seen that the R^2 values at 90% are similar to the results of the baseline models while the results at 10% are similar to the results of the transferred models (Table 9.5). The R^2 of the control models increased with the number of training samples for each covariate group. This trend was more pronounced for small training sets (10–40%) and saturated with increasing size (e.g., Figure 9.7 C,D).

Additional training samples in the mixed-data models improved the accuracies of combined, satellite, and terrain models, as illustrated by the SE. This trend is more pronounced when the proportion of training data from the test state is low (10–20%) and is illustrated by the fact that the SE of control and mixed-data models is not overlapping. The improvement is greatest for satellite models in both states (Figure 9.7 C,D) and combined models in BW (Figure 9.7 B). In contrast to all other groups, mixed-data models decreased to accuracies of soil models in both states (Figure 9.7 E,F). No effect was observed for the climate models (Figure 9.7 I,J).

3.4 Final Prediction Maps

The final SOC map, based on the predictions of the combined models, is presented in Figure 9.8. It is visible that the highest SOC values generally occur in southern BY and are connected to the main streams, originating from the Alps. On the other hand, low SOC values are present in the north and west of BW, as well as in between the carbon-rich streams in southern BY. Magnified scenes of the predictions based on different covariates are presented in Figure 9.9. The maps illustrate local variations of the SOC as well as the influence of different covariates on the prediction results. In the case of BW, the local patterns of the SOC were similar when predicted with satellite (Figure 9.9 D) or combined models (Figure 9.9 F). This is not the case for BY, where clear differences between both models are visible (Figure 9.9 C,E). In contrast to the other maps, the outlines of the polygons are visible in the SOC maps of the soil models (Figure 9.9 A,B). However, this mosaicking effect is barely visible in the SOC maps of the combined models, even though the soil covariates were included. The maps in Figure 9.10 showcase the effect of the transferred satellite models on the local SOC predictions. Sites with dissimilar soil conditions between the states are

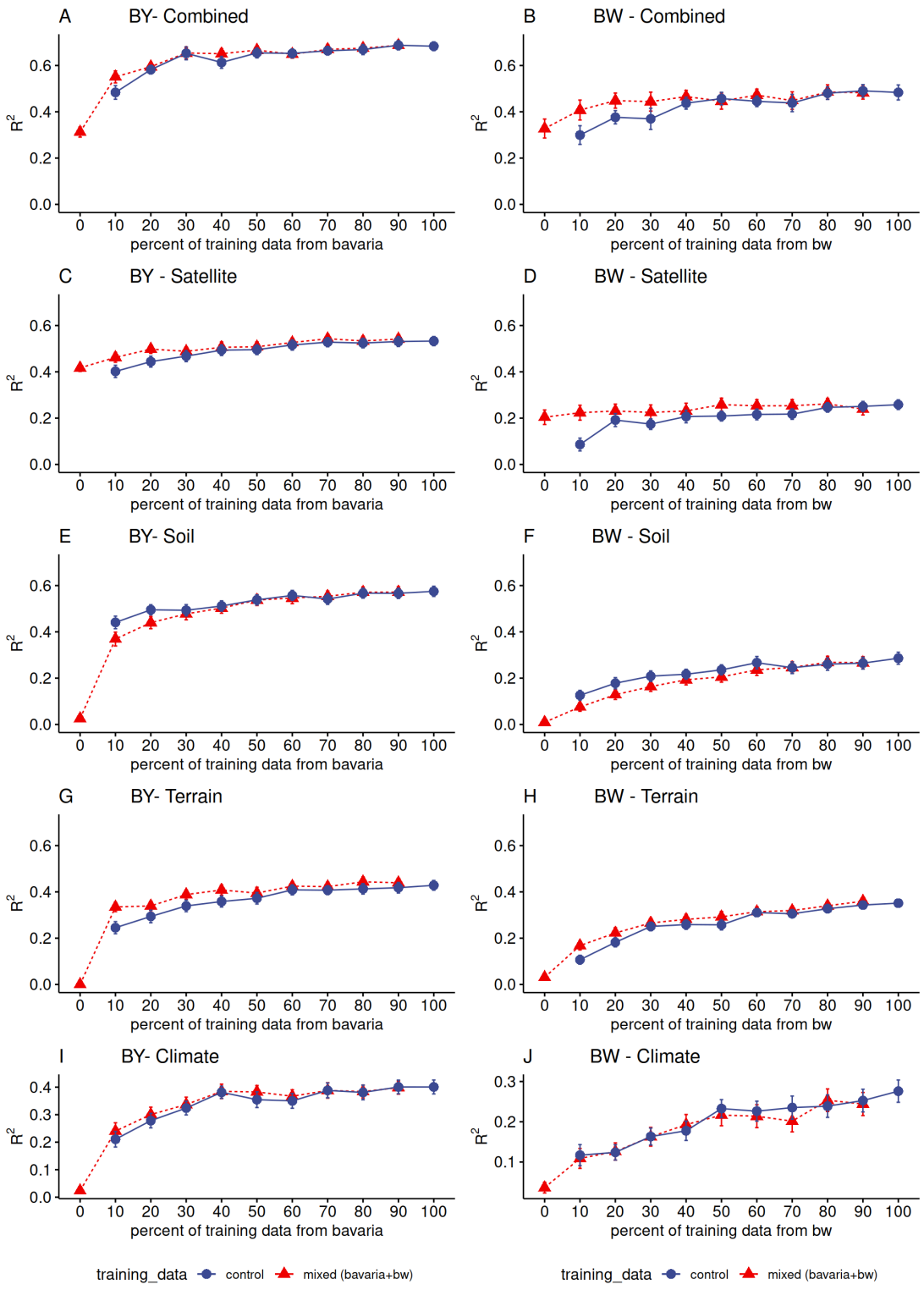

Figure 9.7: Results (R^2) of mixed-data (red) and control models (blue) for each covariate group. Mean values and SE of the 6 times 10-fold cross-validation are shown. The x-axis corresponds to the percentage of training samples from the state in which the models were tested. Control models were trained with the same data but without additional samples from the other state.

compared to sites with similar soil conditions. It is visible that satellite models trained in BY overpredicted high SOC values in BW (Figure 9.10 C,D) while the models trained in BW underpredicted the values in BY (Figure 9.10 A,B). This effect is less pronounced for the sites with similar soil conditions (Figure 9.10 B,D).

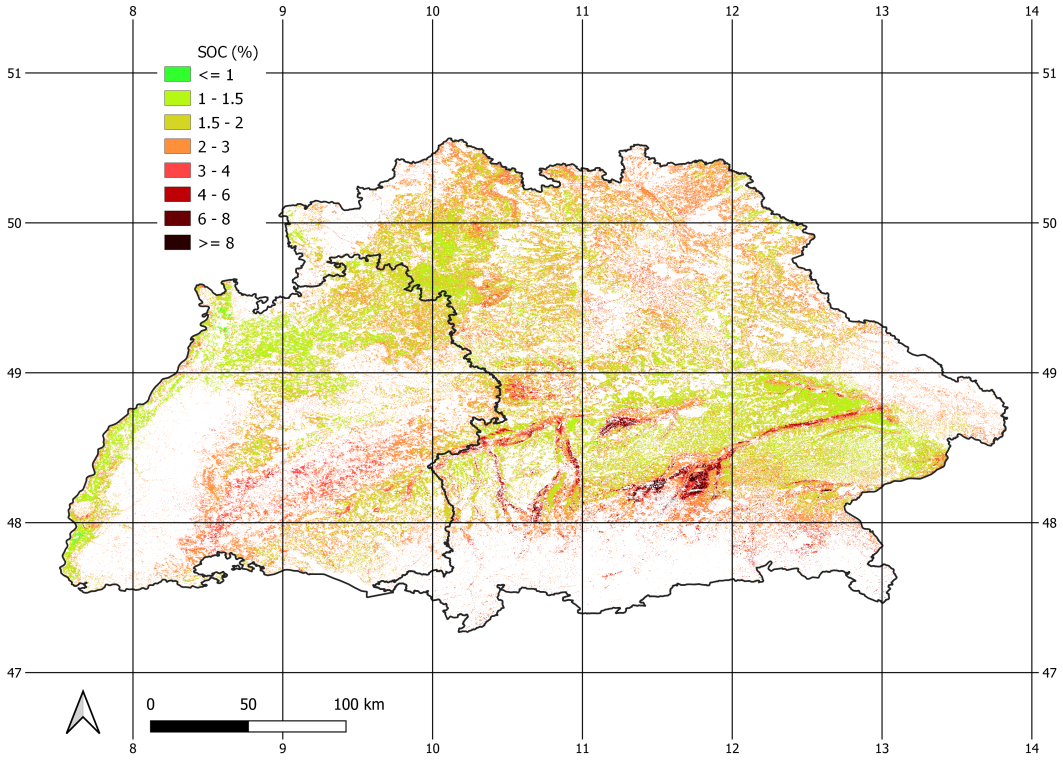

Figure 9.8: Final SOC map. Values were predicted for BY and BW independently, using the combined models of the respective state.

3.5 Similarity of the States

The MESS was calculated for both states using the locations of the training samples and the most important covariates (Figure 9.11). Negative values (red areas) correspond to covariate values that are not represented in the training data. The results show that the mountain ranges are not represented in the training data of either state. This was expected as no soil samples of these regions were included and almost no exposed soils were detected (Figure 9.1 C). Training samples of BW represent most parts of central BY but show low similarity with the most northern parts and some regions in the east (Figure 9.11 A). In turn, the training samples of BY are most similar to regions close to the border of BW but

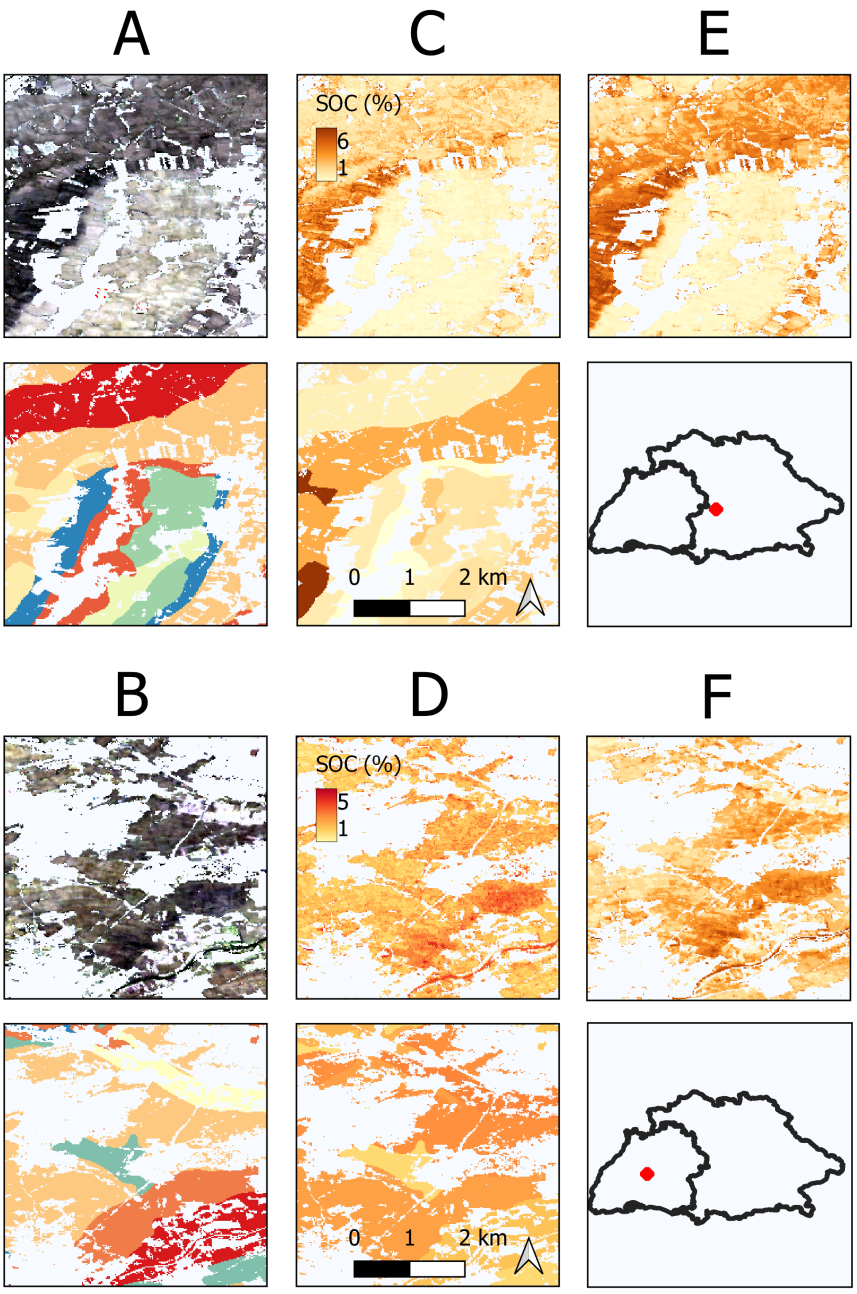

Figure 9.9: Comparison of the results from different models in BY (top) and BW (bottom). Examples of the exposed soil composite and soil types as input covariates (A,B); Respective SOC prediction for satellite and soil models (C,D); SOC predictions of the combined models (E,F). Red dots show the locations within the research area.

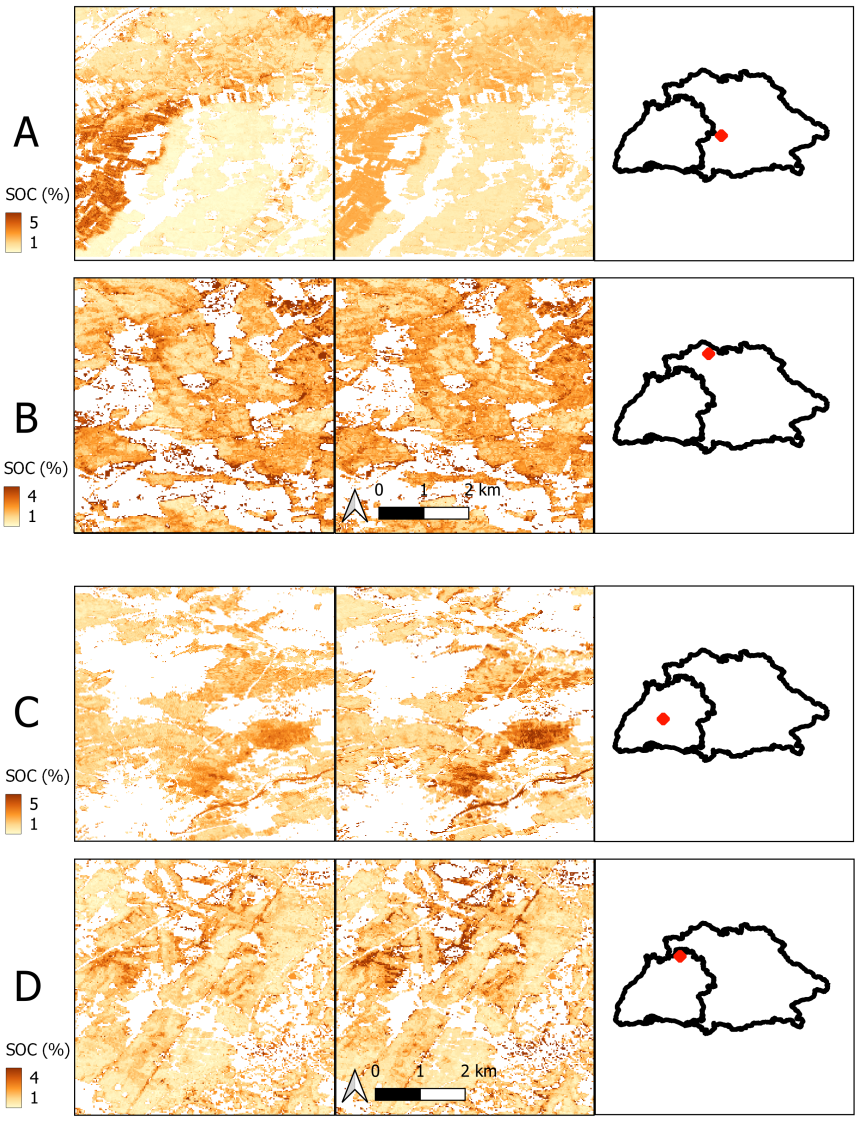

Figure 9.10: Influence of the model transfer on the SOC prediction for satellite models in BY (A,B) and BW (C,D). Maps of non-transferred models (left) and transferred models (right). Sites with dissimilar soil conditions in both states (A,C) are compared to sites with similar soil conditions (B,D). Red dots show the locations within the research area.

do not represent the covariate values of the Rhine Valley, in the west (Figure 9.11 B).

Density plots of important satellite, climate, and terrain covariates at the sampling locations are shown in Figure 9.12. The spectral bands are mostly normally distributed and the peaks are congruent. This is not the case for most of the other analyzed covariates. Especially, the distributions of the terrain covariates are fundamentally different between the states.

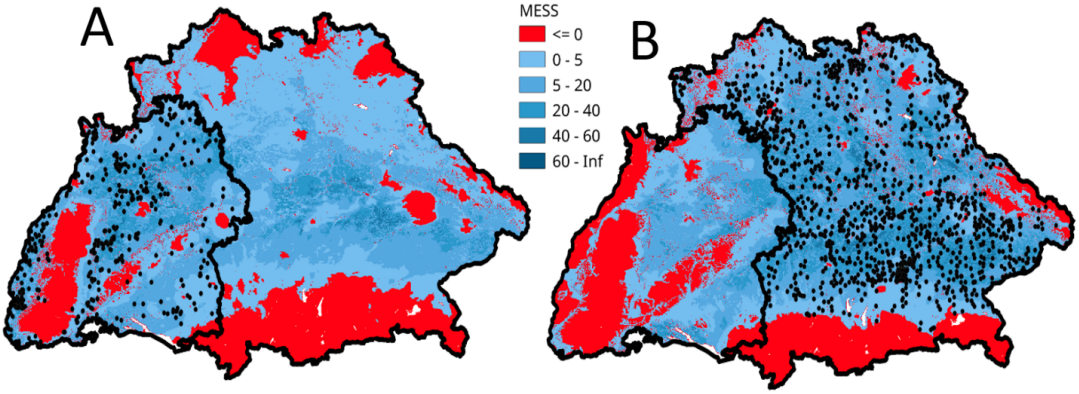

Figure 9.11: Multivariate Environmental Similarity Surfaces (MESS), calculated with the most important covariates and training samples of BW (A) and BY (B). Areas with high model uncertainty are colored in red. Black dots show sampling locations.

4 Discussion

4.1 General Model Performance

We compared DSM models of SOC in two German federal states to assess the transferability of five different covariate groups. To set a baseline, models were first trained and assessed without being transferred. The best results in both states were achieved with combined models that were trained with the most important covariates from all groups (see Table 9.5) (R^2: BY = 0.68, BW = 0.48). This finding is coherent with the concept of DSM and the idea to include a wide range of environmental covariates to represent and predict complex soil properties such as soil type or SOC content. Vaudour et al. (2022) conducted a review to analyze the usage of satellite imagery for the prediction of topsoil SOC.

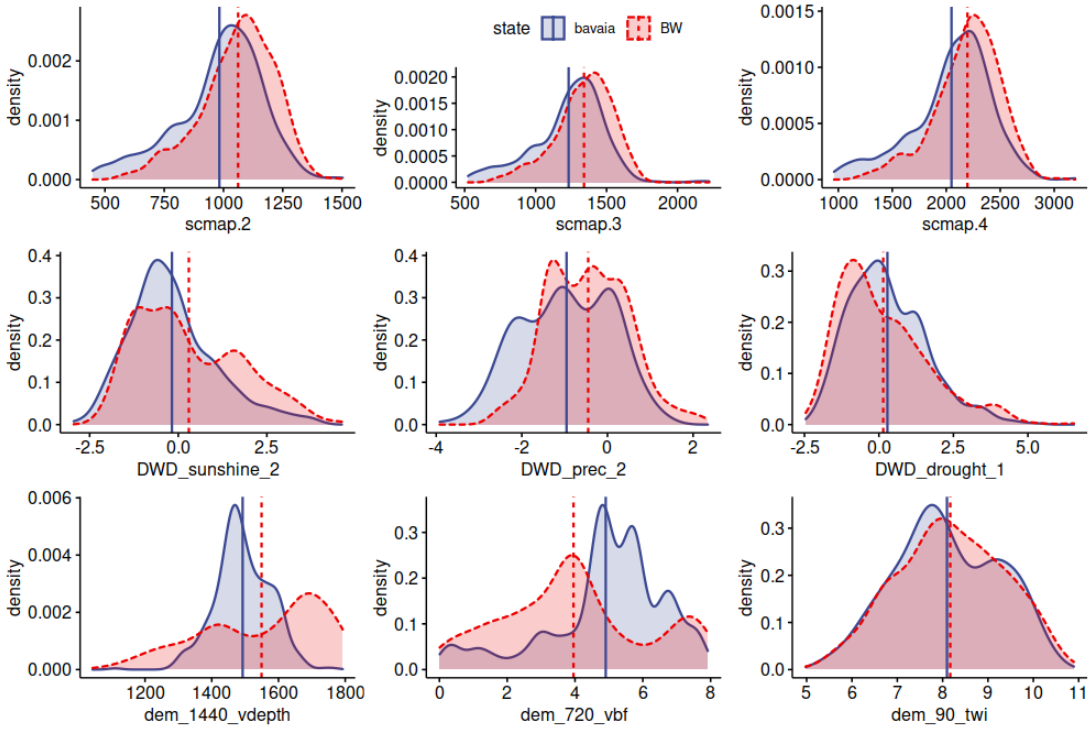

Figure 9.12: Density plots and mean values (vertical line) of important satellite, climate, and terrain covariates at the sampling locations of both states.

Most models were solely based on soil reflectance composites derived from time-series analyses. The majority of studies were conducted in small- to medium-sized regions and R^2 values between 0.5 and 0.7 were reported (Dvorakova et al., 2021; Urbina-Salazar et al., 2021; Vaudour et al., 2021; Zepp et al., 2021a). A lower accuracy was achieved in a model on a continental scale (R^2 = 0.35) (Safanelli et al., 2020) while R^2 values above 0.8 were reported for observations on field level with Sentinel-2 and high-resolution data (Žížala et al., 2019). Considering the relatively large size of the research area, combined models produced good results, especially in BY. However, the RMSE in this study is still relatively high (BY = 1.42, BW = 1.37) and the models should not be used for final SOC predictions without further optimization. As seen in the results of the literature, the error also depends on the size of the research area. Large areas, such as the one in this study, usually cover more diverse landscapes and are therefore more prone to error. Still, the RMSE can be improved by model optimization and hyperparameter tuning. However, this was not the main objective of this study, as the comparability between the different covariate groups was more important to describe the transferability.

Ensemble models based on multiple individual ML algorithms have been successfully used to optimize the prediction of SOC (Taghizadeh-Mehrjardi et al., 2020) and improve the extrapolation of spatiotemporal DSM models (Bonannella et al., 2022). Mixed results were reported for the relative importance of additional covariates in comparison to the satellite images for the prediction of SOC. In one case, the integration of satellite time-series data did not improve the model accuracy (Zeraatpisheh et al., 2022). In other studies, the remote sensing images were the most relevant covariates but the importance also depended on the research area (Fathizad et al., 2022; Fathololoumi et al., 2020; Taghizadeh-Mehrjardi et al., 2020). The results of our study show that the satellite covariates were most important for the combined models in both states, closely followed by the soil covariates (Figure 9.6). Comparing the individual covariates groups, soil models performed slightly better than satellite models in both cases (Table 9.5). This mismatch can be explained by the fact that the satellite models underestimated high SOC values, as illustrated by the LOESS function (Figure 9.4). This resulted in a reduction in the overall R^2 of the satellite models and is less pronounced in soil models, as carbon-rich gleysols are easily segregated from other soil types. As illustrated by the combined models, satellite covariates were generally more important for the prediction of SOC (Figure 9.5). However, the soil models showed better results for areas with high SOC concentrations and therefore produced high R^2 values. This demonstrates the fact the accuracy of remote sensing data can be significantly increased by adding further covariates from different sources, especially soil properties from legacy maps.

4.2 Differences between the States

The models in BY generally produced better R^2 values while the RMSE was slightly lower in BW (Figure 9.3). It has been shown that the R^2 correlates with the SD of the soil samples, which is significantly higher in BY (Mcbratney and Minasny, 2013) (Table 9.2). This can be explained by dissimilarities in the natural soil conditions of both states. High SOC values are mainly present in the North Alpine Foreland Basin of BY and are allocated to high soil moisture (Wiesmeier et al., 2013). Even though bog soils were excluded, carbon-rich Gleysols were more abundant in BY (Table 9.4). As shown in Figure 9.4, this fact also influences the prediction results of the models in BY. Another reason for different prediction results between states might stem from differences in soil data collec-

tion and analysis protocols. The inclusion of soil samples from legacy databases can introduce uncertainties resulting from different sampling methods and imprecise localization. Sampling design is another important factor of DSM (Ma et al., 2020). The proportion of LUCAS samples was higher in BY which led to a better representation of the research area and less spatial clustering (Figure 9.1) (Meyer et al., 2018). From the results in Figure 9.7 (A), it can be seen that only 30% of the samples in BY were necessary to maximize the prediction accuracy of the combined models. This trend is different in BW where the accuracy incrementally increases with sample size (Figure 9.7). The trend can be explained by the spatial clusters in BW and the fact that more samples are necessary to reach the same representation of the research area as in BY. Furthermore, the redundancy of soil information might be higher in BY, as fewer samples are necessary to reach a stagnating accuracy.

The MESS and the density plots of terrain covariates illustrate significant topographic dissimilarities between both states (Figure 9.11 and Figure 9.12). This is especially true for the Rhine Valley in the west of BW which is heavily cultivated and very dissimilar to the sampling locations in BY (Figure 9.1 and Figure 9.11). Overall, the terrain covariates were more important for the models in BW (Figure 9.6). The landscape in BW is diverse and terrain attributes might influence the distribution of SOC to a larger extent than in BY, which is mostly dominated by North Alpine Foreland Basin.

Surprisingly, satellite models had the lowest overall R^2 of all covariate groups in BW (Table 9.5). This result is not in line with the high accuracies that were reported using soil reflectance composites to predict SOC (Vaudour et al., 2021; Zepp et al., 2021a). The differences between satellite models in both states could also be explained by the lower variance of SOC values in BW. As described in Section 2.1, the geologic formations in BW are diverse and affect the soil formations as well as SOC. The multispectral images may not provide enough information to differentiate between the different soil types and detect small variations in SOC content. This argument is supported by the fact that the accuracy of the combined models is significantly higher and that the soil types were identified as important covariates (Figure 9.5). Effects such as soil moisture or crop residue on the generation of the exposed soil composite could be another reason why it is more difficult to predict SOC values with a lower variance (Castaldi et al., 2019a; Dvorakova et al., 2020). These problems may be improved by the availability

of large-scale hyperspectral images (Guanter et al., 2015; Guo et al., 2021; van Wesemael et al., 2021), and comprehensive measurement campaigns to improve the representation of cropland soils (Jacobs et al., 2018). Nevertheless, the multispectral bands were still valuable for the prediction of SOC, as seen from the variable importance rating of combined models (Figure 9.6).

4.3 Transferability of Different Covariate Groups

Satellite imagery has been successfully used to predict the SOC content on many occasions (e.g., Taghizadeh-Mehrjardi et al. (2020); Zepp et al. (2021a)). However, little effort has been made to use remote sensing data to transfer SOC models to extended research areas. The results of this study show that satellite covariates are highly transferable and can be used to improve model extrapolation (Table 9.5). This is supported by the fact that the accuracy of the satellite models improved when soil samples from another state were added to the training data (Figure 9.7 C,D). The prediction maps of the transferred satellite models show that the results are similar to the original models (Figure 9.10). Overall, the results indicate that the transferred models based on the soil reflectance composite can serve as a basis for the prediction of SOC in unknown regions. However, the SOC maps based on the transferred models also illustrate the influence of the input samples. Mean SOC values were generally lower in BW (Table 9.2) and the transferred models underpredicted the SOC in BY (Figure 9.10 A). Differences in the distribution of the input samples should therefore be considered when transferring satellite models. Subsequently, the results can be improved and fine-tuned with the addition of soil samples from the prediction area, as shown by the results of the mixed data models (Figure 9.7 C,D). It has been shown that soil reflectance, especially in VNIR–SWIR region, is closely correlated to the accumulation of organic matter in the topsoil (Gholizadeh et al., 2021; Vohland et al., 2022). For this reason, efforts have been made to adapt SOC models that were trained with laboratory soil spectral libraries to remote sensing data (Castaldi et al., 2019a; Hong et al., 2020). The results of this study suggest that a similar approach could be used to improve the transferability of DSM models.

Terrain covariates have been regularly used to transfer DSM models (de Arruda et al., 2016; Machado et al., 2019; Malone et al., 2016; McKay et al., 2010; Neyestani et al., 2021; Wolski et al., 2017; Zhao et al., 2020) and it was

unexpected that the transferability was lower in comparison to the satellite data (Table 9.5). A possible explanation could be the fact that the terrain attributes of both states are not similar enough (Figure 9.11 and Figure 9.12). The problem was improved by the additional training data. In this case, the mixed-data models were able to adjust for the dissimilarities of the terrain attributes and improve the prediction accuracy in comparison to the control models (Figure 9.7 G,H). Hence, the results suggest that the terrain covariates are only transferable under optimal conditions and if the extended research area is similar to the training data. This is fundamentally different from satellite covariates and should be considered when terrain covariates are used to transfer DSM models.

Models based on soil covariates showed no sign of transferability. One possible explanation could be the natural dissimilarities of soil conditions in both states. This is illustrated by the imbalance of the soil classes in both states, as summarized in Table 9.4. All soil classes must be present in both states for the models to be transferable. However, because of different natural conditions, some soil types are underrepresented in samples of BW. This can introduce biases during the model training which are then transferred to the other state and lead to poor results. In contrast to terrain and combined models, the accuracy of mixed-data models dropped in comparison to the control models (Figure 9.7 E,F). These results suggest that the soil covariates were not transferable and should therefore only be used in a local context. Unlike EO data, underlying soil legacy maps are based on various soil mapping standards and extrapolated measurements. Similar conclusions were drawn in the study of Zhao et al. (2020) who only used information from soil legacy maps to correct the results after the models were transferred to an extended area. To improve the transferability of soil covariates, the soil condition needs to be as similar as possible to reduce the errors resulting from unbalanced classes.

4.4 Limitations and Future Research

The sampling design is an important step to optimize the results of DSM. The inclusion of clustered sampling locations can result in biased predictions and high uncertainties (Hengl et al., 2022; Meyer et al., 2018). Biases might also arise from the fact that samples of the LUCAS campaign and the soil databases are not harmonized as they were collected in different years and with different

sampling protocols. SOC, as a dynamic soil property, is exposed to constant change and it is arguable whether the observed period of 1984–2014 is appropriate for accurate predictions. These factors illustrate the fact that it is necessary to optimize the models and improve the overall results. However, the main goal of this study is not to maximize the prediction accuracy of the models but to compare the transferability of stable covariate groups. Future studies should be directed to test the spatial as well as temporal transferability of remote sensing for the prediction of SOC dynamics. To guarantee comparability, harmonized soil data should be used to optimize the spatial representation and reduce biases. Recently, efforts have been made to improve the sample coverage in Germany with a nationwide Agricultural Soil Inventory (BZE-LW) to improve the quality of future DSM projects (Jacobs et al., 2018; Sakhaee et al., 2022).

Unlike linear regressions, RF models and other tree-based algorithms have problems with the prediction of values outside the range of the training data (Hengl, 2021). It has been shown that ensemble models based on different types of learners produce better results for unknown values and should therefore be preferred for spatial or temporal extrapolation (Bonannella et al., 2022; Hengl, 2021). The next step should therefore be to test the transferability of covariates in the context of ensemble models which also showed promising results for the prediction of SOC (Taghizadeh-Mehrjardi et al., 2020). The success of DSM extrapolation highly depends on the similarity of biophysical conditions in the donor and extended modeling area (Zhao et al., 2020). The MESS (Figure 9.11) and the distribution of the soil types (Table 9.4) illustrate that there are significant differences in the biophysical conditions of both states which influence the prediction accuracy as well as the transferability. Taking this into account, further research should be conducted to describe the similarity of different regions and define thresholds and conditions under which models can be successfully transferred (Roudier et al., 2022). Additionally, the similarity can be increased by defining suitable conditions to reduce the model area. In the case of this study, this can be done by removing certain soil types (e.g., Gleysols) which introduce extreme values and decrease the model-similarity of both states.

4.5 Recommendations for the Transfer and Extrapolation of SOC-Models

The results of our study show that there are significant differences in the transferability of covariate groups. The extrapolation of DSM models can be split into two phases to increase the number of available soil samples while still accounting for the predictive power of non-transferable covariates. For the first step, only covariate groups with a high degree of transferability should be used to train the models. These include information derived from EO data like soil reflectance composites or DEM attributes. This way, soil samples from outside the research can be included to improve the model accuracy (Figure 9.7). Nevertheless, the additional soil samples should be selected to reflect the natural conditions and to be as similar as possible to the research area. The proportion of training data from inside and outside the research area should be chosen to maximize the positive effect of the additional training data. For this study, the biggest accuracy increase was found when around 80% of soil samples came from outside the research area. A similar proportion was found in the study of Zhao et al. (2020) who applied a reverse k-fold cross-validation to estimate how many samples are necessary to maximize the model accuracy in an extended research area. Once the highly transferable covariates were used to predict the soil attributes, further covariates can be used to correct and improve the results. This can be realized by converting the results of the previous step into a covariate layer which can subsequently be combined with other covariates like soil legacy data or climate parameters. For this step, it is important to only include soil samples from the actual research area which corresponds to 20% of the total training samples in this case. This is necessary for including the remaining covariates that are not transferable and therefore do not produce good results with additional training samples from outside the research area. This process can optimize the prediction accuracy while minimizing the number of necessary soil samples in the research area.

5 Conclusions

The transferability of different covariate groups for the prediction of SOC in two federal states in Southern German was tested. The results of this experiment

clearly show the importance of EO data for transferring or extrapolating DSM models regarding SOC prediction. Covariates derived from a soil reflectance composite showed the highest degree of transferability. Models based on multispectral EO data and DEM derivatives were improved by soil samples collected outside the research area. This effect was more pronounced the fewer samples inside the research area were available. A similar trend was found for terrain covariates and combined models. These findings are an important step to improve the precision of DSM models in regions with low availability of soil samples. However, the results of the non-transferred models suggest that covariates from soil legacy data and climate parameters are also important for SOC prediction. Therefore, it is proposed to split the modeling process to utilize the predictive power of transferable and non-transferable covariates. First, the EO covariates can be used to transfer DSM models to areas with low availability of soil samples. Second, additional non-transferable covariates can be combined with local samples of the research area and used to correct the results. Future work should focus on the dissimilarity of donor and recipient areas and define conditions to improve the results of transferred DSM models.

Acknowledgements

We thank the Bavarian Environment Agency (LfU) for providing access to the soil database. We also thank the Japan Aerospace Exploration Agency (JAXA) for providing the Digital Elevation Model.

Chapter 10

Manuscript 4: Spatiotemporal monitoring of cropland soil organic carbon changes from space

Tom Broeg, Axel Don, Martin Wiesmeier, Thomas Scholten, Stefan Erasmi
Global Change Biology, 2024

https://doi.org/10.1111/gcb.17608

Received: 15 August 2024; Revised: 6 November 2024; Accepted: 7 November 2024

Abstract

Soil monitoring requires accurate and spatially explicit information on soil organic carbon (SOC) trends and changes over time. Spatiotemporal SOC models based on Earth Observation (EO) satellite data can support large-scale SOC monitoring but often lack sufficient temporal validation based on long-term soil data. In this study, we used repeated SOC samples from 1986 to 2022 and a time series of multispectral bare soil observations (Landsat and Sentinel-2) to model high-resolution cropland SOC trends for almost four decades. An in-depth validation of the temporal model uncertainty and accuracy of the derived SOC trends was conducted based on a network of 100 long-term monitoring sites that were continuously resampled every five years. While the general SOC prediction accuracy was high ($R^2 = 0.61$; RMSE = 5.6 g kg^{-1}), the direct validation of the derived SOC trends revealed a significantly greater uncertainty ($R^2 = 0.16$; $p < 0.0001$), even though predicted and measured values showed similar distributions. Classifying the results into declining and increasing SOC trends, we found that 95% of all sites were either correctly identified or predicted as stable ($p < 0.001$), highlighting the potential of our findings. Increased accuracies for SOC trends were found in soils with higher SOC contents ($R^2 = 0.4$) and sites with reduced tillage ($R^2 = 0.26$). Based on the signal-to-noise ratio and temporal model uncertainty, we were able to show that the necessary time frame to detect SOC trends strongly depends on the absolute SOC changes present in the soils. Our findings highlight the potential to generate significant cropland SOC trend maps based on EO data and underline the necessity for direct validation with repeated soil samples and long-term SOC measurements. This study marks an important step towards the useability and integration of EO-based SOC maps for large-scale soil carbon monitoring.

1 Introduction

Soils are an essential part of the global carbon cycle and could play an integral role in our mission toward climate neutrality (Amelung et al., 2020). Policies, such as the European Green Deal, heavily rely on sustainable soil management to achieve comprehensive goals regarding soil health, climate protection, and biodiversity (Montanarella and Panagos, 2021). In this context, it is increasingly

important to monitor and report the status of agricultural soils. Advanced soil monitoring systems are needed to track long-term changes and to inform about sustainable management and greenhouse gas emissions (Smith et al., 2020). Soil organic carbon (SOC) is one of the key indicators for soil health and carbon sequestration and has been the focus of current soil monitoring strategies (Lal, 2016). It has been shown that changes in SOC are closely related to agricultural practices (Paustian et al., 2016). Hence, the global demand for fine-scale information on SOC dynamics and management data of agricultural land is increasing (Paustian et al., 2019). Accurate information on long-term SOC changes can be used to verify and track the effects of soil management and help quantify carbon sequestration in carbon farming schemes (European Commission, 2022).

Earth Observation (EO) provides extensive information on land cover and land use of agricultural soils. Starting with the launch of Landsat 5 in 1984, the availability of multispectral satellite data has increased substantially during the last decades. The potential of remote sensing to predict SOC dynamics was already recognized more than a decade ago (Croft et al., 2012). Multitemporal soil samples can be combined with the corresponding EO data to conduct spatiotemporal models and generate information on SOC trends and changes (Hengl et al., 2018; Heuvelink et al., 2021). Based on this, spatially explicit and fine-scale SOC predictions can act as an important resource to support existing soil monitoring programs such as the Land Use/Cover Area Frame Survey (LUCAS) soil survey in Europe (De Rosa et al., 2024). However, this potential has rarely been realized due to challenges with SOC changes being small (<0.5 g kg^{-1} $year^{-1}$), the lack of appropriate ground truth data from long-term soil monitoring, and the low availability of EO data for early years (Croft et al., 2012).

Recent studies have shown that spectral information of bare soils can be used to compute high-resolution, area-wide SOC maps (Broeg et al., 2024a; Möller et al., 2022; Zepp et al., 2021a). Soil reflectance composites (SRC) have been generated by averaging bare soil observations over time to obtain the mean soil reflectance at the pixel level (Demattê et al., 2018; Heiden et al., 2022; Rogge et al., 2018). Despite improving the quality of static SOC maps (Broeg et al., 2023), SRC have been rarely used to predict dynamic SOC trends and changes. This is because SRC are usually generated with data from multiple years to increase the number of bare soil observations and to reduce the noise of the soil signal (Zepp et al., 2023). Instead, several studies rely on composites of the full EO

time series to predict SOC changes (Meng et al., 2024; Paul et al., 2020; Venter et al., 2021; Yang et al., 2023). In this case, however, the presence of vegetation can significantly affect the spectral SOC signal in the composites (Dvorakova et al., 2022).

Several studies have raised concerns about the high uncertainty and low signal-to-noise ratio of EO-based SOC models (Chabrillat et al., 2019; Croft et al., 2012). For spatiotemporal SOC models and the prediction of SOC trends, the signal can be defined as the absolute SOC change in time, while the noise describes the SOC prediction uncertainty in space (Heuvelink et al., 2021; Paustian et al., 2019). In this regard, the high local variability of SOC represents a major challenge to detect SOC trends (Poeplau et al., 2022). Predicting SOC stocks in Argentina from 1982 to 2017, Heuvelink et al. (2021) were able to demonstrate that the spatial SOC variation can be up to one order of magnitude higher than the temporal one. This is especially true for cropland soils, as SOC contents are usually much smaller compared to other land-use types and it can take up to 15 years to measure significant changes (Poeplau and Don, 2013; Smith, 2004). It therefore remains crucial to evaluate the signal-to-noise ratio and the conditions under which spatiotemporal models can predict SOC trends in cropland soils.

Soil data from resampled sites is necessary to directly validate predicted SOC changes and to test the temporal accuracy of spatiotemporal models in an unbiased manner. This is, however, not feasible in most regions due to the low availability of SOC samples from long-term monitoring programs. Meng et al. (2024) used Landsat and soil data from 1984 to 2021 to predict large-scale SOC time series but had to rely on indirect measures to estimate the temporal model accuracy due to limited reference data. For the same reason, Venter et al. (2021) were unable to validate the SOC trend map that was produced for South Africa using Landsat composites and yearly SOC predictions from 1986 to 2019. Because of missing data on SOC change and the lack of direct validation, the usability of spatiotemporal SOC maps is currently very limited (Heuvelink et al., 2021; Meng et al., 2024; Venter et al., 2021). Following the review of Vaudour et al. (2022), the verifiability of predicted SOC changes therefore remains a critical issue, underlining the necessity to integrate soil data from long-term monitoring programs.

Evaluating the results of local and regional soil monitoring, it has been shown

that the drivers behind the cropland SOC dynamics are diverse (Wiesmeier et al., 2019). Recent spatiotemporal models have set a strong focus on land use changes as one important explanatory factor for SOC trends (De Rosa et al., 2024; Paul et al., 2020; Yang et al., 2023). As reported by several studies, however, SOC stocks have been generally decreasing in recent years, including the soils with continuous cropland management and no land use changes (Bellamy et al., 2005; Ciais et al., 2010; Heikkinen et al., 2013). To improve the knowledge about cropland SOC dynamics and provide unbiased and high-resolution maps, SRC time series and variances in the spectral soil signal can be observed and directly linked to SOC changes. Some factors, like tillage or soil texture, can influence the cropland SOC as well as the soil reflectance and must be, therefore, analyzed in detail to improve the prediction of SOC trends (Castaldi et al., 2019a; Dvorakova et al., 2020).

In summary, it remains challenging to predict and validate SOC trends based on EO data and repeated soil samples. In this study, we used data from Landsat and Sentinel-2 satellite image time series to generate a datacube of SRC from 1986 to 2022 and trained a spatiotemporal SOC model based on soil samples from different long-term monitoring sites. We aim to develop and assess EO-based monitoring of cropland SOC trends by addressing the following research questions:

1. Is it possible to accurately predict short-term SOC changes and long-term SOC trends based on remote sensing data?
2. How is the uncertainty of SOC predictions based on SRC changing over time?
3. What factors influence the prediction and accuracy of cropland SOC trends?

2 Material and methods

2.1 Research area

The research area of our study is located in the federal state of Bavaria, southeast Germany (Figure 10.1 a). It is characterized by the Alps in the south, several low mountain ranges in the north, and east, and the lowland of the southern Molasse basin. The elevation ranges from 107 to 2962 m ASL and the climate is located in between maritime (northwest) and sub-continental (southeast) in-

fluences (Figure 10.1 b). In total, cropland accounts for about 34 % of the land area. Grasslands and forests are predominantly located in mountainous regions with increasing altitude. Due to the diverse landscapes and parent materials, various soil classes are present in Bavaria. They include soils with well-developed B horizons (Cambisols and Luvisols), as well as soils with initial soil formation (Leptosols). In some parts, high sand content led to the development of bleached A horizons (Podzols) while Vertisols formed on clay-rich soils. Stagnosols and Gleysols are present in regions with shallow groundwater levels along valleys. In some regions of the area, the large streams originating in the Alps formed lowlands and induced the formation of natural bogs and fens (Histosols). Additional information about the soils in Bavaria can be found in Wiesmeier et al. (2013).

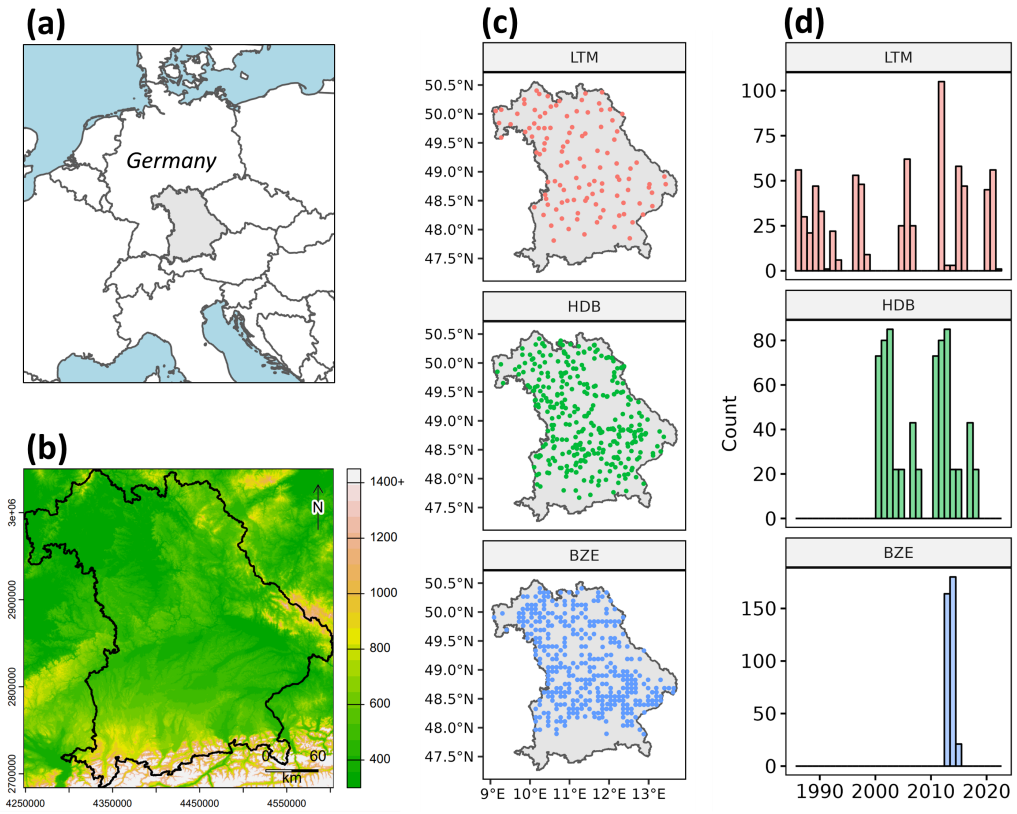

Figure 10.1: (a) Location of the research area in Central Europe. (b) Topographic map of Bavaria (m ASL). Spatial (c) and temporal (d) distributions of the SOC samples, used to train the model. LTM = Long-term monitoring data (1986 – 2022); HDB = Monitoring data from the humus database (2001 – 2018); BZE = German Soil Inventory (2013 – 2015). Map lines delineate study areas and do not necessarily depict accepted national boundaries.

2.2 Soil samples

For our study, we built on comprehensive soil samples from different programs (Table 10.1) (Broeg et al., 2024c). An overview of the sampling locations and temporal distributions is given in Figure 10.1 c,d. We included a total of 100 sites from a long-term soil monitoring program (LTM) that were resampled seven times from 1986 to 2022 (Wiesmeier et al., 2024). The locations were selected to be representative of the climatic and geologic conditions in Bavaria (LfL, 2022). They include 85 sites with continuous cropland management, seven sites with permanent crops, and eight sites with a change of cropland and grassland use within the observation period. The SOC content was analyzed in intervals of approx. 5 years: 1986-88, 1989-93, 1996-99, 2005-07, 2012-14 2015-16, and 2020-22 (Figure 10.1 d). For each site, 1000 m^2 large parcels (~ 30x30 m) were laid out and marked with stationary magnets. The parcels were sampled at a depth of 0 – 15 cm, using four composite samples. A detailed description of the subsample layout is provided in LfL (2022). For laboratory analyses, the samples were dried (40 °C), sieved (2 mm), and ground to 0.2 mm. The SOC contents (g kg^{-1}) were calculated by analyzing the total soil C (Dumas method; Elemental analyzer) and subtracting the carbonate contents (Scheibler method). For our model, median SOC content values were calculated for each sampling date and linked to the central GPS coordinates of the plots.

Table 10.1: Descriptive statistics of the SOC samples (g kg^{-1}), used for model training. N Sites = number of cropland sites used in this study; N Samples = number of training samples; SD = standard deviation; IQR = interquartile range.

Soil Data	N Sites	N Samples	Mean	Median	Min	Max	SD	IQR
ALL	796	1723	17.3	15.6	4.9	126.4	8.8	7.3
LTM	100	696	18.1	16.5	6.6	70.0	8.0	7.5
HDB	331	662	16.4	15.1	4.9	47.7	6.2	7.1
BZE	365	365	17.6	14.7	6.0	126.4	13.0	6.8

Samples from an additional soil organic matter monitoring program (Humus Database = HDB) were used to complement the training data. This program includes a total of 331 cropland locations that were selected to represent the full spectrum of soil textures in Bavaria (Figure 10.1 c). Each of the HDB sites was sampled twice, within exactly ten years. The first sampling campaign was

conducted from 2001 to 2008 (Figure 10.1 d). In each location, 7 m^2 plots were marked with stationary magnets, and six composite samples were taken at a depth of 0 – 15 cm. SOC contents were determined with the same methods as described above.

Samples from the first German Soil Inventory (Bodenzustandserhebung Landwirtschaft = BZE) were added to further improve the spatial coverage of the model (Poeplau et al., 2020a). They include a total of 365 cropland sites in Bavaria, that were selected based on a regular 8x8 km grid (Figure 10.1 c). The sites were sampled from 2013 to 2015 and have so far not been resampled. For each location, a 1 m profile was dug, and a distributed composite sample was collected at a depth of 0 – 10 cm. For laboratory analyses, the samples were oven-dried (40 °C) and sieved to 2 mm. The SOC content (g kg^{-1}) was determined with dry combustion and by calculating the difference between total C and inorganic C contents (Poeplau et al., 2020b).

Furthermore, the LTM sites were grouped based on the mean SOC content (above/below 15 g kg^{-1}; n = 64/36), tillage practice (tillage/reduced tillage; n = 56/44), and main soil texture (loam/sand/silt/clay; n = 47/30/13/10) to analyze the influence of different soil and management factors on the predicted and measured SOC trends. These factors were selected as they were (1) found to be connected to SOC trends in previous studies (Wiesmeier et al., 2024), (2) can significantly influence the soil reflectance signal in EO-based SRC (Demattê et al., 2018), and (3) were available for all 100 LTM sites. The SOC threshold was selected according to Wiesmeier et al. (2013), who reported a median SOC content of 15 g kg^{-1} for cropland soils in Bavaria.

2.3 Measured SOC trends and changes

Based on the seven repeated samples from 1986/88 to 2020/22, linear regression analyses were conducted to derive the long-term SOC trends for each of the 100 LTM sites. Using the sampling years (x) and the measured SOC contents (y), linear models were calculated to obtain the measured SOC trends (i.e. the slopes) in g kg^{-1} $year^{-1}$. In addition, correlation coefficients (Pearson r) and the corresponding p-values were derived from each of the measured linear regressions. Both values, SOC trends (g kg^{-1} $year^{-1}$) and correlation coefficients (Pearson r), were later used to evaluate the temporal accuracy of the predictions (section

2.6.2). For further analyses, the measured SOC trends were categorized into three groups based on the Pearson r and p-values: increasing SOC trends: $r > 0.5$ & $p < 0.1$; decreasing SOC trends: $r < -0.5$ & $p < 0.1$; stable SOC trends: $r > -0.5$ & $r < 0.5$ & $p > 0.1$. In addition to the LTM long-term trends, SOC changes between the two HDB sampling periods (2001/8 to 2011/18) were calculated for each site and used for additional validation of the temporal model accuracy.

2.4 Model input

2.4.1 Remote sensing

Multispectral Landsat (5, 7, 8, 9) and Sentinel-2 (A, B) images were combined to generate a time series of SRC for the observation period from 1986 to 2022. An overview of the workflow is shown in Figure 10.2. Data processing and radiometric corrections were conducted using the Framework for Operational Radiometric Correction for Environmental Monitoring (FORCE) (Frantz, 2019). The FORCE Level-2 module was used to create analysis-ready bottom-of-atmosphere data. It includes atmospheric, topographic, as well as nadir BRDF adjustments that are described in detail in Frantz (2019). Clouds and shadows were removed using an optimized version of the FMASK algorithm (Frantz et al., 2018). We included all observations from 1986 to 2022 that were acquired between March to October and had cloud coverage below 70 %. The images were converted to a common resolution of 30 m using nearest-neighbor resampling. Six compatible spectral bands of the Landsat and Sentinel-2 sensors were used for further analyses: Blue, Green, Red, Near-Infrared (NIR), and Shortwave Infrared (SWIR1/SWIR2). To ensure that the spectral values are consistent over time and different sensors, the Landsat bands were linearly adjusted to the Sentinel-2 sensor using the method described in Okujeni et al. (2024).

In the next step, SRC were generated for each spatial and temporal location of the available soil samples (Figure 10.1) (Broeg et al., 2024c). For this, the bare soil observations were filtered from the cloud-free time series using the Normalized Difference Vegetation Index (NDVI) (Tucker, 1979), the Normalized Burn Ratio 2 (NBR2) (Van Deventer et al., 1997), and the respective thresholds of 0.4 and 0.16 (Broeg et al., 2024a). We defined a time window of ±4 years to derive an average SRC for each sampling date. This was done to increase the number of bare soil observations and to improve the quality of SRC, especially for earlier

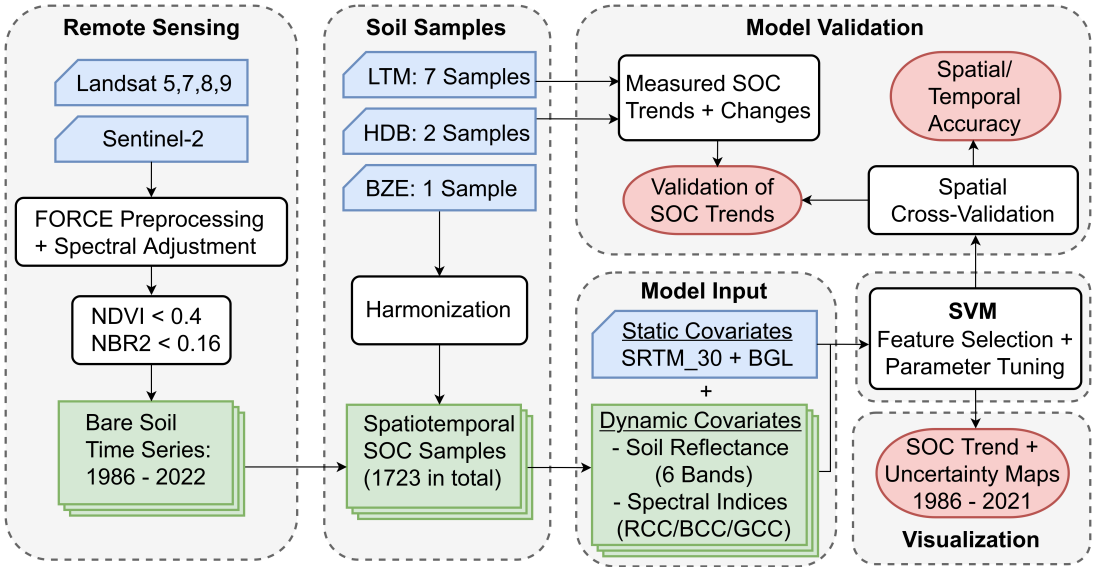

Figure 10.2: (a) Schematic overview of the model workflow: Input data is shown in blue; Intermediate products are shown in green; Results are shown in red.

years of the observation period with less available satellite data (Vaudour et al., 2021; Zepp et al., 2023). Figure 10.3 illustrates the increasing availability of bare soil observations at the sampling locations as well as an example of the time window used to derive the SRC for one soil sample in time. Based on the selected observations, the mean soil reflectance was calculated for each band and used as input for the spatiotemporal model (Figure 10.2). The blue, green, and red chromatic coordinates (BCC, GCC, and RCC) were calculated to provide further information on relative soil color, in addition to absolute RGB values (Gillespie et al., 1987).

2.4.2 Static covariates

Static covariates were added to the SOC model to provide spatial context and complement the dynamic SRC. We included a basic elevation model, provided by the Shuttle Radar Topography Mission with a resolution of 30 m (SRTM_30) (Farr and Kobrick, 2000). Furthermore, we added the main soil-landscape units in Bavaria (BGL) with 16 classes (1:5,000,000), as provided by the Federal Institute for Geosciences and Natural Resources (BGR, 2008). The vector data was rasterized to match the common model resolution of 30 m and the categorical data was numerically encoded using an ordinal scale, provided by the BGR.

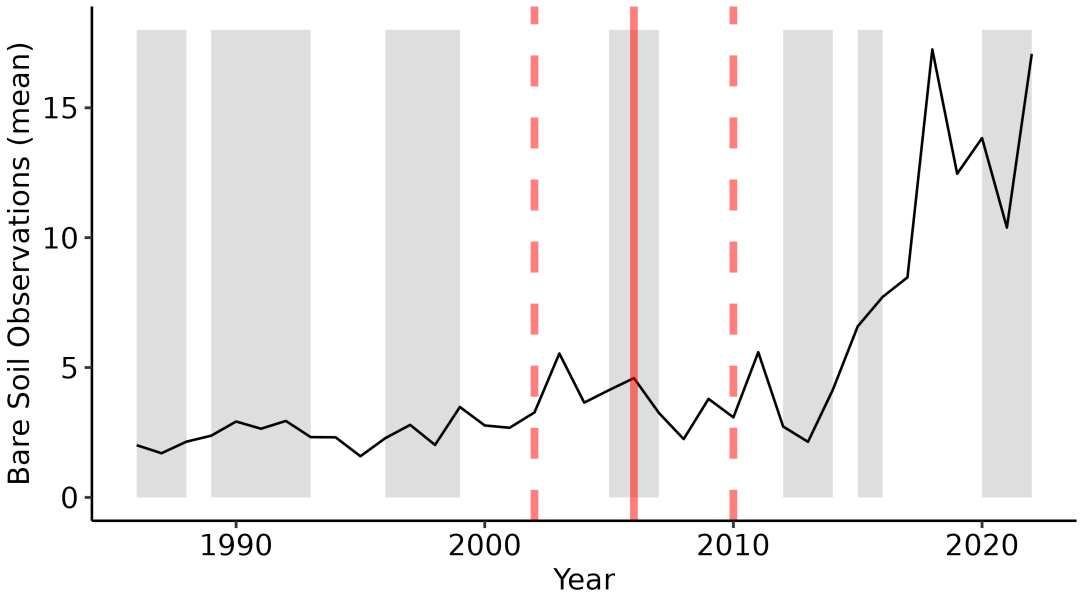

Figure 10.3: Mean number of bare soil observations (Landsat, Sentinel-2) per year from 1986 to 2022 (black line), available at the SOC sampling locations. The seven LTM sampling periods are shown in grey. Red line = example of a single SOC measurement in time (2006); Dashed lines = Time window of ± 4 years, used to derive the SRC for the example sample.

2.5 Spatiotemporal SOC model

We combined the dynamic and static covariates with the available soil samples to train a spatiotemporal SOC model for the observation period from 1986 to 2022 (Broeg et al., 2024c). As described in section 2.4.1, we temporally matched bare soil observations and the dates of the SOC samples to generate an SRC data cube (Figure 10.2). Based on this, a single spatiotemporal Support Vector Machine (SVM) was trained using the SOC samples from all soil datasets and years (Cortes and Vapnik, 1995). A forward feature selection was conducted to determine the best combination of covariates and reduce the risk of overfitting (Meyer et al., 2018). The following features were selected for the final model: RED, RCC, BCC, NBR2, SRTM_30, and BGL. The hyperparameter tuning was carried out using the hyperband method, as described by Li et al. (2017). The kernel of the SVM was set to "radial" and the following hyperparameter settings were selected for the model: Gamma = 0.055; Cost = 5.

2.6 Accuracy Assessment

The accuracy assessment of the SVM was conducted using a spatial 10-fold cross-validation to prevent repeated SOC samples from being used for model training and validation at the same time (Meyer et al., 2018). Based on the results, we calculated the Root Mean Square Error (RMSE), R^2-values, and Lin's Concordance Correlation Coefficient (CCC) of the predictions. The Ratio of Performance to Deviation (RPD) was included to improve the comparability to other studies (Mcbratney and Minasny, 2013).

2.6.1 Temporal Model Accuracy

Repeated SOC samples from the LTM and HDB campaigns were used to analyze the consistency of the temporal model accuracy throughout the observation period. For this, the results of the spatial cross-validation were grouped into the corresponding sampling periods to quantify and compare the temporal model accuracy (RMSE, R^2, CCC, and RPD). This includes seven SOC measurements for the LTM samples from 1986/88 to 2020/2022, and two measurements for the HDB samples from 2001/2008 to 2011/2018. In addition to the accuracy metrics, the prediction residuals of each sample period were compared to check if any temporal biases were present in the results. As the LTM only includes a limited number of sites with high SOC contents, the comparison of the temporal model accuracy was conducted twice, to reduce the influence of outliers on the results: Including all samples; and only considering samples with SOC contents below 40 g kg^{-1}.

The signal-to-noise ratio was estimated based on the LTM time series to further analyze the relationship between model uncertainty and temporal SOC variability. For this, SOC changes between the first (1986/88) and the subsequent measurements (1989 – 2022) were quantified for each sampling year, and site. The standard deviation of the residuals was calculated to describe the temporal SOC change within the observation period. Next, the cross-validated prediction residuals were quantified for each of the seven sampling periods. Based on the results, the standard deviation was calculated to derive spatial prediction uncertainty over time. Last, the signal-to-noise ratios were estimated for each sampling period by dividing the SOC change signal (i. e. standard deviation of the temporal residuals) and the prediction uncertainty (i. e. standard deviation of the predic-

tion residuals). For further analyses, the calculation of the signal-to-noise ratio was repeated for the individual soil groups (mean SOC above/below 15 g kg^{-1}), described in section 2.2.

2.6.2 Accuracy of the predicted SOC trends and changes

The cross-validated predictions were directly compared to the LTM measurements to evaluate the model's ability and accuracy to detect long-term SOC trends and changes. For this, the predicted SOC trends (g kg^{-1} $year^{-1}$) and Pearson correlation coefficients (r) were generated with the same methods described in section 2.3. Calculations were performed for each pixel intersecting the LTM plots (section 2.2) and the median values were taken to summarize the results for each site. The accuracy and significance of the predicted SOC trends and correlation coefficients were derived using the R^2 and corresponding p-values, as well as an ANOVA and Tukey's HSD. Short-term SOC changes were verified using repeated samples from the HDB. As all sites were resampled in the same ten-year interval, measured and predicted SOC changes were directly compared to each other.

2.6.3 Visualization of the model

SOC contents in Bavaria were predicted from 1986 to 2021 to visualize the results of the spatiotemporal model (Broeg et al., 2024c). For this, we generated an SRC every five years (1986, 1991, 1996 2001, 2006, 2011, 2016, and 2021) and applied the spatiotemporal SOC model, described in section 2.5 (Figure 10.2). In addition, uncertainty maps were generated by calculating the standard deviation of the ten SOC predictions, produced by the cross-validation models (section 2.6). Linear regressions (g kg^{-1} $year^{-1}$) were calculated based on the SOC maps from all years to derive SOC trends for each cropland pixel in the research area.

3 Results

3.1 Overall model accuracy

Based on the results of the spatial cross-validation, the developed spatiotemporal SOC model produced an overall R^2 value of 0.61 and an RMSE of 5.6 g kg^{-1}

(Table 10.2 and Figure 10.4). Looking into the different soil datasets individually, the results ranged between 0.55 and 0.69 (R^2), and 4.2 and 7.5 g kg^{-1} (RMSE). According to the CCC (0.77) and RPD (1.73), the BZE data produced the highest prediction accuracy. The repeated soil samples from the LTM and HDB showed similar results with R^2 values of 0.55 and 0.57 and RPD values of 1.44 and 1.49. In general, the HDB samples had the lowest SOC range (Table 10.1) and showed the best RMSE (4.2 g kg^{-1}). As illustrated by the high CCC value (0.72), the predictions were closer to the 1:1 line, compared to the LTM samples (0.65). The regression lines show a general tendency for underpredictions, especially for high SOC contents (Figure 10.4).

Table 10.2: Cross-validation accuracy of the three spatiotemporal SOC datasets (LTM, BZE, HDB), and the combined samples (ALL).

Soil Data	N Samples	R^2	CCC	RMSE	RPD
ALL	1723	0.61	0.72	5.58	1.57
LTM	696	0.57	0.65	5.56	1.44
HDB	662	0.55	0.72	4.15	1.49
BZE	365	0.69	0.77	7.54	1.73

3.2 Temporal changes in model accuracy

Based on the seven repeated LTM samples from 1986/88 to 2020/22, the SOC model exhibits a high temporal consistency for the samples below 40 g kg^{-1} (Figure 10.5). The R^2 shows slight fluctuations over time and ranges between values of 0.48 (1996/99) and 0.61 (2020/22) (Figure 10.5 a). According to the RMSE, the prediction accuracy incrementally improved by around 8 % throughout the observation period (~ 3.9 to 3.6 g kg^{-1}). Similar tendencies are visible for the CCC and RPD, especially for later years after 2006. In comparison, the model produced slightly better results when considering all samples, with R^2 values between 0.57 (2005/07) and 0.67 (2015/16) (Figure S1). While the CCC and RPD showed similar increasing tendencies, the RMSE mostly remained stable with values around 5 g kg^{-1}. Based on the samples below 40 g kg^{-1}, the prediction residuals are normally distributed and reveal a high temporal consistency (Figure 10.5 b). For most years, a slight tendency for underpredictions is visible in the data. Scatterplots, with the results from each sampling period, are shown in

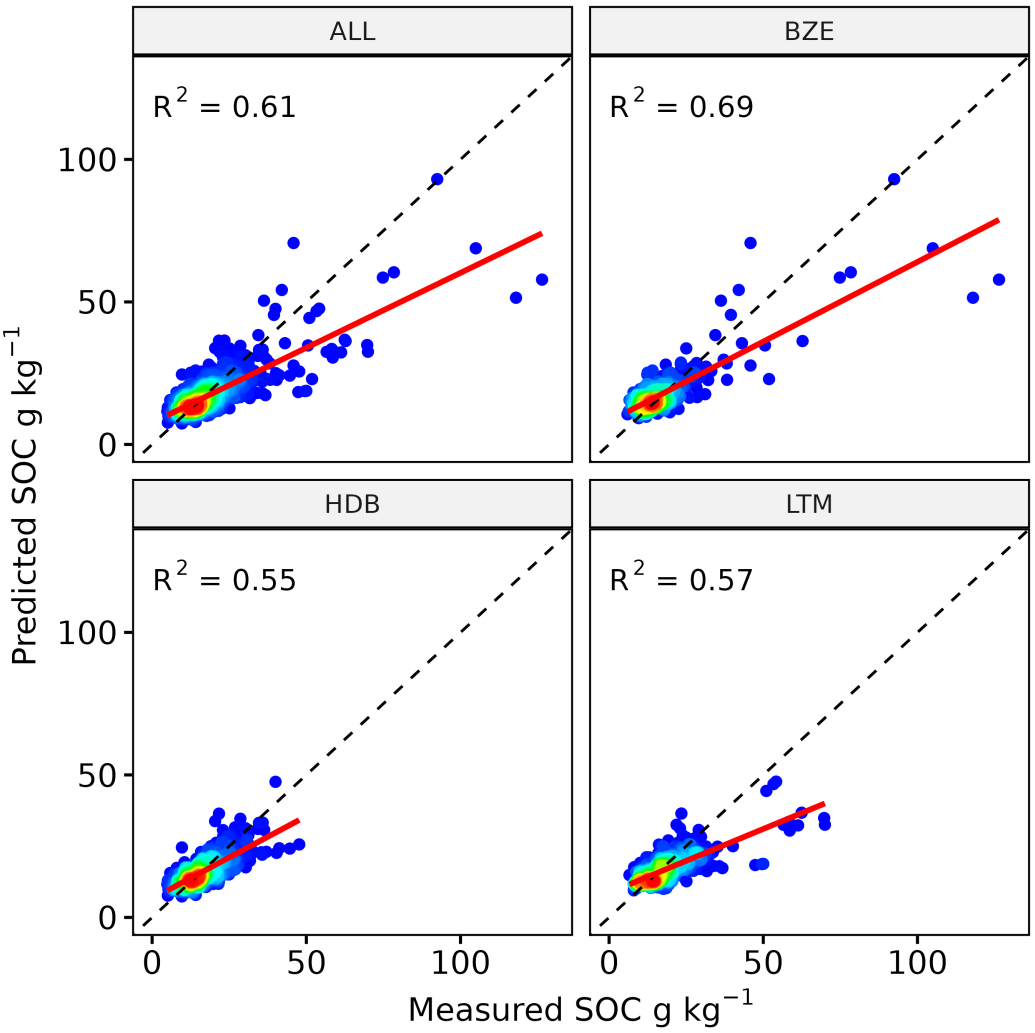

Figure 10.4: Cross-validated predictions of the three spatiotemporal SOC datasets (LTM, BZE, HDB), and the combined samples (ALL). Regression lines are shown in red; 1:1-line is shown in black (dashed).

Figure S2. Comparing the results of the two HDB sampling periods from 2001 to 2008, and from 2011 to 2018, only slight temporal variations are visible (Figure S3). The R^2 values remain mostly constant (0.56/0.55), while the RMSE decreased slightly, from 4.3 to 3.9 g kg^{-1}.

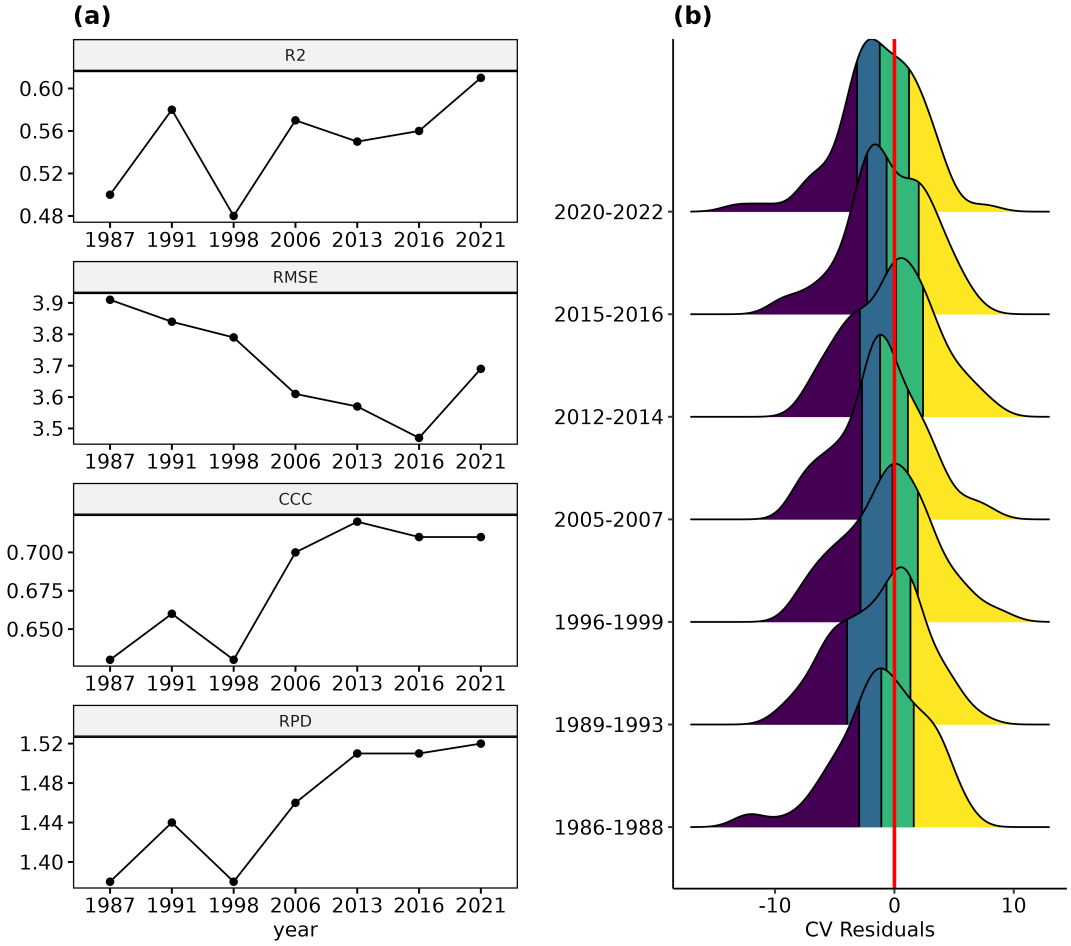

Figure 10.5: (a) Temporal trends of the model accuracy (R^2, RMSE, CCC, and RPD), based on the repeated LTM samples below 40 g kg^{-1}. (b) Distributions of the prediction intervals; Quartiles are illustrated with colors.

3.3 Signal-to-noise ratio

Overall, the measured SOC change steadily increased since the initial SOC samples were taken in 1986/88 (Figure 10.6 a). This is illustrated by the standard deviation of the temporal residuals, which doubled from around 2 (1990) to 4 g kg^{-1} in 2020. In contrast, the standard deviation of the spatial prediction residuals (i.e. the prediction uncertainty) mostly remained stable throughout the

observation period with values around 4 g kg^{-1}. As a result, the signal-to-noise ratio incrementally improved over time and reached a value of around 1 in 2016, roughly 30 years after the initial SOC samples were collected (Figure 10.6 b).

For further analysis, the calculation of the signal-to-noise ratio was repeated for the LTM sites with mean SOC contents above and below 15 g kg^{-1} (section 2.2). In both cases, the results had similar but slightly shifted trends (Figure 10.6 a). Sites with mean SOC contents above 15 g kg^{-1} showed larger SOC changes over time (temporal residuals) and generally produced higher signal-to-noise ratios. The opposite is true for the remaining sites with SOC values under 15 g kg^{-1}. Although the spatial residuals were much lower, the temporal SOC changes were too small and resulted in ratios under 1, even for long time series of 35+ years (Figure 10.6 b).

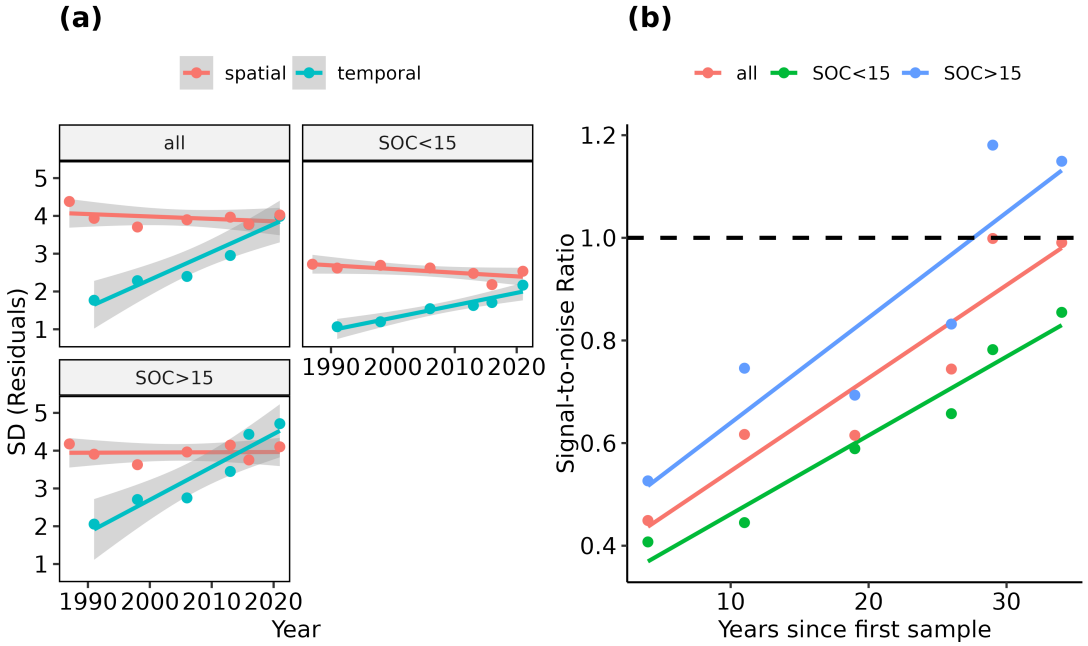

Figure 10.6: Standard deviation (SD) of the spatial (prediction uncertainty) and temporal (SOC changes) residuals of the LTM samples (a). The plots were grouped according to the mean SOC contents of the sites: above/below 15 g kg^{-1}. Signal-to-noise ratio throughout the observation time, as derived from the temporal and spatial SOC residuals (b).

3.4 Prediction of SOC trends and changes

Out of the 100 observed LTM sites, exactly 50 showed increasing and decreasing SOC trends (Figure 10.7 a,b). In around 70 % of all sites, the Pearson r values were below -0.5 or above 0.5. A similar distribution was produced by our model. Here, 47 sites showed positive trends while the r values were below -0.5 or above 0.5 in around 60 % of the cases. The direct validation, however, only revealed a low ($R^2 = 0.16$) but significant ($p < 0.0001$) accuracy (Figure 10.7 c). The measured SOC trends showed a normal distribution and median values close to 0 g kg^{-1} $year^{-1}$ (Figure 10.7 b). For most LTM sites, the values were in the range between -0.2 and 0.2 g kg^{-1} $year^{-1}$, with few outliers in both directions. In general, the predicted values showed a similar distribution and range, with a median value slightly below 0 g kg^{-1} $year^{-1}$. Similar to the r values, a direct comparison of the trends only yielded a low R^2 value of 0.17 (Figure 10.7 d).

Measured and predicted SOC trends were classified into three groups to further analyze the results: decreasing, stable, and increasing (section 2.3). In general, it was possible to differentiate between increasing and decreasing trends with high significance ($p < 0.001$) (Figure 10.8 a). While significant results were also visible between the stable and decreasing classes ($p < 0.05$), no differences were found between the stable and increasing classes. As illustrated by the confusion matrix in Figure 10.8 b, misclassifications between the decreasing or increasing classes occurred in only 5 % of the observed sites.

The results of the short-term SOC changes, based on the HDB plots, are shown in Figure S4. Within the 10-year intervals, most sites showed SOC changes between -4 and 4 g kg^{-1}. The measured values were normally distributed and showed a median value close to 0 g kg^{-1}. Although the model generally produced a similar distribution and range, no significant correlation between the measured and predicted SOC changes was found.

3.5 Influence of additional factors on long-term SOC trends

The LTM sites were categorized based on their mean SOC content, tillage intensity, and soil texture to further analyze the results of the long-term SOC trends (section 2.2). In general, the sites with a mean SOC content above 15 g kg^{-1} showed lower correlation coefficients (Figure 10.9 a), and a broader distribution

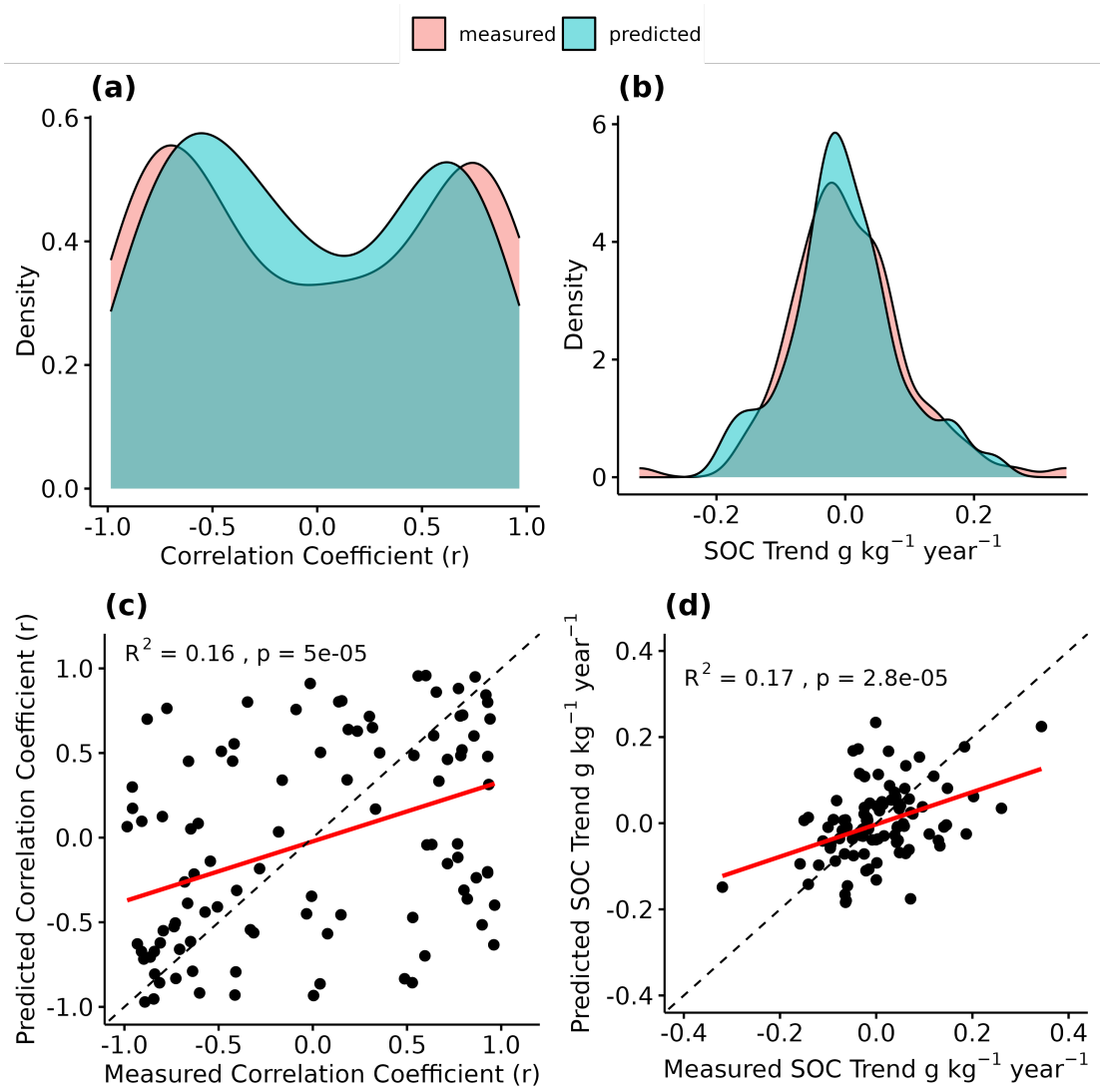

Figure 10.7: Distribution of the measured and predicted correlation coefficients (Pearson r) (a) and SOC trends (g kg^{-1} year^{-1}) (b). Direct comparison of the results (c+d). Regression lines are shown in red; 1:1-lines are shown in black (dashed).

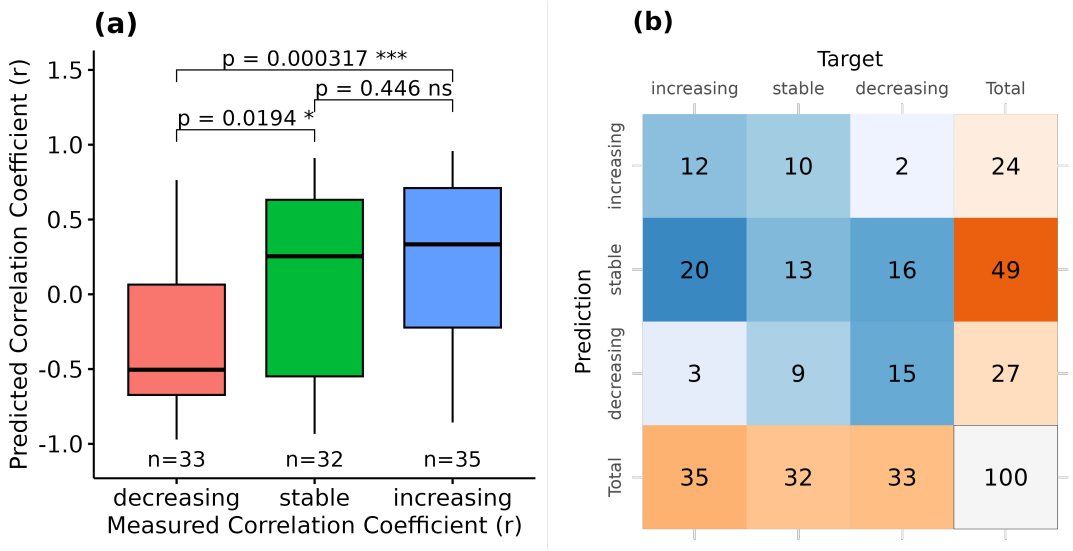

Figure 10.8: Boxplots of the measured and predicted correlation coefficients (Pearson r) (a). Decreasing: r < -0.5; Stable: r > -0.5 and < 0.5; Increasing r > 0.5. Confusion matrix of the predicted and measured correlation coefficients (Pearson r) (b).

of SOC trends (Figure 10.10 a). In comparison, sites with SOC contents below 15 g kg^{-1} often showed SOC trends close to 0 g kg^{-1}, indicating no or little SOC changes (Figure 10.10 a). In general, the model produced much better results for the LTM sites with mean SOC contents above 15 g kg^{-1} (R^2 = 0.27/0.4), while the remaining sites (SOC < 15 g kg^{-1}) showed no significant results (Figure 10.9 c & Figure 10.10 c).

Overall, the presence of reduced tillage resulted in a higher number of positive SOC trends (Figure 10.9 b). This was the case for both, the predicted and measured values. In comparison to the sites with intensive tillage, the trends showed a broader value distribution and higher absolute SOC changes (Figure 10.10 b). In general, the spatiotemporal model produced much higher accuracies for the sites with reduced tillage, with R^2 values of 0.26 and 0.38 (Figure 10.9 d & Figure 10.10 d).

Compared to the other factors, the influence of the main soil texture on the results is less clear (Figure S5). Loam and sand soils, which include most of the LTM sites, show similar results to the overall accuracy (R^2 = 0.12/0.16). In contrast, LMT sites with clay soils produced a significantly higher (R^2 = 0.6), while silt soils produced a lower accuracy (R^2 = 0.09). Both of these classes are,

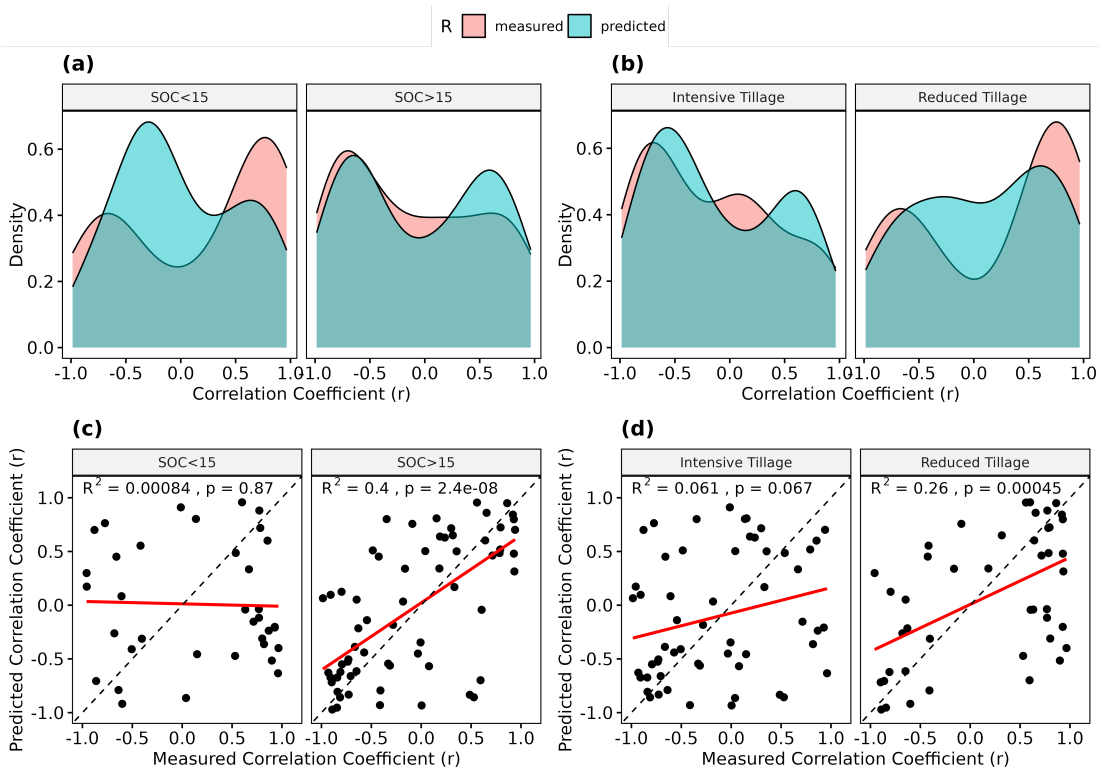

Figure 10.9: Distribution of the measured and predicted correlation coefficients (Pearson r), grouped by mean SOC content (above/below 15 g kg^{-1}) and tillage regime (intensive/reduced) (a+b). Corresponding scatterplots of the results (c+d). Regression lines are shown in red; 1:1-lines are shown in black (dashed).

however, most likely influenced by the low number of available sites (n = 10/13).

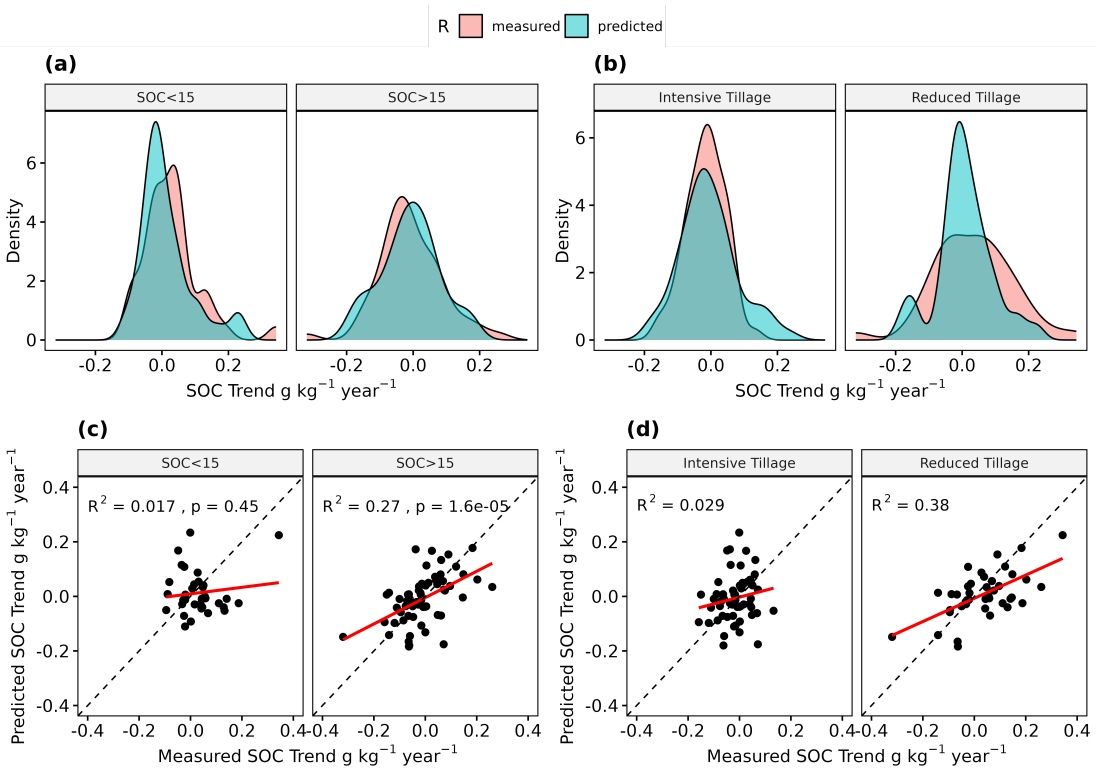

Figure 10.10: Distribution of the measured and predicted SOC trends (g kg^{-1} year^{-1}), grouped by mean SOC content (above/below 15 g kg^{-1}) and tillage regime (intensive/reduced) (a+b). Corresponding scatterplots of the results (c+d). Regression lines are shown in red; 1:1-lines are shown in black (dashed).

3.6 SOC predictions and trend map

The spatiotemporal model was used to predict the SOC content in Bavaria every five years from 1986 to 2021. Looking at the most recent SOC map in 2021, the highest SOC contents are present in the south and connected to the large stream, originating in the Alps (Figure 10.11 a). A similar spatial distribution is visible in the corresponding uncertainty map (Figure S6). An example of the differences, visible between the individual SOC maps, is provided in Figure S7. Based on the individual predictions, SOC trends were calculated for the full observation period from 1986 to 2021 (Figure 10.11 b). While increasing trends are present across all regions, the resulting map shows that the SOC of most cropland soils in Bavaria is stable or slightly decreasing. Regional patterns as well as small-scale differences are visible throughout the research area. As illustrated in Figure 10.11

b, contrasting SOC trends were predicted at a field scale, even for neighboring cropland sites with similar SOC contents.

4 Discussion

4.1 Overall model accuracy

We combined repeated soil samples from 1986 to 2022 and EO data to train a spatiotemporal SOC model for Bavaria. The overall R^2 value of 0.61 is comparable to static large-scale SOC maps in Germany (Broeg et al., 2024a), and Bavaria (Zepp et al., 2021a), illustrating that it is possible to construct spatiotemporal SOC models without compromising prediction accuracy (Chen et al., 2022). Overall, the three soil data sets yielded varying results (Figure 10.4 & Table 10.2). Comparing the R^2 values of the HDB (0.57) and BZE (0.69), it is clear that the higher range and standard deviation of the BZE samples (Table 10.1) significantly influenced the results. While BZE was sampled on a regular grid (Figure 10.1 c) and included more SOC values above 40 g kg^{-1}, the HDB sites were selected to represent the soil textures in Bavaria. In both cases, the median SOC values are similar (Table 10.2), and the CCC values (0.77/0.72) support the conclusion that the overall accuracy was high (Table 10.2). Although the LTM includes a lower number of sites, they express a high spatial representativeness and are evenly distributed across the research area (Figure 10.1 c). Considering that the samples were collected within a time frame of almost four decades, including years with limited EO data (Figure 10.3), the model produced a good but slightly lower overall accuracy (CCC = 0.65 and RPD = 1.44). Similar to previous studies, we found a limited accuracy and tendency for underpredictions at high SOC contents (Figure 10.4) (Feeney et al., 2022; Zepp et al., 2021a). Considering the low availability of SOC samples above 40 g kg^{-1}, it can be concluded that additional soil data is needed to improve the prediction accuracy for high SOC contents. Further tests are necessary to evaluate if EO data provides sufficient information on the full range of SOC values, including organic soils, or needs to be extended by other covariates (Feeney et al., 2022).

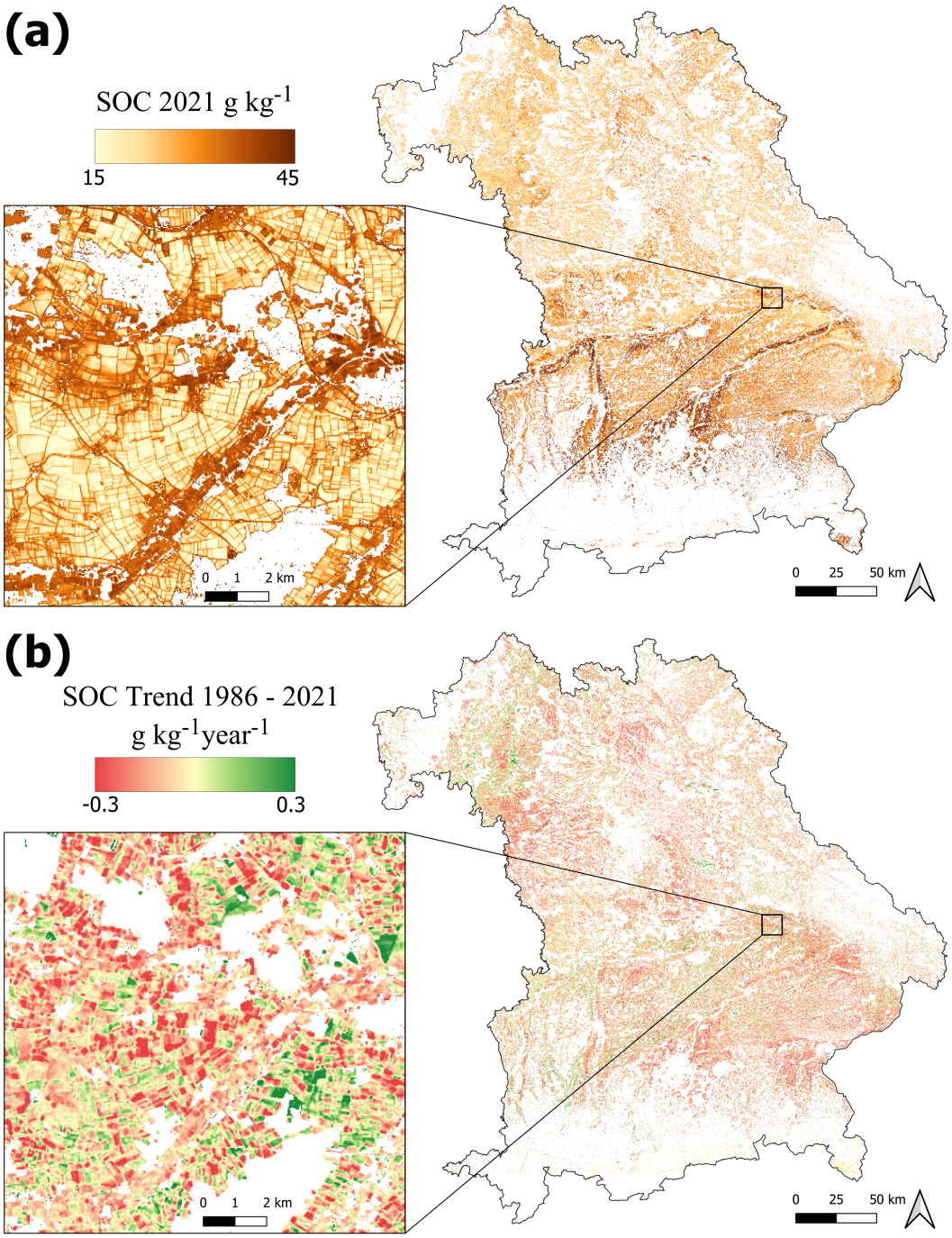

Figure 10.11: (a) Map of the predicted cropland SOC content in 2021 (g kg^{-1}). (b) SOC trend map (g kg^{-1} $year^{-1}$), calculated based on the individual SOC predictions from 1986 to 2021. Map lines delineate study areas and do not necessarily depict accepted national boundaries.

4.2 Temporal changes in model uncertainty

Quantifying the temporal model uncertainty is an important prerequisite in understanding and improving the prediction of SOC trends. As pointed out by Croft et al. (2012), variances between different satellite sensors represent a challenge to the transferability and accuracy of spatiotemporal SOC models. Accordingly, spectral adjustments were conducted in this study and several other works and used to preprocess the remote sensing data for the SOC models (Meng et al., 2024; Venter et al., 2021). Without a validation based on repeated SOC samples, however, it is difficult to provide unbiased information about the consistency of the temporal model accuracy. Our results show that the model accuracy and residual distribution are comparable for all seven sampling periods from 1986 to 2022 (Figure 10.5). This is an important prerequisite, as a temporal bias could result in artifacts when deriving the SOC trends based on the model predictions.

With an RPD of 1.44, the validation of the LTM samples revealed a slightly lower accuracy compared to the other soil data sets (Table 10.2). Taking into account the temporal model accuracy of the LTM samples, different explanations for this can be derived: First, improvements of the RMSE and CCC over time (Figure 10.5) seem to coincide with the strongly increasing number of bare soil observations (Figure 10.3). This is especially true for the years after 2015, following the launch of the Sentinel-2 satellites, and could have positively affected the quality of the SRC. For the more recent sampling periods, the RPD of the LTM samples increased to values above 1.5 which are comparable to the other soil data sets (Figure 10.5 & Figure S1). Second, slight adjustments in the sampling design of the LTM plots could have led to better results in later years. This is supported by the fact that the coefficient of variation for the subsamples has been decreasing throughout the years (LfL, 2022). In summary, our results underline that the generation of SRC based on a moving window is suitable to increase the temporal model consistency, especially for years with varying availability of EO data. The slightly lower accuracy in the early years could have, however, negatively affected the detection of long-term SOC trends. Further research is necessary to test if the accuracy can be improved without sacrificing on the temporal model resolution.

4.3 Signal-to-noise ratio

Using the repeated LTM samples, we were able to show that the signal-to-noise ratio of measured SOC changes and model uncertainty significantly increased throughout the observation period (Figure 10.6 b). These findings highlight the large potential of our approach to monitoring SOC trends but also underline the necessity to analyze long observation periods. Based on our estimates, a minimum number of 25 years was necessary to obtain signal-to-noise ratios with values above 1. Considering that SOC changes in cropland soils are generally very slow (Ciais et al., 2010), these findings support the conclusion that it is easier to predict long-term SOC trends rather than short-term changes. Based on experiments, Smith (2004) demonstrated that in soils with low carbon inputs, up to 15 years are necessary to measure significant changes. This could be one reason why we were unable to verify the SOC changes in the HDB samples that were repeated after 10 years (Figure S4). Furthermore, we were also able to demonstrate a connection between the mean SOC contents and the ability to predict SOC trends. In soils with contents above 15 g kg^{-1}, the measured SOC changes outweighed the higher prediction uncertainty and increased the signal-to-noise ratio (Figure 10.6). In contrast, soils with lower SOC contents also produced a lower signal-to-noise ratio, increasing the necessary time frame and prediction accuracy to detect changes. Summarizing the results, we were able to demonstrate three key factors that influence the ability to predict SOC trends and should be reported when applying spatiotemporal SOC models: (1) The mean SOC contents and absolute SOC changes; (2) the temporal model uncertainty; and (3) the length of the observation period.

4.4 Prediction and validation of SOC trends and changes

There is an increasing demand for EO data and spatially explicit information on SOC trends with existing and upcoming carbon farming schemes and their role in contributing to carbon removal and climate change mitigation (Amelung et al., 2020; Lehmann et al., 2020; Paustian et al., 2019). To increase the useability of EO-based models, it is necessary to validate the predictions with measured SOC trends and changes. Due to the low availability of large-scale soil inventories with high-quality repeated soil samples, however, recent studies rely on indirect measures to validate the accuracy of the predicted SOC changes (Heuvelink et

al., 2021; Meng et al., 2024; Venter et al., 2021). Our findings highlight that the general model uncertainty is insufficient to verify spatiotemporal SOC models and could lead to an over-optimistic interpretation of the results. Although the cross-validation produced good results in all cases ($R^2 = 0.61$), the direct validation of the predicted long-term SOC trends revealed a much lower accuracy ($R^2 = 0.16/0.17$) (Figure 10.7). This is the case even though the predicted and measured SOC trends share a nearly identical distribution of values.

Despite the high uncertainty, our results underline that EO data is a valuable asset for deriving long-term trends in cropland soils. As shown in Figure 10.8 b, the model rarely predicted trends opposite to the measurements, illustrating that a clear SOC signal is present in the data. In general, the predicted r values showed significant differences for sites with increasing and decreasing SOC trends (Figure 10.8 a). Using the validated spatiotemporal model, it is possible to generate high-resolution information on the cropland SOC dynamics from the year 1986 onwards. As illustrated in Figure 10.11 b, SOC trends can be derived and analyzed at a field scale. Even though the results show that the signal is not strong enough to detect SOC changes in all cases, they significantly enhance the data density and understanding of the cropland soil dynamics within the observed timeframe. Thus, for the first time, it will be possible to estimate long-term SOC trends for large regions in very high resolution.

4.5 Influence of additional factors on long-term SOC trends

Considering the high uncertainty associated with the predicted SOC trends, it is necessary to further analyze and understand the contributing factors. In general, the mean SOC contents of the sites influenced the SOC trends as well as the accuracy of the predictions. As illustrated in Figure 10.9, the observed LTM sites with low SOC contents predominantly showed increasing trends, while the opposite is true for sites with contents above 15 g kg^{-1}. Similar results have been reported in other studies and can be partly linked to historic land use changes that have influenced SOC dynamics and trends for decades (Bellamy et al., 2005; Heikkinen et al., 2013; Lettens et al., 2005). Furthermore, sites with SOC contents above 15 g kg^{-1} showed higher accuracy in the predicted trends ($R^2 = 0.4$), while no significant results were found for the remaining sites (Figure 10.9). Similarly to the signal-to-noise ratio (Figure 10.6), absolute SOC

changes may be too small to be detectable by the remote sensing data, resulting in low accuracy. This is supported by the distributions of the predicted and measured values in Figure 10.10 and the fact that SOC trends were often close to 0 g kg^{-1} $year^{-1}$. In general, our findings show that the model is more likely to detect the changes in soils with larger C-stocks, which are also at higher risk of SOC loss (Bellamy et al., 2005).

Additionally, we analyzed the influence of tillage practices on the predicted and measured SOC trends. Several studies reported increasing SOC contents on surface soil layers, following the implementation of reduced tillage measures on cropland soils (Bai et al., 2019; Haddaway et al., 2017; Ogle et al., 2019). It is often argued that this is due to the accumulation of the SOC in the uppermost centimeters of the topsoil, while the SOC content below is decreasing with little effect on the overall SOC stocks of the total soil profile (Baker et al., 2007; Meurer et al., 2018). As the SOC samples were taken at a depth of 0-15 cm, sites with reduced tillage showed increasing SOC trends in most cases (Figure 10.9). Most of these SOC trends were correctly predicted as increasing and the accuracy of the predictions is significantly higher, compared to the sites with conventional tillage (Figure 10.9). This result is surprising since reduced tillage can introduce organic soil cover (litter, crop residuals, etc.) and influence soil reflectance beyond the SOC change signal (Dvorakova et al., 2022). It is possible that these changes in the surface conditions lead to higher SOC predictions and generally increase the chances of predicting positive trends. Our findings underline the importance of tillage on the measured and predicted SOC trends if only the upper centimeters of the soils are observed. Overall, the results show that EO is a powerful tool for deriving topsoil SOC changes, while these do not correspond to the total SOC stocks in all cases.

4.6 Impact, limitations, and further developments

Carbon removal from the atmosphere with C sequestration in soils can result in negative greenhouse gas emissions in agriculture and is developing as a promising business model to support efforts for climate change mitigation (Don et al., 2024b; European Commission, 2022). The monitoring of SOC in carbon farming programs has been outlined as a major challenge for success and to avoid greenwashing (Paul et al., 2023). The transaction costs for measuring SOC changes

are often too high and jeopardize the establishment of carbon farming on a large scale (Don et al., 2024b). Our results show that the integration of remote sensing data is a promising option to reduce the cost of SOC monitoring and generate high-resolution information on cropland SOC dynamics beyond static SOC maps. Based on the comprehensive network of 100 LTM sites, we were able to provide a detailed verification of the predicted SOC trends and changes (Vaudour et al., 2022). These are the first steps in creating an integrated high-resolution soil monitoring system that directly profits from the exploding availability of EO data (Paustian et al., 2019; Smith et al., 2020).

SRC reduce the satellite signal to bare soil observations and provide direct information on the soil state. Our findings demonstrate that they contain valuable information on SOC changes and should be a central part of spatiotemporal soil models. An increasing number of bare soil observations (Figure 10.3) offer the opportunity to further improve the signal-to-noise ratio and model accuracy (Figure 10.5). However, multiple challenges regarding the prediction and validation of SOC trends remain: Even though we had access to comprehensive soil monitoring data with seven measurements from 1986 to 2022, we were not able to capture and describe non-linear SOC trends if they occurred. Similar to Venter et al. (2021), we exclusively relied on linear models to predict and validate the long-term SOC trends. It has been shown, however, that SOC trends can change over time and alternate between SOC gains and losses (Janzen, 2006; Wiesmeier et al., 2019). To improve the accuracy and predict non-linear trends, it would be therefore necessary to increase the temporal resolution of the model and soil monitoring data. Due to the limited availability of EO data in earlier years, we had to rely on relatively broad intervals (± 4 years) to generate the dynamic SRC. This issue could, however, significantly decrease in the future following the increasing number of satellites and bare soil observations.

Still, our results based on the repeated HDB samples show that the prediction of short-term SOC changes based on EO data remains challenging and is currently not possible (Figure S4). These findings underline the necessity of an integrated soil monitoring program to detect SOC changes in carbon farming schemes with smaller time scales (e.g. five years). Important pillars for this can be EU-wide (e.g. the LUCAS Soil Survey), and national (e.g. the BZE) soil inventories, supported by comprehensive parcel databases on cropland management such as the Land Parcel Identification System (LPIS). Process modeling

can integrate soil management data such as organic fertilization, crop rotation, or tillage, and improve the estimations of short-term SOC changes. Stratified soil sampling programs, with strong support from remote sensing, can help to scale the results down to a field level. In summary, SRC can provide comprehensive data on cropland soils and present a great opportunity to improve the information on long-term SOC trends. Due to its high spatial and temporal resolution and increasing availability, EO can hardly be substituted by other data sources and should therefore be integrated into cropland soil monitoring programs.

Acknowledgements

The project “KlimaFern” is funded by the Federal Ministry of Food and Agriculture as part of the German Climate Protection Programme 2022. We would like to thank the Bavarian State Research Center for Agriculture (LfL) for organizing and providing data on the long-term soil monitoring programs.

Supporting Information

Additional supporting information can be found online at https://doi.org/10.1111/gcb.17608

Chapter 11

Manuscript 5: Unveiling year-round cropland cover by soil-specific spectral unmixing of Landsat and Sentinel-2 time series

Felix Lobert, Marcel Schwieder, Jonas Alsleben, Tom Broeg, Katja Kowalski, Akpona Okujeni, Patrick Hostert, Stefan Erasmi

Remote Sensing of Environment, 2025

https://doi.org/10.1016/j.rse.2024.114594

Received 16 September 2024; Received in revised form 2 December 2024; Accepted 29 December 2024; Available online 9 January 2025

Abstract

Croplands are essential for food security but also impact the environment, biodiversity, and climate. Understanding, monitoring, modeling, and managing these impacts require accurate, comprehensive information on cropland vegetation cover. This study aimed to continuously monitor the state and vegetative processes of cropland, focusing on the assessment of bare soil and its cover with photosynthetic vegetation (PV) and non-photosynthetic vegetation (NPV) at the national level. We employed regression-based unmixing techniques using time series of Sentinel-2 and Landsat imagery to quantify cover fractions of NPV, PV, and soil during the agricultural growing season. Our approach extends existing spectral unmixing methods by incorporating a novel soil-specific unmixing process based on a soil reflectance composite. The extension accounts for variations in the spectral characteristics of soils which is particularly relevant for large-scale monitoring of annually cultivated croplands, as the spectral soil properties can vary considerably at national level and periods of bare soil are frequent. All cover fractions were predicted with mean absolute errors between 0.13 and 0.19. Introducing soil-specific unmixing reduced the mean absolute error of the predictions for soil by 11.3% and NPV by 15.1% without compromising PV predictions, particularly benefiting areas with bright soils. These findings demonstrate the efficacy of our method in accurately predicting crop cover throughout the cultivation period and underline the added value of incorporating the soil adjustment into the unmixing workflow. The contributions of this research are twofold: first, it provides essential data for the continuous monitoring of cropland cover, supporting agricultural carbon cycle and soil erosion modeling. Second, it enables further investigation into cropland management practices, such as cover cropping and tillage, through time series analysis techniques. This work underscores the potential of advanced spectral unmixing methods for enhancing agricultural monitoring and management strategies.

1 Introduction

Approximately a quarter of the European Union's (EU) land is covered by cropland making it the second largest land cover type in the EU after woodlands (Eurostat, 2024). While being the foundation of food security, croplands can also

have significant negative environmental, biodiversity, and climate impacts (Dudley and Alexander, 2017; Kross et al., 2022). Among the effects are greenhouse gas (GHG) emissions, soil erosion, nitrogen leakage, and habitat loss. However, the net impact of these effects is not always negative and largely depends on management practices and use intensity (Lal, 2009; Poeplau and Don, 2015; Tscharntke et al., 2012).

Continuous vegetation or residue cover of agricultural soils is an essential management practice for addressing several environmental issues. The sequestration of carbon in the soil, effectively removing CO_2 from the atmosphere, is significantly enhanced when fields are, e.g., continuously cultivated with cover crops instead of leaving fields as bare fallow (Johnson et al., 2007; Lal, 2009; Poeplau and Don, 2015). Reducing bare soil periods also mitigates soil erosion by maintaining the soil structure (Cerdan et al., 2010; Panagos et al., 2015). Additionally, it can prevent nitrogen leakage into ground- and surface water, thereby protecting water quality and promoting sustainable agricultural practices (Hively et al., 2020). To accurately assess the overall impact of cropping on the environment, comprehensive information on cropland cover is essential, and improved data quality can help reduce uncertainties in the modeling of soil factors and processes such as carbon dynamics (McClelland et al., 2021; Öttl et al., 2024; Seitz et al., 2023).

Earth observation (EO) enables frequent, broad-scale monitoring of cropland and has proven effective for tasks such as producing annual crop type maps (Blickensdörfer et al., 2022; Pham et al., 2024), delineating agricultural parcels (Tetteh et al., 2023), or analyzing crop phenology (Bolton et al., 2020; Lobert et al., 2023). Some studies also directly focused on monitoring specific cropland management practices. Among them are the detection of cover crops (Schulz et al., 2021), the identification of bare soil (Mzid et al., 2021), and the assessment of soil erosion (Vrieling, 2006). While providing valuable information, most products are static snapshots, lacking the ability to map nuanced variations in crop dynamics and growing periods (McClelland et al., 2021). Binary classification can oversimplify cropland cover characterization by only distinguishing between bare soil and cover crops, overlooking other nuanced land cover states, such as mulch, spontaneous vegetation, or volunteer grains.

A generic and versatile concept of describing the dynamics of agricultural

land cover with EO data is the analysis of fractional cover time series (Kowalski et al., 2022; Lewińska et al., 2020). Within this concept, the spectral signal captured by the satellite sensor is modeled as a mixture of defined endmembers according to their ground cover fractions. Unmixing spectral signals into cover fractions has a long history in satellite remote sensing. Starting from spectral mixture analysis (SMA) through area-weighted linear combinations of endmember spectra (Adams et al., 1986) or its variants such as Multiple Endmember Spectral Mixture Analysis (MESMA; Roberts et al., 1998), recent approaches make use of regression-based unmixing (Atkinson et al., 1997; Carpenter et al., 1999). Here, machine learning (ML) models estimate the cover fractions based on prior training using quantitative samples, which consist of mixed spectra labeled with their corresponding mixing fractions. To overcome the challenge of scarce training data, regression-based unmixing using synthetic training data generated from endmember spectra (Okujeni et al., 2013) provides a robust solution.

On agricultural land, the three main cover fractions are photosynthetic vegetation (PV), non-photosynthetic vegetation (NPV), and bare soil. The time series of these fractional cover estimates provide quantitative and interpretable units and were shown to support the investigation of environmental and management impacts on agricultural land (Guerschman et al., 2020). Among them are the mapping of crop residues (Barnes et al., 2021), short-term vegetation loss (Lewińska et al., 2020), and soil properties (Bouroubi et al., 2014). However, existing unmixing studies for agricultural land often cover limited study areas (Coulibaly et al., 2021; Yue et al., 2020) or only short periods (Li et al., 2016; Meusburger et al., 2010; Pacheco and McNairn, 2010), while wall-to-wall cover and high resolution long-term fractional cover time series are needed to support monitoring tasks and the evaluation of agricultural policies. Okujeni et al. (2024) demonstrated that the PV, NPV, and soil fractional cover time series retrieval method for grasslands, initially proposed for Sentinel-2 data by Kowalski et al. (2023), is both transferable to the Landsat archive and applicable on a national scale in Germany. This method's adaptability to multiple satellite sensors provides two significant advantages for large-scale monitoring activities. First, utilizing all available Sentinel-2 and Landsat observations maximizes potential revisit times (Lewińska et al., 2024), which is particularly important in regions with frequent cloud cover, such as the winter season in Germany (Mzid et al., 2021). Second, this allows to go back in time prior to the Sentinel era facilitating

multidecadal monitoring applications.

In contrast to grassland, where moments of pure bare soil states rarely or never occur (Okujeni et al., 2024), we are faced with significantly more and longer bare soil periods in cropland. The wide variety of soils and factors influencing soil reflectance, like, soil type and moisture, lead to distinct spectral signatures and can cause confusion between NPV and PV when dealing with spectral unmixing (Guerschman et al., 2015; Yue et al., 2020, 2019). Yue et al. (2019) examined this issue in detail and proposed a dynamic soil endmember selection based on soil moisture. The method resulted in an improvement of their rice residue cover estimations on four test fields in China with the same soil type. However, spatially and temporally accurate information on soil moisture and other soil parameters are required, which are not always available at the desired resolution. Furthermore, they added that soil moisture only influences the soil spectrum to a certain extent and that the spectrum also depends on the soil color, organic carbon content, and texture. Although their approach represented a valuable initial step in addressing the issue, further research is necessary to account for the variability of soils in spectral unmixing.

The overall goal of this study is the development of an approach for continuously monitoring the cropland cover with NPV and PV for large areas, by adapting a regression-based unmixing method originally developed for estimating fractional cover time series of grassland (Kowalski et al., 2023; Okujeni et al., 2024) utilizing all available Sentinel-2 and Landsat 8/9 imagery in Germany for 2011–2023. We extended the method by developing unmixing models specific to different soil reflectance groups that we identified from a soil reflectance composite of Germany following Broeg et al. (2024a). With this, we address the challenge of a large variety of factors that affect soil reflectance (Yue et al., 2019) and the often high spectral similarity of soil and NPV in multispectral data, depending on soil color, organic carbon content, and texture (Verrelst et al., 2023). We aim to answer the following research questions:

1. How accurately can we estimate the fractional cover of PV, NPV, and soil for cropland with regression-based unmixing of harmonized Sentinel-2 and Landsat imagery?
2. How much does incorporating soil reflectance groups in the unmixing approach improve the results?

3. How do multi-year time series of fractional cover predictions vary depending on different cropland cover conditions throughout the season?

2 Study area and data

2.1 Study area

The research was conducted across Germany, which spans approximately 357,592 km^2. Around half of this area is utilized for agricultural purposes, which splits up into roughly 70% cropland and 30% permanent grassland (Federal Statistical Office, 2024), which was not considered in this study. The country's landscape is diverse, featuring the flat North German Plain, the undulating Central Uplands, and the mountainous regions of the Alps in the south. Germany's soils accordingly also vary widely, including highly productive loess soils in the central and eastern regions, sandy soils in the North, and clayey/loamy soils in the southern areas. The country also encompasses multiple climate zones: the northern and northwestern parts experience a maritime climate with mild temperatures and moderate to high rainfall, influenced by the North Sea and the Baltic Sea. The eastern and southeastern regions have a continental climate, marked by hot summers and cold winters. Central Germany exhibits a transitional climate that combines elements of both maritime and continental climates, offering moderate temperatures and variable precipitation. The Alpine region in the south is characterized by cooler conditions, higher precipitation, and frequent snowfall, leading to shorter growing seasons (Deutscher Wetterdienst, 2024).

2.2 Remote sensing imagery

We used an available datacube of harmonized Sentinel-2 and Landsat 8/9 data for Germany, comprising all scenes with cloud cover below 75%. Landsat data were acquired as Level-1TP, while Sentinel-2 data was obtained as Level-1C. The scenes were pre-processed and corrected for radiometric and geometric effects using the Level 2 processing system of the Framework for Operational Radiometric Correction for Environmental Monitoring (FORCE; Frantz (2019)). All data were projected and resampled to a common grid with a pixel size of 10 m. For Landsat imagery, nearest neighbor resampling was used. All data were then stored

in the FORCE data cube structure consisting of non-overlapping 30 × 30 km tiles, creating an analysis-ready dataset (ARD). We further improved the spectral consistency across the sensors in our datacube by adjusting the reflectance values of all Landsat sensors and bands to the corresponding reflectance values of Sentinel-2 using slopes and intercepts from a reduced major axis regression derived by Okujeni et al. (2024). Bands for atmospheric correction, panchromatic and thermal bands, and bands not common to all sensors used were not further considered. This resulted in 6 bands covering the blue, green, red, near-infrared (NIR), and shortwave-infrared (SWIR 1 and 2) wavelengths.

2.3 Reference data

Aerial imagery

We derived fractional cover reference data from a very high resolution (VHR; 30 cm) aerial imagery mosaic for Germany (acquired between 2019 and 2022) provided by the German Federal Agency for Cartography and Geodesy (BKG) to validate our method. Following the approach proposed by Kowalski et al. (2023), we sampled Sentinel-2 pixels on croplands in Germany, where cloud-free Sentinel-2 observations were available with less than five days deviation from the acquisition date of the aerial image at that specific point. The sampling was stratified by the dominating crop types in Germany: winter wheat, winter barley, winter rye, rapeseed, silage and grain maize, spring barley, oat, and potato which were taken from annual German-wide crop type maps (Blickensdörfer et al., 2022; Schwieder et al., 2024). Furthermore, we ensured that a variety of different fraction mixtures were included by equalized random sampling observations from bins with 0.2 step size based on the Normalized Difference Vegetation Index (NDVI; Tucker, 1979) and the SWIR ratio ($\rho_{SWIR1}/\rho_{SWIR2}$; see Kowalski et al., 2022). For each sampled 10 m pixel, a grid of 100 sub-pixels of 1 m was constructed to assist the labeling process. By visually interpreting the VHR imagery, the dominant cover fraction (NPV, PV, or soil) was assigned to each sub-pixel. Subsequently, the fractional covers for each 10 m pixel were calculated through aggregation. We removed samples located on field edges to other land use types like streets, residential areas, or hedges. Our sampling design resulted in 201 labeled Sentinel-2 pixels distributed across Germany's croplands (Figure 11.1 A).The different acquisition dates of the VHR mosaic furthermore led to coverage of data from multiple years

and seasons (Figure 11.1 B/C).

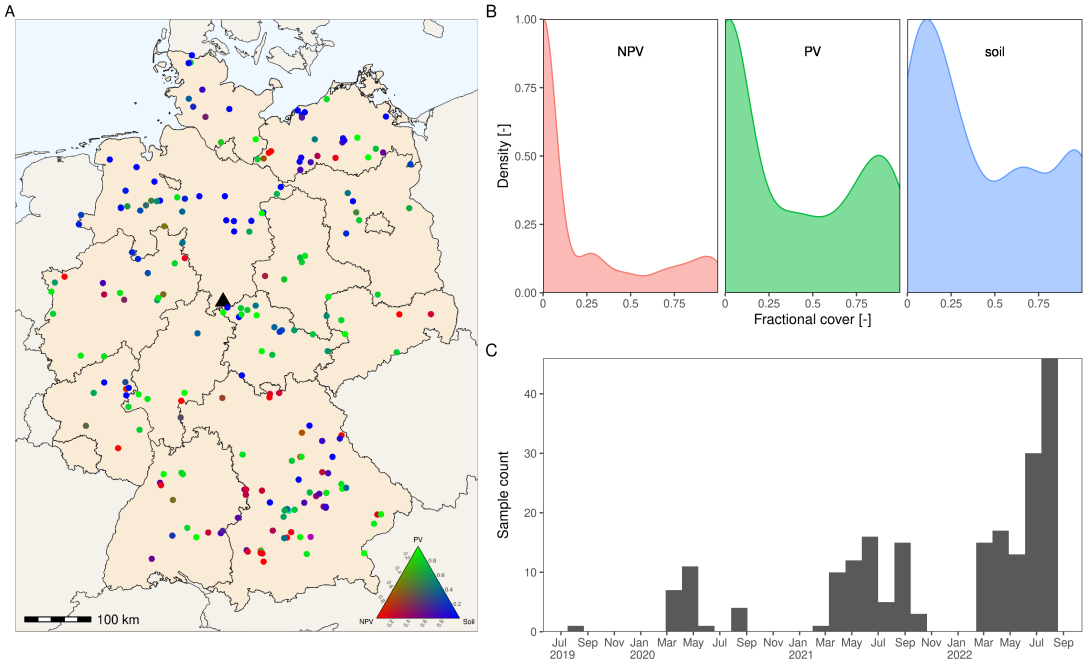

Figure 11.1: Locations of the Sentinel-2 pixels that were used for the validation of this study colored by their labeled cover fractions (A; the black triangle marks the area for which field observations were available), distribution of labeled fractional covers (B), and temporal distribution of the samples (C).

Field observations

We additionally had access to data from a field survey that was carried out between July 2023 and May 2024. During this period, cropland fields near the city of Göttingen in Lower Saxony (Figure 11.1 A) were visited repeatedly. The observation frequency ranged from up to weekly visits during phases of rapid plant growth and phenological development to bimonthly visits during the vegetative dormant phase in winter. Photos of the field cover were taken during each visit. Next to the cover photos, we had access to field boundaries and main crop types, and, if present, cover crop types were known for the observed fields.

3 Methods

We employed regression-based unmixing using synthetic training data from spectral libraries Okujeni et al. (2013). Specifically, we adapted the generalized

workflow for fractional cover time series retrieval for grasslands by Kowalski et al. (2023) and Okujeni et al. (2024) for monitoring croplands. We extended the approach to build unmixing models specific to different soil reflectance groups. This approach involves multiple steps that are explained in the following subsections:

1. creation of a soil reflectance composite for Germany and grouping of soils according to their spectral properties
2. compilation of a spectral library containing pure spectra, hereafter referred to as endmembers, representing NPV, PV, and soil for different crops and soil reflectance groups
3. synthetic mixing of the endmembers to create soil-specific training datasets for regression modeling
4. training of regression models and prediction of NPV, PV, and soil fractional cover time series based on Sentinel-2 and Landsat imagery
5. validation of the predicted fractional cover time series by a quantitative comparison with reference data derived from VHR imagery and a qualitative comparison with photos from field observations

3.1 Mapping soil reflectance groups

We divided the soils in Germany into different spectral groups to account for their spectral variability in our unmixing approach. This was based on a soil reflectance composite we derived following Broeg et al. (2024a). We used all available Sentinel-2 and Landsat 8/9 imagery from our datacube between 2011 and 2023 and applied a dynamic threshold-based function using the NDVI and Normalized Burn Ratio 2 (NBR2; Van Deventer et al., 1997) to the time series of each pixel. NDVI shows high values for PV and lower values for soil and NPV. While NDVI alone is not reliable for detecting bare soil (Demattê et al., 2018), NBR2 was shown suitable for separating bare and dry soils from wet soils and NPV (Dvorakova et al., 2023). We used the thresholds established by Broeg et al. (2024a) for detecting bare soil in the context of soil organic carbon estimation and excluded observations with an NDVI above 0.45 or an NBR2 above 0.16. We discarded values greater than the 15th percentile of the remaining NBR2 values of each pixel to further minimize any non-bare soil influence, like NPV or soil moisture. We then derived the soil reflectance composite by pixel-wise

averaging all remaining observations. We constructed soil reflectance groups by randomly sampling 250,000 points from the soil reflectance composite and applying the K-means algorithm (Lloyd, 1982). Based on iterative testing, we found that using six groups provided the best balance between capturing the variability in soil reflectance and maintaining meaningful distinctions between different soil types. Consequently, we assigned every pixel of the soil reflectance composite to its nearest soil reflectance group centroid, ensuring that each pixel was classified according to the most representative soil reflectance signature within our predefined groups.

3.2 Spectral library compilation

We developed an image spectral library using two different strategies, with PV and NPV endmembers obtained separately from soil endmembers. PV and NPV endmembers were based on the triangular feature space concept, wherein a triangular space spans along the two dimensions of SWIR ratio and NDVI, with endmember candidates positioned on the vertices of the space (Guerschman et al., 2009; Kowalski et al., 2022). We stratified our endmember selection along annual national crop type maps (2017 to 2019; Blickensdörfer et al., 2022) and created a comprehensive, multitemporal feature space by sampling 10,000 points per year and crop type to represent the heterogeneity of cropland cover in Germany. For each location, we extracted all Sentinel-2 observations of the respective years, to exploit the higher spatial resolution compared to Landsat, which increases the probability of identifying pure spectra. After constructing an NDVI/SWIR ratio feature space for each crop, we identified candidate spectra by selecting the greenest pixels for PV, which corresponded to observations with the highest NDVI and lowest SWIR ratio values. For NPV, we chose the driest pixels, characterized by observations with the lowest NDVI and SWIR ratio values. We summarized all potentially pure PV and NPV spectra by calculating the 25th, 50th, and 75th percentile spectra to reduce computational costs while maintaining the class-wise spectral variability. A detailed description of this procedure can be found in Kowalski et al. (2023). Soil endmembers were taken from the soil reflectance composite. The reflectance values were summarized to the 25th, 50th, and 75th percentiles for each soil reflectance group analogous to the PV and NPV spectra using the derived soil reflectance group map.

3.3 Synthetic training data generation

The compiled spectral library served as the foundation for creating synthetically mixed spectra and associated mixing proportions that were used as training data in the regression-based unmixing method. The approach for generating synthetic mixtures was originally introduced by Okujeni et al. (2013) and has been continuously enhanced. We followed the descriptions in Cooper et al. (2020) and Okujeni et al. (2021), which are based on a randomized mixing strategy to create separate sets of synthetic training data for each cover fraction of interest to train single-output regressors. We created 1000 synthetic mixtures for each cover fraction based on the spectral library. Each synthetic mixture comprised two to three endmembers from the three cover fractions randomly sampled from the library. These were then linearly combined with random fractions assigned to each endmember class, ensuring the fractions always sum up to 1. The resulting synthetic spectrum was then added to the training dataset, with the six spectral bands as input variables and the share of the target cover fraction as the label. Additionally, we included a shade spectrum in the synthetic mixtures to represent direct and structural shade components (Shimabukuro and Smith, 1991). The shade endmember, with a near-zero reflectance of 0.01 across all bands, was treated like any other endmember during the mixing step but was not considered as target fraction. We created two different types of synthetic training datasets for each cover fraction to assess the influence of soil variety in the unmixing procedure:

1. Soil-specific synthetic training datasets: Separately for each soil reflectance group, considering only the soil endmembers from the same soil reflectance group to train soil-specific models.
2. Global synthetic training datasets: Considering all soil endmembers simultaneously to train a global model.

We replicated each training set generation five times. This enabled us to train an ensemble of regression models and aggregate the results afterward. This approach resulted in one global and six soil-specific training data compositions for each cover fraction and a total of 105 training datasets after replication.

3.4 Regression-based Unmixing

The training data sets were used to train global and soil-specific Support Vector Regression (SVR) models for the prediction of fractional cover time series based on Sentinel-2 and Landsat data. We used a grid search combined with 10-fold cross-validation to determine the optimal *gamma* and *c* parameters for each SVR model as model parameterization strategy following Van der Linden et al. (2015). The trained models were applied to each image in our datacube which covered the reference data. The global model was applied to all pixels, while the soil-specific models were applied to pixels based on their grouping in the soil reflectance map. Predictions from the ensemble of models trained on the replicated training sets were averaged per pixel. The predicted values were normalized per pixel by dividing each fractional cover by the sum of the fractional covers. This ensured that the fractional covers always sum up to one. For the plots covered by the available field observation data, we subsequently summarized the predictions for all pixels within the field boundaries to the median for each observation. To mitigate edge effects, we applied a negative buffer of 20 m.

3.5 Validation

We validated our predictions against the reference samples derived from VHR imagery (see Section 2.3) based on the mean absolute error (MAE) and the coefficient of determination (R^2). The MAE gives an overview of the general error magnitude in the same unit as the predictions, while R^2 offers insight into how well the model predictions explain the variation in the reference data. These calculations were performed for each reference sample and its nearest corresponding prediction within a $\pm$ 5-day window. Together, these metrics enabled us to compare the performance across different cover fractions and the global or soil-specific unmixing models. Following this quantitative evaluation, we plotted the time series for the plots covered by the available field observation data alongside a selection of cropland cover photos, enabling the visual interpretation of their correspondence with the actual cover observed in the field. Only for visualization purposes, time series were smoothed and interpolated using locally estimated scatterplot smoothing (loess; Cleveland et al., 1992) using a span parameter of 0.25.

4 Results

4.1 Mapping soil reflectance groups

The soil reflectance composite reflects the overall diversity of soils in Germany (Figure 11.2 A). Dark soils, with higher organic carbon contents, are visible in central and Northwest Germany, while brighter sand-rich soils with low carbon contents are present in the Northeast and East. The results of the k-Means algorithm summarize these patterns into soil reflectance groups with similar spectral properties (Figure 11.2 B). Darker soils are categorized into groups 1 to 3, while the bright, yellowish soils belong to groups 5 and 6. Soils with medium brightness are found in group 4. The summarized spectral signatures of these groups built the basis for extracting the soil endmembers that were used for the soil spectral library compilation (Figure 11.3). Here, next to the general brightness gradient across all bands, some differences could be observed. While most groups show a similar slope between SWIR1 and SWIR2, soil group 3 shows considerably less slope in this part of the spectrum. This is especially of interest since the spectra of the soil groups 2 and 3 are nearly identical in the visible and near but show high differences in the SWIR. We have also found differences in the red range that make the course from green to NIR either straight (groups 4 to 6) or with a bend (groups 1 to 3). In total, the soil reflectance composite covers 97% of croplands in Germany identified from the CTM with sufficient bare soil observations. Further information on the number of identified bare soil observations per pixel is presented in the appendix (Figure 11.11).

4.2 Accuracy of fractional cover predictions

Scatterplots of reference data against predicted fractions offer a detailed insight into both modeling setups (Figure 11.4). For the NPV fraction, both models reveal high scattering. We also observed a tendency to overestimate fractions near zero. Nevertheless, MAE and R^2 improved with the soil-specific model, indicating overall better estimates. For PV, both the global and soil-specific models produced similar results and showed the least scattering of all three cover fractions. While R^2 was identical for both models, MAE was only marginally impaired with the soil-specific model. Both models showed nearly identical point patterns for PV, indicating consistent performance across both approaches. For

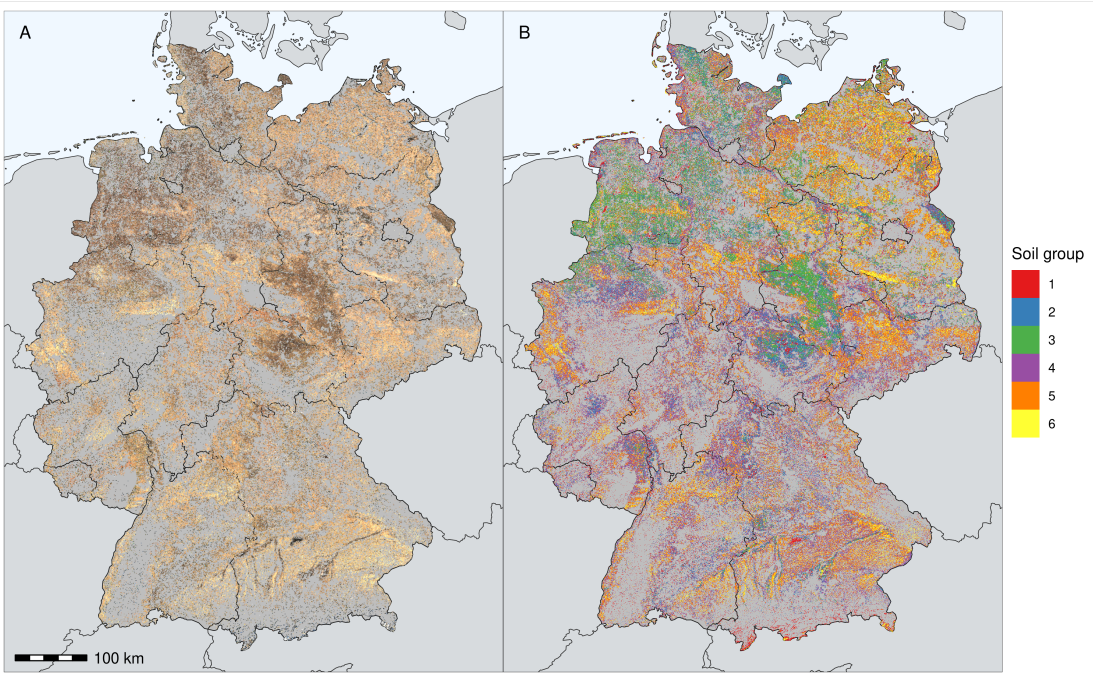

Figure 11.2: Soil reflectance composite of German croplands in RGB true-color representation (A) and map of the derived soil reflectance groups (B).

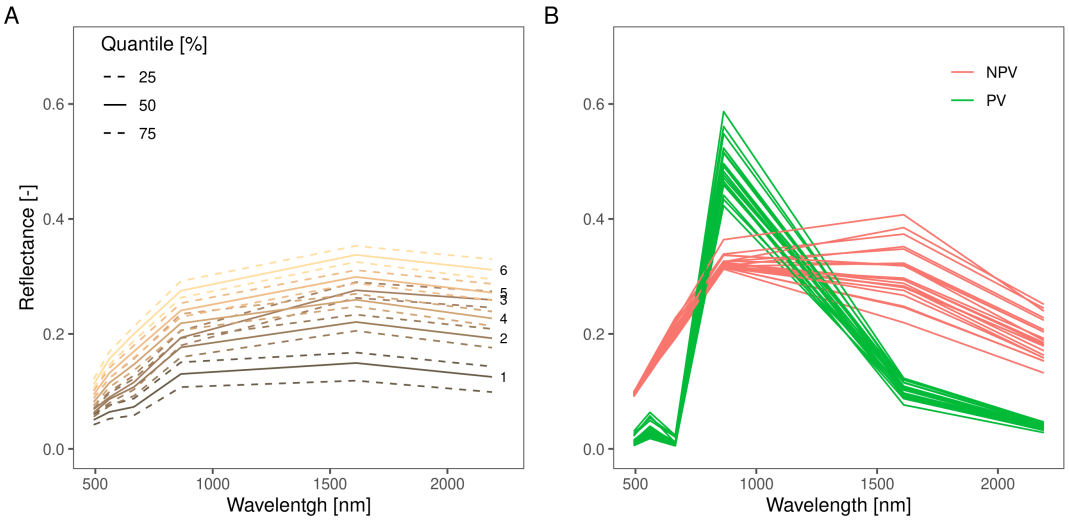

Figure 11.3: Quantile spectra of extracted soil endmembers for different reflectance groups, with line colors representing the average RGB true color of each group (A), and spectra of PV and NPV endmembers across all crop types and quantiles included in the spectral library (B).

the soil fraction, both models again show scattering, and an underestimation of values particularly close to 1. The soil fraction also benefited from the soil-specific modeling approach. The global model showed an MAE of 0.159, while the soil-specific model reduced the MAE to 0.141. The scatterplots confirmed this improvement, indicating better performance, particularly at higher soil fractions, which is shown by the regression line that is much closer to the 1:1 line.

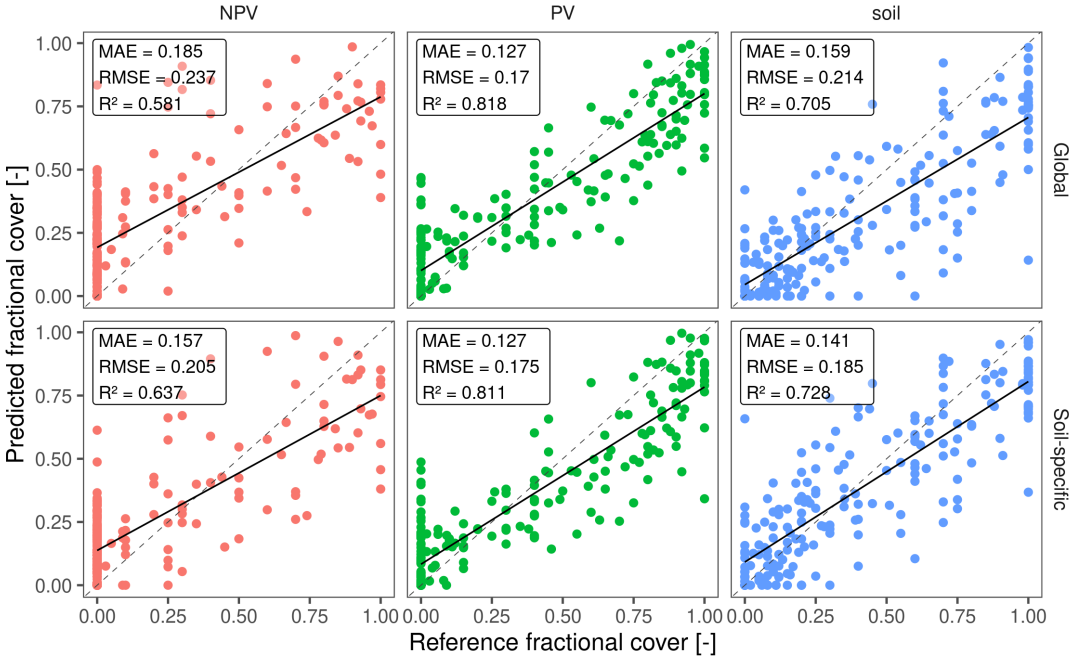

Figure 11.4: Scatterplots of predicted and reference fractional covers with regression line (solid) and 1:1-line (dashed) for NPV, PV, and soil for the global (top) and soil-specific (bottom) models.

The error metrics and improvements with soil-specific models varied across soil reflectance groups (Figure 11.5). For NPV, the global model showed the highest errors in soil reflectance groups 4 and 5, with MAEs of 0.22 and 0.21. The soil-specific models improved the predictions for these groups, reducing the MAE by up to 0.04, while smaller reductions were seen in groups with generally lower errors in the global model, where MAEs ranged from 0.15 to 0.17. For PV, the global model's errors showed little variation across soil reflectance groups, with a slightly higher MAE in group 3. The soil-specific models had minimal impact on the PV fraction, with only a slight decrease in MAE for soil reflectance groups 3 and 5, where errors were already relatively low with the global model. For the soil fraction, the global model had the highest MAEs in groups 1 and 5, while

group 6 had the lowest. The soil-specific models improved accuracy, particularly in groups 1 and 5, where the MAE decreased by up to 0.03, addressing the largest initial errors. Results stratified by crop type are presented in the appendix (Figure 11.12).

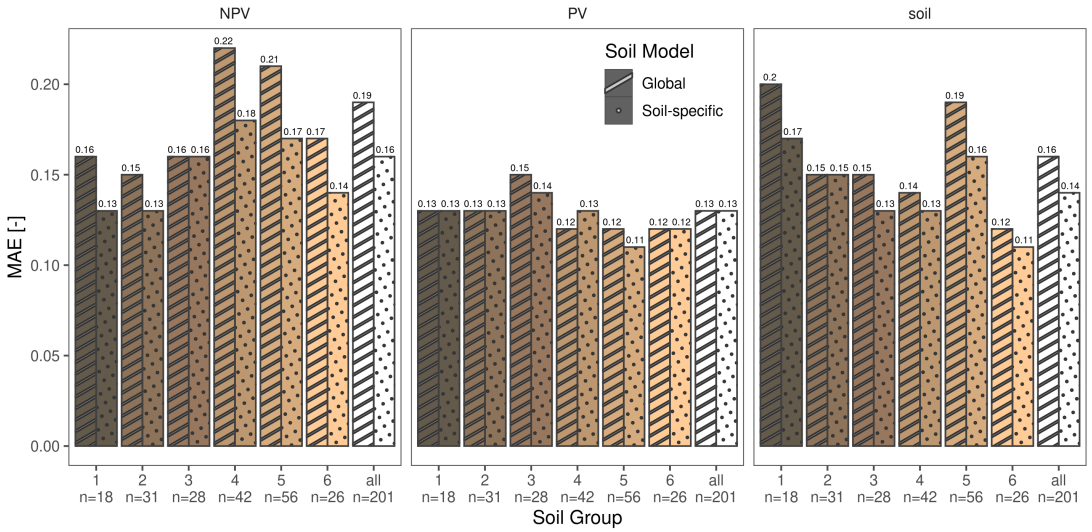

Figure 11.5: Mean absolute error for the three cover fractions stratified by soil reflectance group for the global and soil-specific unmixing models. The fill color of the bars represents the average RGB true color of the respective soil reflectance group.

When analyzing the soil fraction in more detail, we observed that the soil-specific model had varying impacts on the prediction accuracy across different soil reflectance groups. While for most soil reflectance groups the MAE decreased, observing the scatterplots revealed a more nuanced picture (Figure 11.6). In the case of bright soils, there was a notable adjustment of the regression line towards the 1:1 line, particularly at higher soil fractions close to 1. This was reflected by a substantial increase in R^2 for groups 3 to 6. In contrast, predictions for dark soils (groups 1 and 2) exhibited a decrease in R^2, while a slight general upward shift of all values is visible.

4.3 Fractional cover maps

Spatial patterns of the predictions are illustrated for an exemplary area in Southern Germany in the federal state of Bavaria (Figure 11.7). A prediction for the whole study area can be found in the appendix (Figure 11.13). The false-color

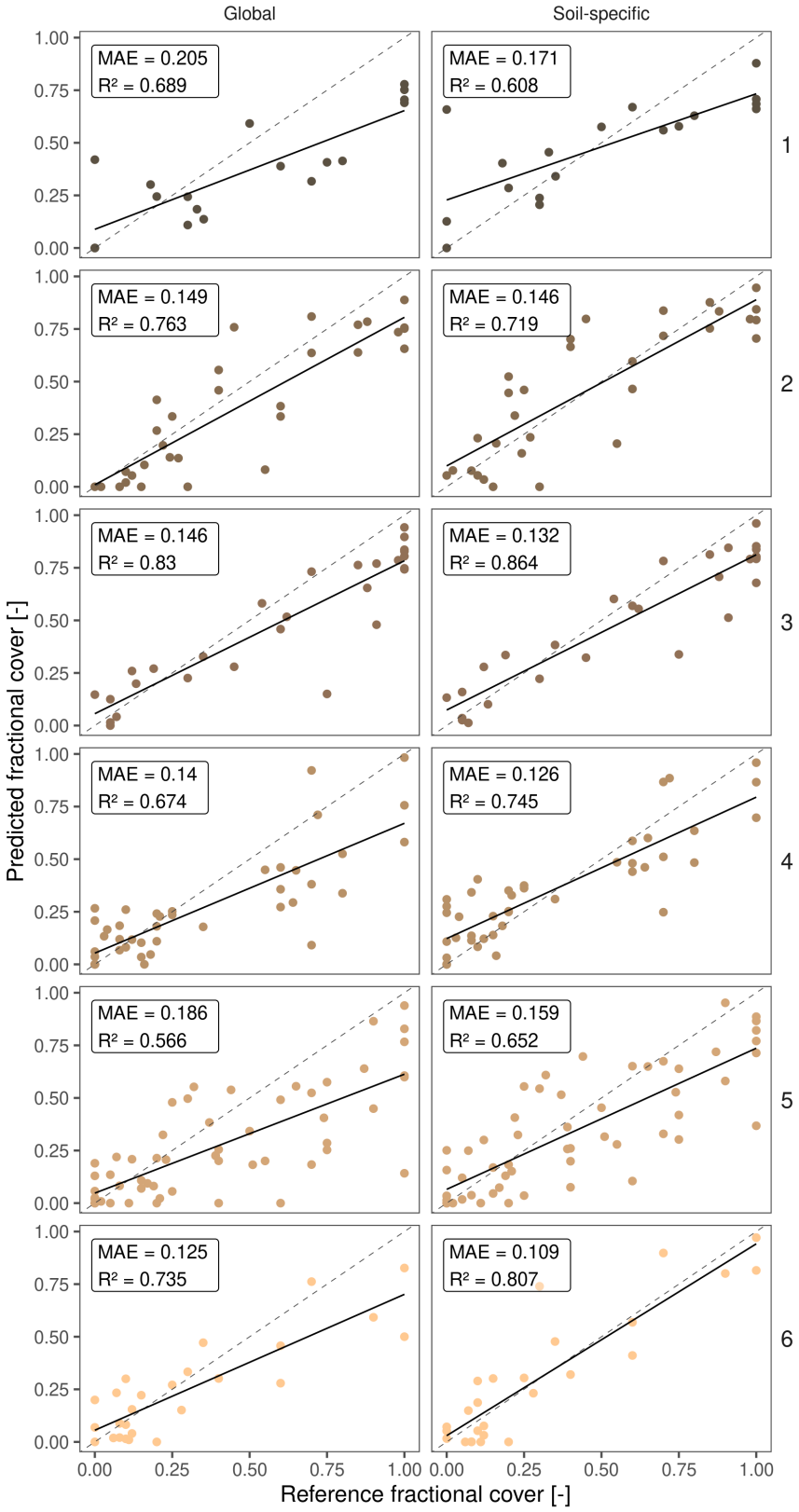

Figure 11.6: Scatterplots between predicted and reference soil cover fractions stratified by soil reflectance group (rows 1 to 6) for the global and soil-specific unmixing models (columns). Dot colors refer to the average RGB true color of the respective soil reflectance group.

composite uses SWIR2 in the red channel, NIR in the green channel, and the red band in the blue channel and is scaled between the 3rd and 97th percentile per band. As a result, bare soil is represented by magenta tones due to its relatively high reflectance in both the SWIR2 and red bands, combined with lower reflectance in the NIR band. PV appears green because it reflects strongly in the NIR band while having lower reflectance in the SWIR2 and red bands. NPV is shown in blue because it has lower reflectance in the SWIR2 and NIR bands but higher reflectance in the red band, leading to a dominance of blue tones in the composite. The soil reflectance group map shows the variation in soil reflectance across this region. Darker soils (groups 1 to 3) dominate the southern regions, while brighter soils (groups 4 to 6) are more prevalent in the North and Northwest.

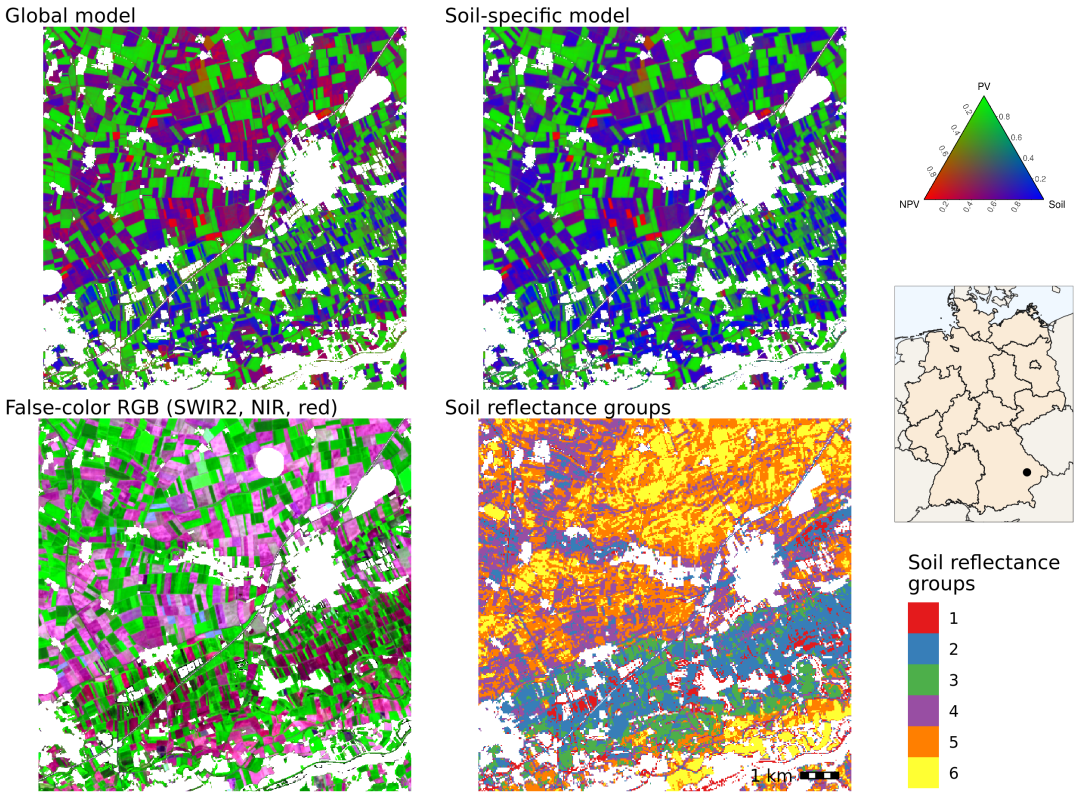

Figure 11.7: Predictions from the global and soil-specific unmixing models and false-color composite based on a Sentinel-2 scene from May 28, 2023, together with the soil reflectance groups for a site in Bavaria.

In the global model predictions, we observe frequent mixtures of NPV and soil (purple pixels) in the North and Northwest, alongside fields dominated by

PV (green) and NPV (red). In the southern regions, fewer NPV-soil mixtures are present, with fields more commonly characterized by pure pixels of soil, PV, or NPV. This pattern aligns with the distribution of soil groups, where brighter soils (groups 4 to 6) tend to have more NPV-soil mixtures, while darker soils (groups 1 to 3) show more pure soil pixels.

In contrast, the soil-specific predictions produce a more consistent map. Fields in the North and Northwest that exhibited NPV-soil mixtures on pure soils in the fraction maps under the global model are now showing more realistic fractional covers that better match the false-color composite. Overall, the influence of soil reflectance groups on the predictions and spatial patterns of cover fractions is less pronounced in the soil-specific model, leading to a more uniform distribution across the area.

Another interesting finding is, that fields with high NPV fractions in the global predictions do not change in the soil-specific ones, while other areas, especially those with high soil fractions, exhibit an adjustment through the consideration of soil reflectance groups in the prediction. This matches our observations from Figure 11.6, that especially high soil fractions on bright soils are underestimated with the global model, while the soil-specific models can account for that and predict high soil fractions on bright soils better.

The zoom-ins in Figure 11.8 enable a more detailed comparison of soil-specific and global modeling results along with coincident orthophotos across four landscape subsets and for different phases of the growing season. Panels A and B illustrate landscapes with small to medium-sized agricultural fields after the growing season for winter crops and at the beginning of ripening for summer crops. The harvested cropland is characterized by tillage activities leading to diverse patterns of bare fields with varying amounts of crop residues. Both the global and soil-specific models effectively capture the spatial variability evident in the orthophotos, particularly in distinguishing between areas of bare soil and vegetation. When comparing the soil-specific with the global predictions, we observe higher shares of NPV mixtures with soils predicted by the global model. In contrast, the higher soil cover depictable in the aerial photos is also represented by higher soil fractions in the unmixing results. This is consistent with the patterns seen in Figure 11.7. Panel C shows a site with larger fields during stem elongation of winter crops and seedbed preparation for summer crops, where fields

are predominantly covered by PV and soil. Here, notable intra-field differences are distinctly visible, demonstrating the models' sensitivity in distinguishing different cover fractions even on a fine scale. There are no noticeable differences between the global and soil-specific predictions. Panel D features a site with smaller fields towards or after harvest for winter crops and during the vegetative phase for summer crops. Fully matured winter crops show high NPV fractions and fields with summer crops are fully covered by PV. Both models capture the borders and differences between the fields, even in areas that have been partially harvested. The only notable difference is the higher soil share predicted by the soil-specific model. Notably, fields with high NPV fractions are not reduced in the soil-specific predictions, underscoring that soil-specific modeling does not simply result in reduced NPV predictions.

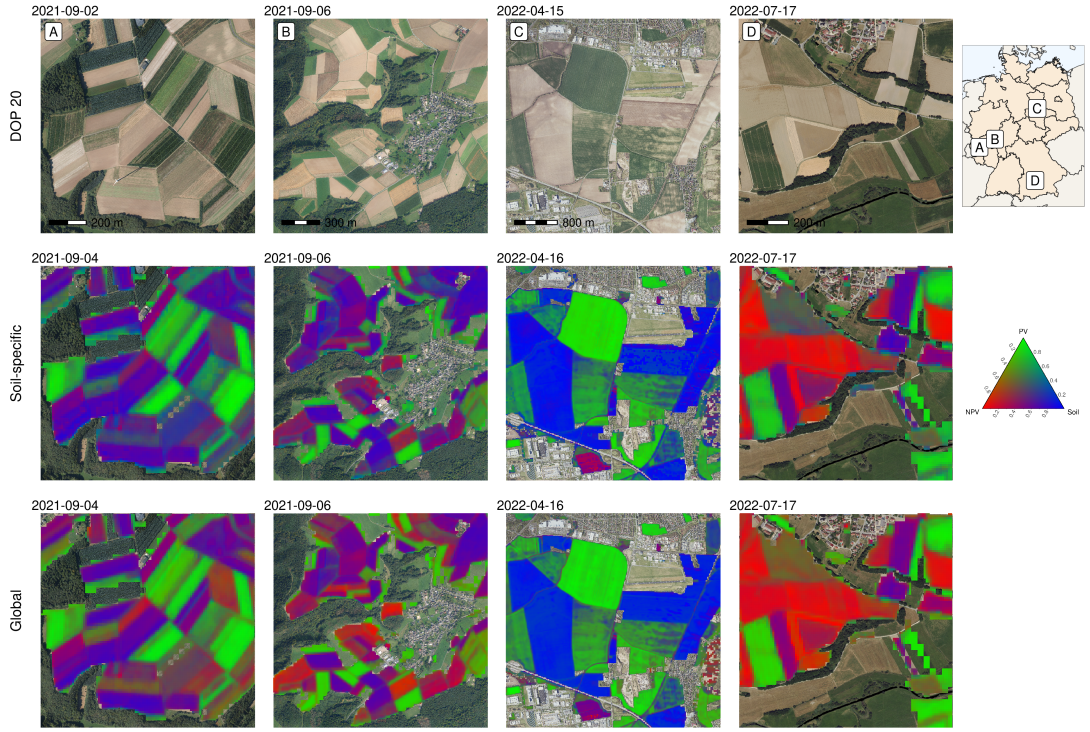

Figure 11.8: Orthophotos (20 cm; © GeoBasis-DE / BKG [2024]) and corresponding cover fraction estimates from global and soil-specific models based on the closest Sentinel-2 observation for different sites in Germany.

4.4 Evaluating fractional cover time series

The computation of fractional cover maps for the plots covered by the field observation data based on the available Sentinel-2 and Landsat images enabled the generation and qualitative analysis of fractional cover time series. Figure 11.9 and Figure 11.10 show examples of such time series for different soil reflectance groups and two-year crop sequences.

Figure 11.9 shows a field belonging to soil reflectance group 6 initially cultivated with winter wheat in 2023, followed by a cover crop during winter and the seeding of a summer crop in spring 2024. At the beginning of June, the field exhibited nearly full PV cover, with minimal soil and NPV presence. This was followed by a rapid decline in PV and a corresponding increase in NPV. This also becomes visible from image A in mid-July, where the winter wheat had reached full yellow ripeness. Up to this point, the time series are very similar to those in Figure 11.10.

Following tillage at the end of August, the green-up of a cover crop was visible in September, as shown in image D. This quickly reduced the estimated share of bare soil. The predicted increase in NPV during this time does seem counterintuitive and appears to be an overestimate. By November, field visits confirmed the cover crop had fully grown, as seen in images E and F. Starting in December, the NPV share began to increase, reaching a maximum of over 0.8 by the end of January, which is caused by the dying off of the cover crop. During this period, the estimated share of PV steadily decreased, hitting a minimum in February. This transition is evident in images G and H, which show only residues of the cover crop and bare soil being left from mid-February on. The estimated soil share also began to rise, slightly lagging behind the changes in NPV, and reached a maximum of nearly full cover by May. This increase corresponds with images I and J, which reveal mulch tillage and the sowing of a summer crop. Differences between global and soil-specific modeling become more evident here, compared to Figure 11.10. Especially in May where nearly full soil cover is visible from the field observations, the global model predicts around 30% NPV and only 70% soil, while the soil-specific model corrects these values to less than 10% NPV and 90% soil. Major differences between both models are also visible during September 2023 and January 2024. Figure 11.10 shows a field belonging to soil reflectance group 3 cultivated with winter wheat in 2023, followed by a bare fallow

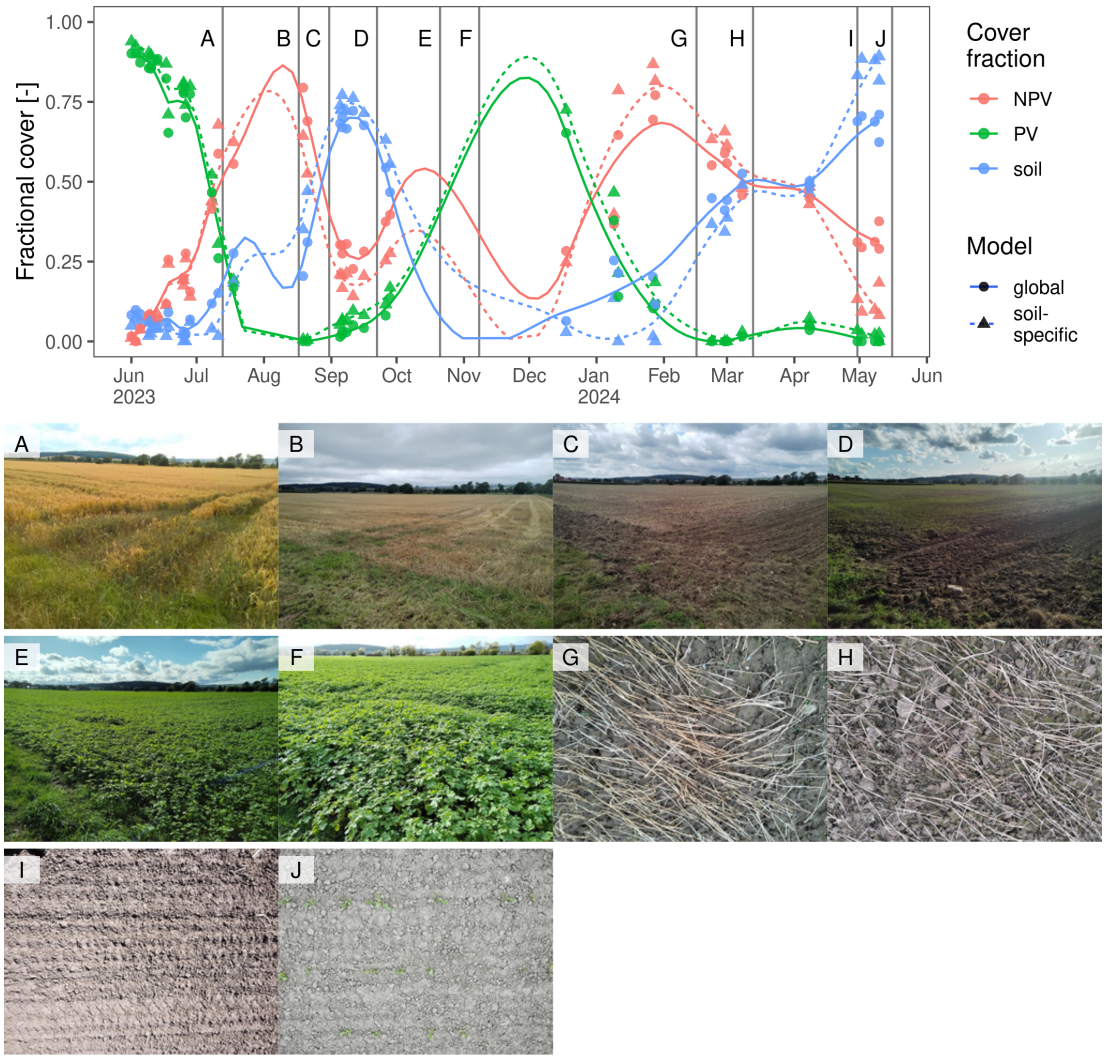

Figure 11.9: Predicted fractional cover time series for photosynthetic vegetation (PV), non-photosynthetic vegetation (NPV), and soil for a field initially cultivated with winter wheat in 2023, followed by a cover crop during winter and the seeding of a summer crop in spring, supplemented with photos from field visits. A-F show landscape perspective, G-J show overhead perspective. The field belongs to soil reflectance group 6. Points represent the original unmixing results, lines show a smoothed trend (only for visualization).

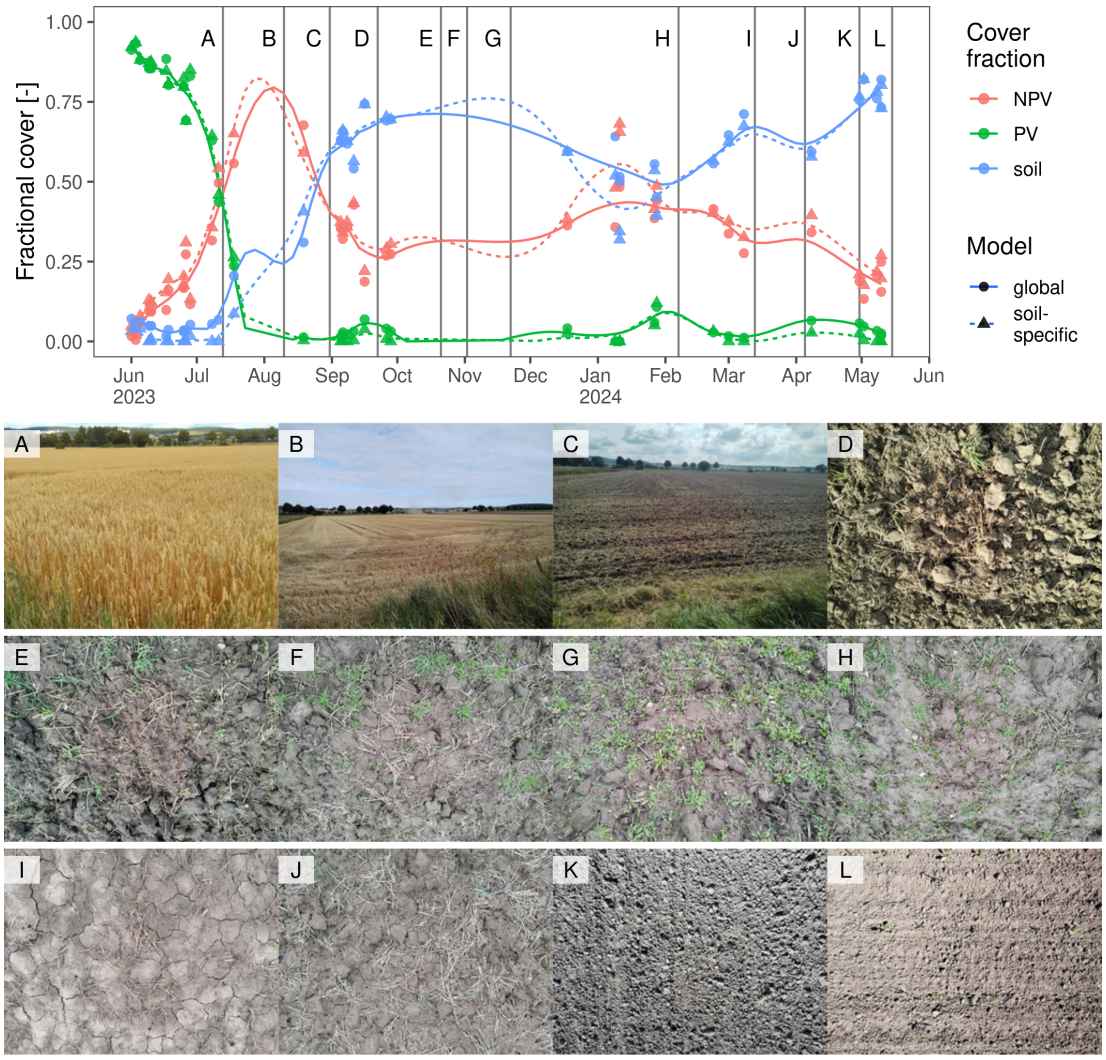

Figure 11.10: Predicted fractional cover time series for photosynthetic vegetation (PV), non-photosynthetic vegetation (NPV), and soil for a field initially cultivated with winter wheat in 2023, followed by a bare soil period during winter and the seeding of a summer crop in spring, supplemented with photos from field visits. A-C show landscape perspective, and D-L show overhead perspective. The field belongs to soil reflectance group 3. Points represent the original unmixing results, lines show a smoothed trend.

period after harvest and during the following winter. At the beginning of June, the models predicted nearly full PV cover, with little to no soil or NPV presence. This was followed by a rapid decrease in PV in favor of NPV, as observed in image A from mid-July, where the wheat had reached full yellow ripeness. Harvesting occurred in August, resulting in a state of stubble fallow as shown in image B, with around 70% NPV, 30% soil, and nearly zero PV share estimated. By the end of August, initial tillage practices were conducted, leading to higher estimated soil shares afterward (image C). Starting in September, very sparse spontaneous vegetation appeared on the field, as can be observed in image D. The spontaneous vegetation showed only little growth between October and February (images E to H). During this period, the PV share rose gently, while NPV and soil remained relatively stable at around 0.5, yet fluctuated during this period. From February onwards, soil share began to rise rapidly, corresponding to the absence of visible green vegetation in images I to L, and a decreasing amount of NPV, particularly from images J to K, where tillage was visible. By mid-May, estimated soil cover was above 0.8, while NPV estimations were around 0.2, even though no NPV was observable in images K and L, which suggests an overestimation. Differences between the global and soil-specific predictions were less pronounced compared to Figure 11.9 and no notable pattern was visible. Only during January 2024, global predictions rank soil over NPV while the opposite is the case for the soil-specific predictions.

5 Discussion

5.1 Fractional cover estimates

We developed and evaluated an approach for predicting the fractional cover of NPV, PV, and soil across German croplands over time. Building on the methodology developed by Kowalski et al. (2023) and Okujeni et al. (2024) for grassland monitoring, we adapted the approach of regression-based unmixing to croplands and incorporated a novel extension that accounts for the impact of soil properties in predicting fractional cover. Overall, quantitative validation against multi-year reference data derived from VHR aerial imagery demonstrated the model's successful transferability from grassland monitoring to cropland applications. We found satisfactory agreements for all cover fractions between the cropland refer-

ence data and the fractional covers from our model predictions. However, the different cover fractions and models need a detailed and differentiated interpretation.

While the absolute values differed, we found the same ranking for the accuracy of the three cover fractions as in other studies. In the context of grasslands, Okujeni et al. (2024) reported MAEs of 0.149, 0.067, and 0.135 for NPV, PV, and soil, respectively. Kowalski et al. (2023) found similar but lower values of 0.137, 0.065, and 0.122 for the same fractions. Studies conducted in other contexts also showed variability in accuracy metrics, while showing the same ranking. For example, Guerschman et al. (2015) reported RMSE values of 0.162, 0.112, and 0.130 for NPV, PV, and soil, respectively, while unmixing Landsat data in Australian pastoral and agricultural settings. Laamrani et al. (2020) observed R^2 values of 0.84 and 0.7 for PV and NPV, respectively, in their work on green vegetation and crop residues in Canadian corn and soybean fields. Compared to grasslands, arable land is subject to a greater variability of fractional cover due to different management and growth patterns within and between crop types. While grasslands, e.g. in drylands, also present challenges such as low vegetation cover, soil background, etc. (Smith et al., 2019), we still anticipated a reduced accuracy for our PV and NPV fractional cover predictions compared to grasslands. In Germany, plowing and tillage are regular practices in cropland cultivation, whereas such management practices are highly restricted on permanent grasslands. Extended periods without PV cover, dominated by mixtures of NPV and soil, as well as prolonged slow emergence of PV after sowing, are common on arable land and characterize the phenological cycle of most crops and all practices related to crop residue management. In general, grasslands only experience such states during droughts and mowing. This led to our samples covering the whole gradient of fraction mixtures, while the reference samples for the soil fraction on grasslands used by Okujeni et al. (2024) were concentrated around 0 to 0.2 and 0.8 to 1.

NPV and soil both exhibited lower performance than PV, indicating confusion between these cover fractions. This phenomenon has been noted in several studies on spectral unmixing and can be attributed to the high similarities in the spectral signatures of NPV and soil, which differ primarily in the SWIR region (Delegido et al., 2015). Similar results were also reported by Li et al. (2016) for open woodlands and grasslands in Northern China using multispectral GF-1 wide-field view data. They explained that certain combinations of bare soil and

NPV result in ambiguous spectral signatures that are difficult to resolve, a limitation also highlighted by Okin et al. (2001) in their seminal study on the practical limits of hyperspectral vegetation discrimination. This issue aggravates when the intra-class variability within the spectral library is high, e.g., when multiple soil types are involved. Then, mixtures of dark soils with a small proportion of the generally brighter NPV cover may be misidentified as pure bright soils.

Our results also revealed that, for the same fraction, some value ranges were underestimated while others were overestimated, a pattern observed in other studies as well. Guerschman et al. (2015) noted similar findings, with low NPV and soil shares being overestimated, while higher shares tended to be underestimated. This trend was also observed by Dennison et al. (2019) in their study on cropland unmixing using simulated hyperspectral data. This might be due to the visually interpreted reference fractions. Although being commonly used to validate fractional cover when in-situ field data is unavailable (e.g., Mandl et al., 2024; Nill et al., 2022; Schug et al., 2024), this method can introduce uncertainty in labeling. Very low fractional cover with NPV or PV cannot always be clearly recognized even in aerial images with a resolution of 20 cm. While a fractional cover of 0 is then labeled during interpretation, even low degrees of coverage can already change the remote sensing signal and are reflected by a predicted fractional cover by the model, as also observed by Kowalski et al. (2023). The same counts for very high and full degrees of cover by NPV or PV.

5.2 Soil-specific Unmixing

The derivation of a soil reflectance composite for Germany provided the foundation for implementing the soil-specific unmixing approach, including both the development of the spectral library and the predictions of soil-specific cropland cover. This approach allows us to extract soil endmembers that represent the variability of soils across the study site, which is particularly valuable when no soil spectral database is available. Here, we even preferred this strategy over soil databases or official soil maps, because it supports not only developing soil-specific models but also soil-specific predictions, by determining the best-fitting model for each pixel based on the soil group map. The high spatial resolution of the soil reflectance composite of 30 m further enabled us to account for spectral soil differences even within a single field - something that official maps usually cannot

achieve due to their lower resolution. Additionally, the soil reflectance composite provides the actual spectral signature of the pixel. This directly targets the challenge of determining the appropriate endmember for each pixel, especially for soil, as concluded by Li et al. (2016). Using this information addresses the problem of ambiguities in the unmixing of cropland with different soils.

The spectral signatures of the soil reflectance groups showed a high variability. The brightness gradient, especially in the visible spectrum, could indicate a general gradient of organic carbon content in the soil, with more organic carbon content absorbing more visible radiation (Udelhoven et al., 2003). Iron oxides were found to be related to the red and NIR parts of the spectrum (Chabrillat et al., 2019; Richter et al., 2009), where we also found a gradient between the spectra (Figure 11.3). Another interesting finding was a less negative slope between SWIR1 and SWIR2 for soil group 3 compared to group 2, although being nearly identical in the visible and NIR. This could indicate that group 3 has a lower clay content, which is known to absorb especially in the SWIR2 region (Chabrillat et al., 2019).

The utilization of soil-specific models resulted in improvements in the NPV and soil fractional cover predictions. PV showed little to no changes attributed to this model change. For NPV, the overestimation was reduced while for soil we observed a reduced underestimation. This indicates that the soil-specific model was able to reduce the often-described uncertainties between NPV and soil depending on soil brightness, organic carbon content, and decomposition status of the NPV/organic matter (Verrelst et al., 2023). It is important to note that the improvements in soil accuracy did not come at the expense of NPV, and vice versa. At the same time, this method extension did not harm the PV predictions. Laamrani et al. (2020) used Landsat 8 data and linear spectral unmixing to estimate cropland soil cover with corn residue, soybean residue, and green cover in Canada, finding no significant improvement when accounting for soil types. They attributed this to their sampling being conducted in winter, when the soil was uniformly very wet and close to field capacity, making the differences between soil types less pronounced. Our year-round sampling may explain why we could not confirm their findings.

Training a model with all soil spectra combined reduces the number of bands with significant reflection differences between soil and NPV, due to the wide vari-

ation in soil reflection values. As can be taken from Figure 11.3, when all soil spectra are considered at the same time, the primary differences from NPV are concentrated in the SWIR2 region and the slope between SWIR1 and SWIR2, which weakens the overall decision-making basis for the SVR models. In contrast, using a stratified model that includes only the soil spectra relevant to the pixel of interest excludes irrelevant spectra, thereby strengthening the model's decision basis and improving accuracy. For example, soil group 1, despite being visually distinguishable from NPV spectra, exhibited the highest MAE when using the global model, possibly due to the global model's focus on the SWIR region. However, with the soil-specific model, we observed remarkable improvements in MAE for soil group 1, possibly because the model considered more bands when irrelevant soil spectra were excluded during training. This does not support the findings from other studies observing global models to be similar or even superior compared to models stratified by geographical regions or timeframes (Dudley et al., 2015; Kowalski et al., 2023). The differences between those and our findings could be attributed to several key differences in focus and methodology. First, unlike our study, which concentrated on croplands, both studies focused either on deriving fractional cover times in grasslands or the species-unmixing in forests and rangelands, presenting environments that are difficult to compare. Additionally, while we stratified our models solely based on soil types, we used a global library comprising PV and NPV spectra for all crop types and multiple years across Germany. This could indicate a combination of the strengths coming from both stratification and generalization in our approach, leading to more accurate estimates.

Although the soil groups had varying sample sizes and distributions, potentially limiting the robustness of estimates and conclusions about the relationship between spectral features and soil-specific unmixing performance, our results showed overall improvement across all soil types. The scatterplots per soil type reinforced this (Figure 11.6), neither revealing a major decrease in performance for any soil group nor indicating that only specific groups benefited from the soil-specific approach. This consistency underscores the effectiveness of our method. While the mentioned constraints of generating reference data through visual interpretation of VHR imagery may affect absolute accuracy, this approach remains suitable for the intended purpose of providing a consistent basis for the relative comparison of the two modeling approaches.

5.3 Monitoring of cropland with fractional cover time series

The available field photos provided a unique data source for qualitative validation and were valuable for comparing time series fractional cover estimates to observed vegetation dynamics and management events on the field. They also helped assess the suitability of the time series for real-world monitoring tasks beyond the comparison with singular mono-temporal reference fractions. For the two winter wheat fields with field photos available, we observed very similar trajectories of the fractional cover estimates from the beginning of our observation period until harvest, indicating consistency between the same crop types. Remarkable differences, however, were observed during winter. Depending on whether the field was left as bare fallow during winter or a cover crop was cultivated and removed in spring before sowing a new spring crop, we found distinct patterns in the time series that accurately reflected the field cover changes during these measures, even if only minor field activities, such as spontaneous greening, are visible in the PV time series.

Some time series did not reach 0 or 1 despite observing full PV cover or bare soil in the field photos. This indicates that the overall uncertainties observed in the quantitative evaluation are also present in the specific time series for our test fields. NPV sometimes only reached low values when PV was the dominant cover fraction (Figure 11.10), while soil dominance often led to NPV being overestimated. This finding highlights that there is still some confusion remaining between these two classes, though no confusion was noted between NPV and PV. Despite these uncertainties in estimating exact cover fractions, the method has proven well-suited for capturing the dominant cover type and its temporal development.

Germany also experienced a very cloudy and rainy autumn and winter in 2023/24. This is reflected in a low density of cloud-free observations that we observed in the satellite data time series. Since winter is among the most important periods for soil cover monitoring, this might be a potential shortcoming of the datasets. However, we have also seen that even under these unfavorable weather conditions, specific management practices like cover crops and bare fallow periods could be traced in our time series.

6 Conclusion

Our study confirms the effectiveness of spectral unmixing using Sentinel-2 and Landsat data for cropland monitoring. The method delivers reliable time series of NPV, PV, and soil fractional cover, offering valuable insights for continuous monitoring of cropland management, agricultural carbon cycle assessments, and soil erosion modeling. By integrating Landsat data, our approach enables multi-decadal monitoring, extending the analysis back before the Sentinel era.

A key advancement in our research is the introduction of a soil-specific unmixing approach, which enhances the accuracy of fractional cover estimates by selecting the most appropriate model based on the soil spectral signature of each pixel. This method proved particularly effective across diverse soil types, as demonstrated in our multi-year, Germany-wide validation. The reliance on free and open remote sensing data, rather than official soil maps, makes this approach transferable to other regions. The qualitative evaluation of fractional cover time series, supported by repeated field cover photos, underscored the practical applications of these time series and highlighted crucial information needs that can be derived from it.

Looking ahead, future research should explore the potential of hyperspectral time series to further refine these estimates. Additionally, future field campaigns should consider data collection at identical points with constant acquisition geometries to enable highly accurate derivation of cover fractions from the images. This study lays the groundwork for applying our method to large-scale assessments of cropland cover, providing an essential database for modeling carbon emissions and sequestration related to agricultural land use at the national level and beyond.

Acknowledgements

This work is part of the joint project ‘Monitoring of biodiversity in agricultural landscapes’ (MonViA), funded by the German Federal Ministry of Food and Agriculture.

Appendix

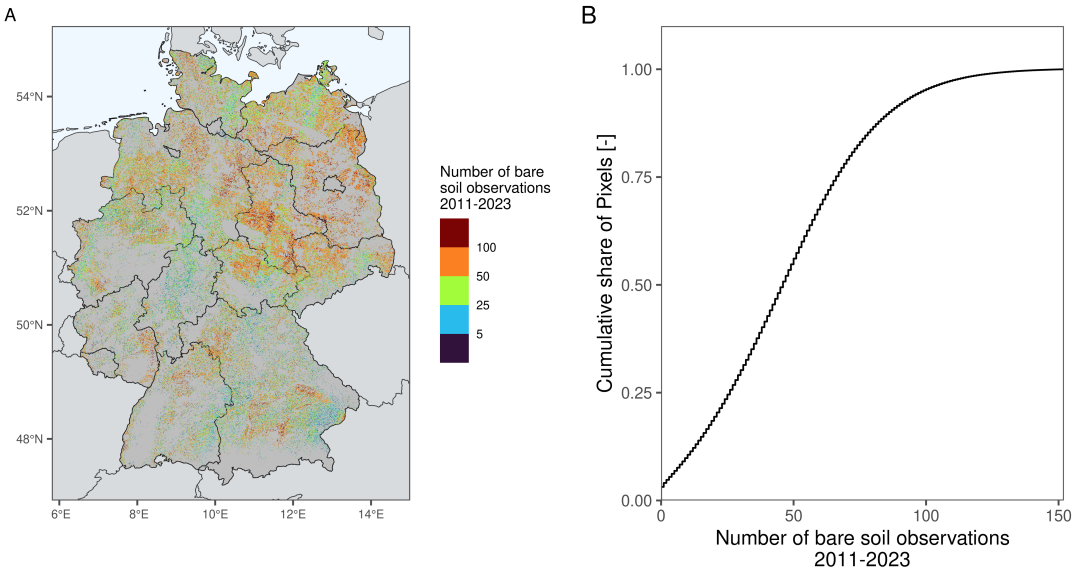

Figure 11.11: Map (A) and cumulative distribution plot (B) showing the identified number of bare soil observations per pixel during the soil reflectance composite creation.

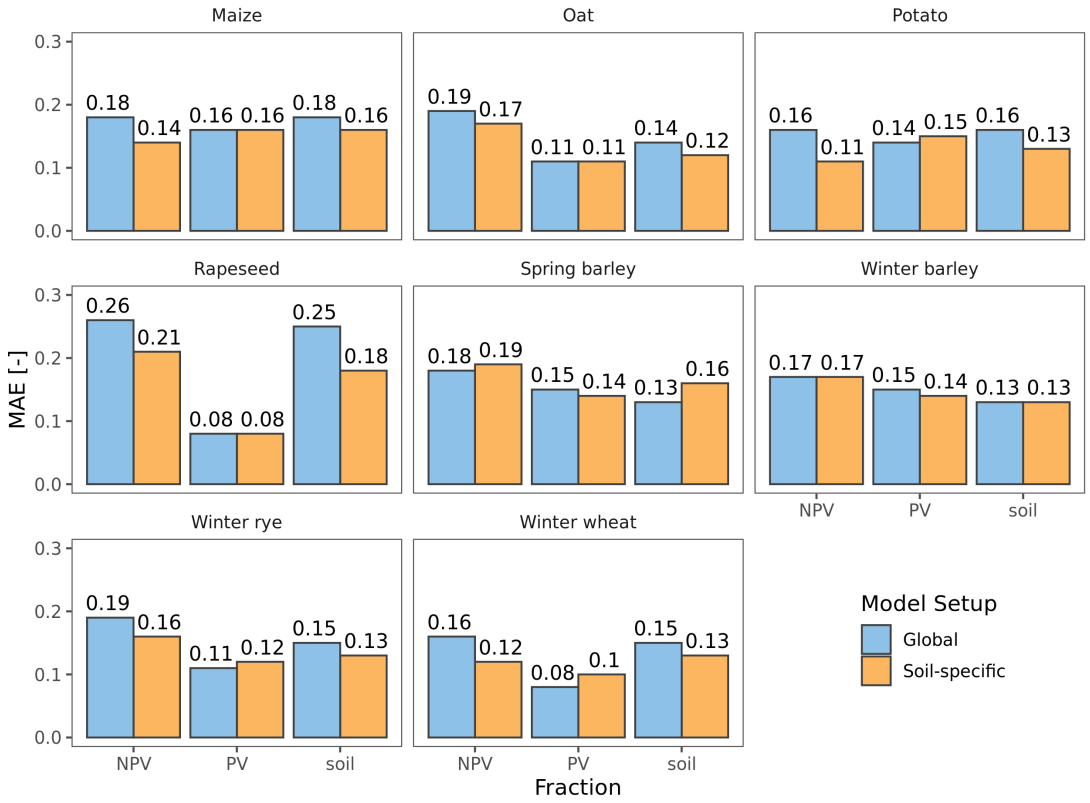

Figure 11.12: Mean absolute error for the three cover fractions stratified by crop type for the global and soil-specific unmixing models.

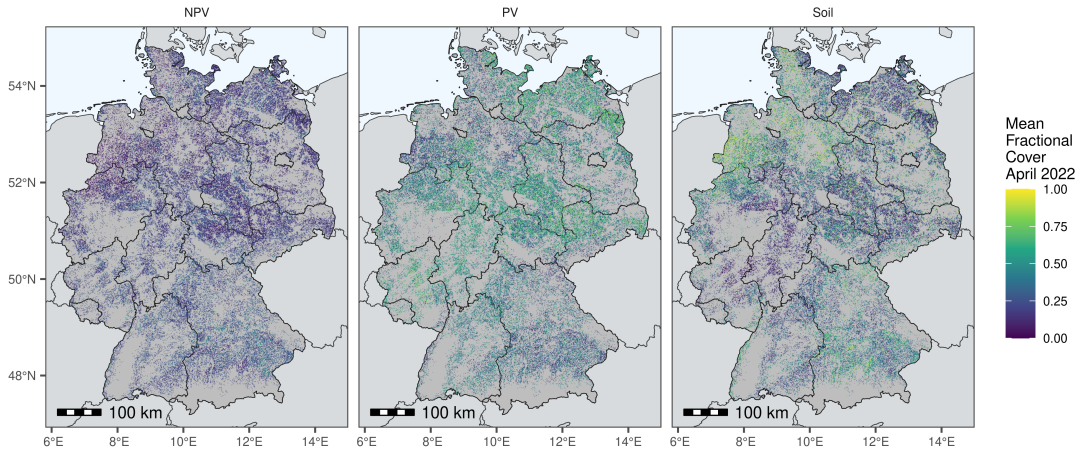

Figure 11.13: Predicted fractional cover from the soil-specific model for the whole study area. To ensure complete coverage, all predicted fractions were averaged for an entire month (April 2022).

Chapter 12

Manuscript 6: Leveraging multi-decadal satellite time series to estimate grassland age on national scale

Lukas Blickensdörfer, Tom Broeg, Felix Lobert, Marcel Schwieder, Patrick Hostert, Stefan Erasmi

Under review at the International Journal of Applied Earth Observation and Geoinformation

Abstract

Grasslands are biodiversity hotspots and provide a wide range of ecosystem services in the agricultural landscape. One important factor that drives ecosystem services is grassland age, as it influences the carbon sequestration potential but also supports biodiversity and the resilience of grassland to disturbances. Satellite image time series enable grassland age estimates, given the long-term coverage of virtual constellations and methodologies able to deal with variable satellite data availability. We present an approach to estimate grassland age in temperate regions of intensive agriculture using multidecadal time series of Landsat and Sentinel-2 data and bare soil occurrence to indicate agricultural land use periods. Every clear-sky satellite observation from 1986-2023 in grassland is classified as bare soil or non-bare soil and normalized to seasonal bare soil occurrence frequency. We then calibrate a rule set to derive grassland age based on the temporal bare soil frequency patterns. We demonstrate our approach at national scale for Germany. Grassland establishment was detected with an F-score of 78.92 ± 0.64, and differentiated from grassland persistent throughout all analyzed years with $99.21 \pm 0.02\%$ overall accuracy. The mean absolute error for grassland age is 1.3 years, and estimates remain reliable throughout low observation densities. Regional and temporal patterns align with political changes due to the collapse of the German Democratic Republic and adaptations in European agricultural policies since the 1990s. Our study demonstrates the potential of long-term image time series from satellite constellations to generate reliable data on grassland age for e.g., biodiversity or ecosystem service modelling.

1 Introduction

Grasslands support high levels of biodiversity (Petermann and Buzhdygan, 2021) and provide a wide range of ecosystem services including provisional (food and water supply), regulating (climate and water regulation, carbon sequestration, soil erosion control) and cultural (aesthetic appreciation and cultural heritage; Bengtsson et al. (2019)). The type and intensity of grassland management directly affect ecosystem function and service provision. A critical dimension is grassland age, which reflects its temporal continuity at a given location. For this study, we define it as years since initial grassland establishment after land

use (LU) change. Grassland age influences the provision of services like carbon sequestration, but also biodiversity, and resilience to disturbances (Emde et al., 2024; Wagg et al., 2022) and is therefore a crucial parameter for monitoring and informing conservation and agricultural policy.

In Europe, grassland persistence and conversion are primarily driven by changes in political framework conditions and particularly by regulations and incentives in agricultural policy (European Commission, 2020; Prishchepov et al., 2017). For instance, the collapse of socialism in Eastern Europe caused large-scale abandonment of agricultural land (Prishchepov et al., 2017), which was later countered by recultivation initiated through Common Agricultural Policies of the European Union (CAP; Pazúr et al. (2020)). At the same time, regional hotspots of decreasing or increasing grassland area are prevalent throughout Europe (Pazúr et al., 2020), emphasizing the need for accurate, large-scale, wall-to-wall monitoring.

Data on grassland use history for large areas remains sparse. Data should span multiple decades to capture large-scale grassland dynamics and provide consistent national and continental information. Agricultural statistics can offer insights into regional trends, but are rarely spatially explicit and can be challenging to harmonize across administrative boundaries. Satellite image time series can provide frequent, spatially high-resolution LU information over large areas and complement agricultural statistics. Particularly, virtual constellations that combine different generations of Landsat sensors and Sentinel-2 have enabled the mapping of agricultural LU and grassland management through dense time series data (Blickensdörfer et al., 2022; de Beurs et al., 2016; Schwieder et al., 2022). A challenge of working in the era before operational Sentinel-2 data was available, i.e., before 2016, is the limited availability and temporal irregularity of clear-sky observations (Lewińska et al., 2020). For general land cover classifications, spectral-temporal metrics (Frantz et al., 2023) or per-pixel function fitting (Zhu and Woodcock, 2014) have proven to be promising approaches. In the agricultural landscape focusing on cropland, Pham et al. (2024) use a deep learning-based framework and augmented training data. However, to our knowledge, approaches tailored to the specific characteristics of grassland for estimating grassland age and management history have not been explored.

Within the agricultural landscape, satellite image time series of grassland

exhibit unique characteristics, especially the absence of bare soil for seedbed preparation or post-harvest tillage. Thus, looking back in time, a sudden increase in bare soil frequency on grassland likely indicates preceding cropland use and the timing of grassland establishment. We use this rationale to map grassland age. While mapping of bare soil is frequently used to derive bare soil composites for cropland to predict soil parameters (Broeg et al., 2024a; Zepp et al., 2023), it has also been applied to characterize tillage practices in cropland (Azzari et al., 2019). The bare soil reflectance signal is typically dominant for weeks to months following tillage in temperate agricultural systems. This enables detecting soil management activities even in periods with relatively low observation density. We here propose to adapt these bare soil mapping approaches to the grassland domain to allow robust analysis of grassland age dating back to 1990.

The overall aim of this study is thus to explore the potential of long time series of optical satellite constellation data for determining the year of last grassland conversion after 1990 and to develop a workflow for mapping the history of grassland use across large areas. More specifically, we aim to answer the following research questions:

- How accurately can we derive grassland age in temperate regions from long time series of combined Landsat and Sentinel-2 data using bare soil occurrence as an indicator of LU?
- What are the spatial and temporal patterns of grassland age at the national scale?

2 Study area

The Central European country Germany serves as study area. It has an area of about 357,000 km^2 and covers three bioclimatic regions – oceanic, continental, and alpine – resulting in a temperature and precipitation gradient from North to South. Those environmental conditions contribute to a high diversity in grassland distribution and management. Grassland management covers the typical range of Central European grassland systems, which fall in the broad categories of meadows (regularly mown), pastures (grazed), mown pastures (grazed and mown), and extensive grassland (not actively managed through grazing or mowing). The North German Plain and the Alpine Foreland have the highest shares of grassland (Smit et al., 2008). In these regions, grassland is mostly used for dairy farming

as meadows or intensely managed as mown pastures with high productivity. In 2024, about 4.7 million ha (28.5% of agricultural land) were under permanent grassland use (grassland not included in the crop rotation for five years or longer; Umweltbundesamt (2024)). Since 1991, the total area of permanent grassland decreased by about 12% from 5.3 million ha, with most grassland conversions occurring before 2014. Since 2014, the grassland area has remained mostly stable. This can be attributed to introducing "greening" requirements in the CAP as a prerequisite for subsidy payments in 2013: permit requirements for the conversion of permanent grassland and a ban on the conversion of permanent grassland under special protection.

3 Data and methods

To map grassland age, we first map the occurrence of bare soil within the Landsat 5-9 and Sentinel-2 time series from 1986 to 2023 (Figure 12.1; sections 3.3 and 3.4). We use a random forest classification model trained on reference data derived from the European Land Use/Cover Area frame Survey (LUCAS). Second, we calculate a relative measure of bare soil frequency over time by summarizing the occurrence of bare soil per season. From the bare soil frequency time series, we estimate grassland age through a defined set of rules (section 3.5). This builds on the assumption that permanent grassland usually does not show states of bare soil. We calibrate the rule set using regionally available agricultural data from the European Integrated Administration and Control System (IACS). Finally, a holdout sample of the IACS data is used to assess the quality of grassland age estimates (section 3.6). The workflow is applied within a grassland mask derived from remote sensing-based agricultural LU maps (Blickensdörfer et al., 2022; Schwieder et al., 2024). The mask includes all permanent or temporary grassland areas between 2017 and 2023. While temporary grassland is part of a crop rotation cycle and therefore considered cropland, its inclusion minimizes the omission of grassland due to classification errors in the agricultural LU maps. Cropland areas within the mask will be identified as such by our mapping procedure.

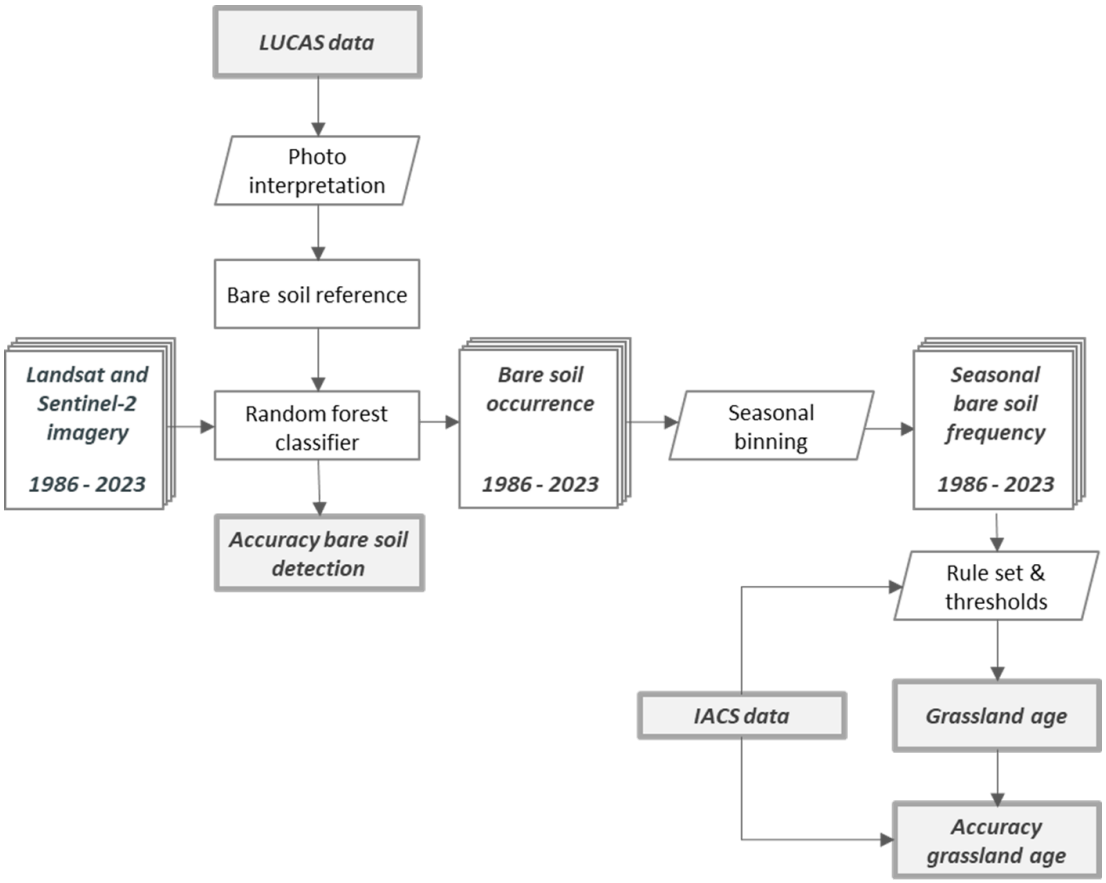

Figure 12.1: Overview of the proposed approach.

Table 12.1: Average clear-sky observations (CSO) for distinct periods of satellite data availability.

	1986–1998	1999–2011	2012	2013–2014	2015–2016	2017–2020	2021–2023
Satellite	Landsat 5	Landsat 5 Landsat 7	Landsat 7 Landsat 8	Landsat 7 Landsat 8 Sentinel-2A	Landsat 7 Landsat 8 Sentinel-2A Sentinel-2B	Landsat 8 Landsat 9 Sentinel-2A Sentinel-2B	Landsat 8 Landsat 9 Sentinel-2A Sentinel-2B
CSO-count							
Dez.–Feb.	0.6	0.9	0.2	1.2	1.9	5.1	5.2
Mar.–May	2.3	3.6	2.4	3.6	6.4	14.1	15.9
Jun.–Aug.	2.9	4.1	1.4	5.5	9.3	14.6	16.5
Sep.–Nov.	1.7	3.4	1.9	2.8	6.6	10.7	10.9

3.1 Landsat and Sentinel-2 data

We used all scenes from Landsat 5-9 and Sentinel-2 for 1986–2023 with up to 75% cloud coverage. Landsat 7 images acquired after 2020 were excluded due to the influence of orbit drift on data quality (Qiu et al., 2021). Data preprocessing was done in the Framework for Operational Radiometric Correction for Environmental Monitoring (FORCE; Frantz (2019)). The preprocessing consisted of atmospheric (Buchner et al., 2020), topographic, BRDF, and adjacency effect correction (Frantz et al., 2016a; Zhu et al., 2015). Sentinel-2 images were spatially co-registered to Landsat (Rufin et al., 2021). Clouds and cloud shadows were masked (Frantz et al., 2018; Zhu et al., 2015). Reflectance values of the Landsat imagery were adjusted to reflectance ranges of Sentinel-2, following Okujeni et al. (2024) by using slopes and intercepts from a major axis regression (Smith, 2009). Table 12.1 summarizes the average number of clear-sky observations by distinct virtual satellite constellations.

In addition to the spectral bands shared among Landsat generations and Sentinel-2 (blue, green, red, near-infrared (NIR), and shortwave-infrared (SWIR1, SWIR2)), we derived indices that proved valuable for bare soil mapping (Broeg et al., 2024a; Mzid et al., 2021; Sorenson et al., 2021; Zepp et al., 2023) and mapping of tillage practices (Azzari et al. (2019); Table 12.2).

We considered Landsat data since 1986, despite our target period starting in 1990, as our approach uses a multi-year moving window to derive grassland age. Analyzing data from 1986 ensured that the moving window boundary conditions did not influence 1990 estimates.

Table 12.2: Spectral indices computed from Landsat and Sentinel-2 bands.

Indices	Formula	Source
NBR2	$\frac{\text{SWIR1}-\text{SWIR2}}{\text{SWIR1}+\text{SWIR2}}$	Broeg et al. (2024a), Zepp et al. (2023), Azzari et al. (2019), Sorenson et al. (2021)
NDVI	$\frac{\text{NIR}-\text{RED}}{\text{NIR}+\text{RED}}$	Broeg et al. (2024a), Sorenson et al. (2021), Mzid et al. (2021)
BCC	$\frac{\text{BLUE}}{\text{RED}+\text{GREEN}+\text{BLUE}}$	Broeg et al. (2024a)
RG_diff	$\text{RED} - \text{GREEN}$	Broeg et al. (2024a)
NDWI	$\frac{\text{GREEN}-\text{NIR}}{\text{GREEN}+\text{NIR}}$	Sorenson et al. (2021)
NDI7	$\frac{\text{NIR}-\text{SWIR2}}{\text{NIR}+\text{SWIR2}}$	Sorenson et al. (2021), Azzari et al. (2019)
PV_BLUE	$\frac{\text{NIR}-\text{RED}}{\text{NIR}+\text{RED}} + \frac{\text{NIR}-\text{BLUE}}{\text{NIR}+\text{BLUE}}$	Zepp et al. (2023)
PV_IR2	$\frac{\text{NIR}-\text{RED}}{\text{NIR}+\text{RED}} + \frac{\text{NIR}-\text{SWIR2}}{\text{NIR}+\text{SWIR2}}$	Zepp et al. (2023)
BSI	$\frac{(\text{SWIR1}+\text{RED})-(\text{NIR}+\text{BLUE})}{(\text{SWIR1}+\text{RED})+(\text{NIR}+\text{BLUE})}$	Mzid et al. (2021)
EVI	$\frac{2.5\cdot(\text{NIR}-\text{RED})}{\text{NIR}+6\cdot\text{RED}-7.5\cdot\text{BLUE}+1}$	Azzari et al. (2019)
GCVI	$\frac{\text{NIR}}{\text{GREEN}} - 1$	Azzari et al. (2019)
NDI5	$\frac{\text{NIR}-\text{SWIR1}}{\text{NIR}+\text{SWIR1}}$	Azzari et al. (2019)
CRC	$\frac{\text{SWIR1}-\text{GREEN}}{\text{SWIR1}+\text{GREEN}}$	Azzari et al. (2019)
STI	$\frac{\text{SWIR1}}{\text{SWIR2}}$	Azzari et al. (2019)

3.2 Reference data

3.2.1 LUCAS data for bare soil detection

We used a dataset derived from LUCAS to train and validate the model for mapping bare soil occurrence across the Landsat/Sentinel-2 time series. LUCAS is the most comprehensive land cover database in Europe (Gallego and Bamps, 2008) and conducted in a three-year cycle since 2006. We used the harmonized field survey data from 2006-2018 published by d'Andrimont et al. (2020). We drew 2,500 sample plots from 2006, 2009, 2012, 2015, and 2018. We sampled grassland as well as cropland categories to increase the representation of bare soil, as its occurrence is rare in grassland but frequent in cropland. We excluded points without any clear-sky observation within two weeks of the surveying date and extracted corresponding spectral features for the remaining sample. Based

on the LUCAS landscape photos, each sample was interactively labeled as bare or non-bare soil by visual interpretation using the open-source software Label Studio (Tkachenko et al., 2024). Samples on land cover boundaries were excluded. The resulting dataset served as reference for bare soil detection in the satellite time series. The sample was distributed randomly throughout the years, however, lower image availability enforced a smaller sample in 2012, while the ample imagery from Landsat 8 and the inclusion of Sentinel-2 resulted in a higher sample density in 2018 (Figure 12.2). The regional bare soil observation share in the sample is influenced by the timing of the survey campaign and local agricultural management.

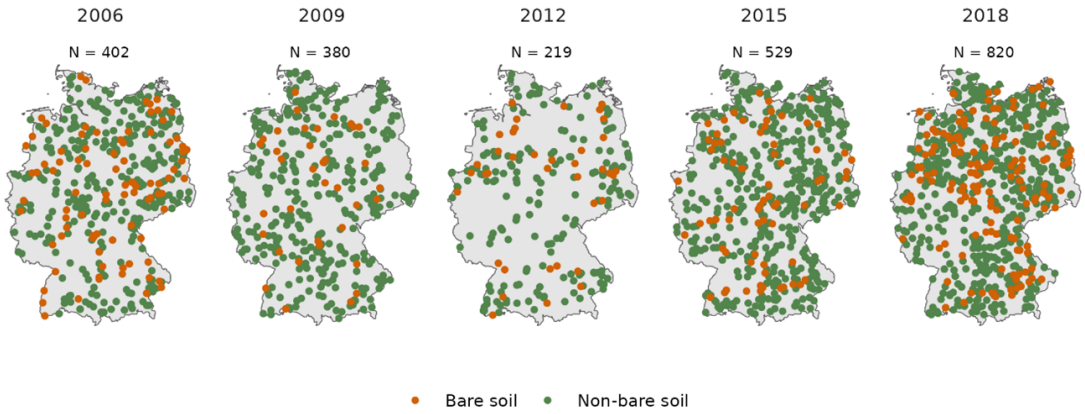

Figure 12.2: Spatial sample distribution from LUCAS data.

3.2.2 IACS data as reference for grassland age

To derive and validate grassland age, we used IACS data, developed by the European Union to manage CAP subsidies (Tóth and Kučas, 2016). The IACS includes annual parcel-level LU information submitted by farmers and landowners. We had data access for selected years and federal states in Germany (Table 12.3). However, the data use is limited by potential digitization errors or false claims, and only includes subsidized agricultural land. From the IACS data, we derived a point sample indicating grassland age for field areas present across all available years. Therefore, the IACS data was processed as follows: To reduce boundary-related digitization errors, a 20m negative buffer was applied to each parcel. We then computed the geometric intersection of parcel geometries across all years, producing sub-parcel units with complete temporal coverage. One sample point was randomly placed in each sub-parcel and attributed with the corresponding

annual LU from IACS. Points with grassland use in the final year (2022; 2021 for Hesse) were selected. If non-grassland use occurred earlier in the time series, grassland age refers to the first year after the last instance of non-grassland use.

IACS distinguishes between permanent and temporary grassland due to specific subsidy regulations. In the IACS time series, periods of temporary grassland often precede permanent grassland use. In this case, the change from temporary grassland to permanent grassland does not necessarily represent a management change, but rather a farmer's decision on which subsidies to apply for. The same applies to fallow land. To better approximate the year of LU change, we adjusted the grassland age to the start of the preceding temporary grassland or fallow period, if present. The resulting sample was randomly split into training (70%) and validation (30%) subsets. These are referred to as the IACS-training and IACS-validation sample (Table 12.3).

Table 12.3: Reference data available for grassland areas that remain stable over the analysis period (persistent grassland) and areas for which grassland age is derived from the IACS data (recent grassland).

Federal state	Years	IACS-training sampleGrassland persistent/recent	IACS-validation sampleGrassland persistent/recent
Brandenburg	2006–2022	72,910 / 3,951	31,247 / 1,693
Baden-Wurttemberg	2010–2022	81,500 / 1,180	34,929 / 506
Bavaria	2010–2022	324,210 / 9,660	138,947 / 4,140
Hesse	2010–2021	127,609 / 3,602	54,690 / 1,544
Lower Saxony	2011–2022	110,073 / 2,055	47,174 / 880
North Rhine-Westphalia	2010–2022	68,267 / 1,250	29,257 / 536
Rhineland-Palatinate	2010–2022	52,064 / 3,270	22,313 / 1,402

3.3 Bare soil detection

To detect bare soil occurrence throughout the time series, we classified each clear-sky observation as bare or non-bare soil using a random forest model trained on the LUCAS data from the five survey years. The random forest is known for its powerful predictive performance while handling data with high dimensionality and multi-collinearity well (Belgiu and Drăguţ, 2016; Breiman, 2001). However, large feature sets increase computational cost and reduce interpretability of results (Azzari et al., 2019). We therefore applied a two-step feature selection

following Azzari et al. (2019) based on a mutual information threshold of 0.9 (Ross, 2014) and a sequential forward selection as implemented in the R-package mlr3 (Wright, 2024). Based on the out-of-bag error (OBB-error; Breiman (2001); Lyons et al. (2018)), the subset with the best performance and fewest features was selected for the final classification.

The class size imbalance in the LUCAS sample, with fewer bare soil sample points, can lead to biased predictions, as random forest minimizes overall rather than class-wise error (Chen et al., 2004). To address this, we applied class weights of 0.4 for the bare soil and 0.6 for the non-bare soil class. The weights were determined empirically by testing different weight combinations: weight of bare soil class $w_{BareSoil} = 0.2,\ 0.3,\ 0.4,\ 0.5$; weight of non-bare soil class $w_{nonBareSoil} = 1 - w_{BareSoil}$. Iterative testing showed 300 trees to yield the best trade-off between computation time and OBB-error of the bare soil class (Chen et al., 2004; Lyons et al., 2018).

Bare soil pixel counts were aggregated by meteorological season (Dec. – Feb., Mar. – May, Jun. – Aug., Sep. – Nov.) to create an equidistant time series. To enhance comparability across varying observation densities, seasonal bare soil counts for each pixel were normalized by the number of available clear-sky observations. This produced a time series of seasonal bare soil observation frequency since 1986 (Figure 12.3).

3.4 Evaluation of bare soil detection

Based on the LUCAS reference data, we derived two sets of accuracy metrics: 1) All-years-accuracies, based on the predictions of the final classification model, which is trained on the complete reference data set. This provides accuracy estimates of the actual mapped output for the five years with LUCAS data. 2) Folded-years-accuracies, based on models trained on reference data from all years but one. The model was applied to the leave-out year, and predictions were assessed using the left-out reference data. This provides an estimate of model generalization to years without reference data.

For both approaches, we estimated overall, user's, and producer's accuracy with standard errors (Olofsson et al., 2014), focusing on the bare soil class. To estimate a combined accuracy representative for all years, we followed a stratified estimation approach (Stehman, 2014), and included the sensor type as strata to

account for the years each sensor provided data since 1986 (Table 12.1).

3.5 Generating grassland age information

We used the sparse occurrence of bare soil on grassland to estimate grassland age from the bare soil frequency time series. However, multiple consecutive years of grassland use do not equate zero bare soil occurrence due to potential false positive bare soil classifications and rare soil management events. To address this, we calibrated a rule set to derive grassland age using the IACS-training sample (Figure 12.3):

- A minimum seasonal bare soil share (*max_share_BS*) to filter potential bare soil classification errors,
- a maximum count of seasons with bare soil observations on grassland (*max_count_BS*),
- and the length of a temporal window to derive these counts (*temp_window_seasons*).

First, seasons with bare soil frequencies below *max_share_BS* were excluded from the time series (Figure 12.3, marker 1). Second, seasons when bare soil occurred but the *max_count_BS* threshold was not exceeded within the temporal window of *temp_window_seasons* length were considered management events on established grassland (e.g., grass sod renewal) and removed from the time series (Figure 12.3, marker 2). Finally, moving backward through the filtered time series, the first season with bare soil occurrence was assigned as the year of grassland establishment (Figure 12.3, marker 3). Many grassland areas in Germany were established before 1990 and no establishment year should be assigned in such cases. Therefore, we further assigned two classes based on the predicted establishment year: persistent grassland (established before the analysis period) and recent grassland (established within the analysis period).

All three parameters, *max_count_BS*, *temp_window_seasons*, and *max_share_BS* were chosen from a set of initial values using a grid search and the IACS-training samples (Table 12.4). The rule set parametrization was evaluated against samples of the IACS-training set using overall accuracy (Stehman, 2014), F-score (Sokolova and Lapalme, 2009), and the concordance correlation coefficient (CCC, Akoglu (2018); Lin (1989)). First, we selected those parametrizations that showed no significant difference in overall accuracy and F-score of recent grassland compared to the highest overall accuracy and

F-score. From these, we selected the final parameter set by the highest CCC comparing predicted and reference years for samples with correctly assigned recent grassland.

Table 12.4: Rule set parameters tuned in the grid search.

Parameter	Values
max_count_BS [count]	$\in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
temp_window_seasons [seasons]	$\in \{4, 8, 12, 16, 20, 24, 28, 32\}$
max_share_BS [%]	$\in \{0, 10, 20, 30, 40, 50\}$

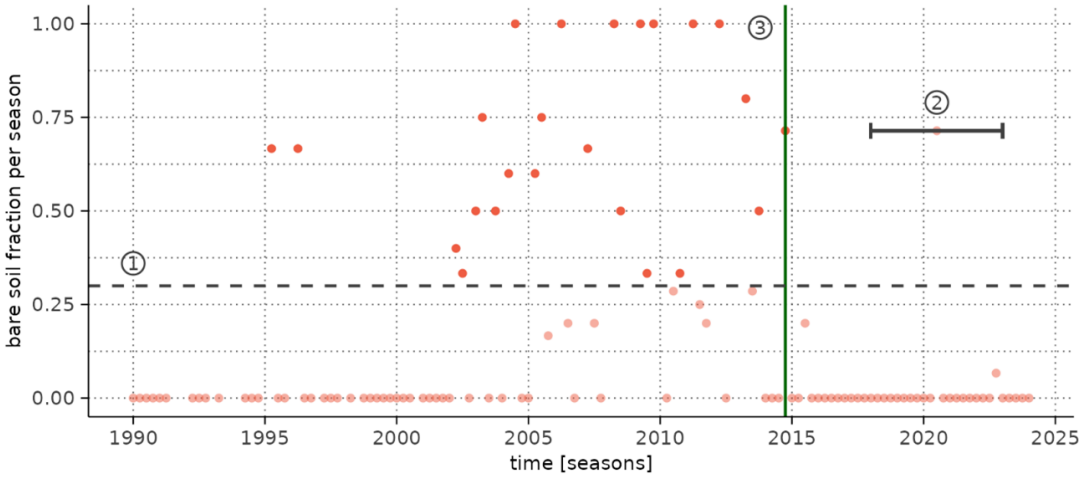

Figure 12.3: Schematic representation of a bare soil observation frequency time series and the rule set parameters used to derive grassland age. The markers represent 1) the threshold of bare soil share per season, parameterized as *max_share_BS*, 2) the moving window to filter management events parameterized by *temp_window_seasons* and *max_count_BS*, and 3) the derived year of grassland establishment.

3.6 Accuracy assessment of grassland age

To evaluate the accuracy of grassland age estimates during the rule set parametrization and the final validation, we first assessed the accuracy of persistent and recent grassland. Following, the estimated grassland age was evaluated for correctly identified recent grassland. We used the classification metrics user's,

producer's and overall accuracies (Stehman, 2014), and the F-score (Sokolova and Lapalme, 2009) with corresponding standard errors to assess accuracy for persistent grassland and recent grassland. The predicted grassland age was then evaluated against the reference year of IACS-validation data for samples correctly identified as recent grassland using the mean absolute error (MAE). The MAE is the average error irrespective of its sign and in the same unit as the prediction, in our case, years, which enables intuitive interpretation of results. As the analyzed grassland parcels might have undergone multiple LU changes before grassland establishment, we differentiated between parcels with a single versus multiple LU changes to gain a better understanding of the reliability of the approach.

3.7 Analysis of grassland age distribution

To analyze spatial and temporal patterns of grassland age, the predictions at the pixel level were aggregated to the regional scale. For this, we use Soil-Climate Regions (Roßberg et al., 2007), which divide Germany into 50 subregions with relatively homogeneous soil and climate conditions. We excluded the latest 5 years to reduce the influence of temporal grassland on the results. We computed the share of persistent grassland by region and the share established each year after 1990 for the entire study area. Additionally, we derived the share of grassland established in bi-annual intervals in each region for the period after 1990.

4 Results

4.1 Detection of bare soil

The bare soil classification throughout the time series had an all-years overall accuracy of 92.34 ± 0.6%. The feature selection revealed that only few features were needed to classify bare soil reliably throughout the time series. From the 15 tested indices, the combination of NBR2, RG_diff, NDWI, and NDI5 provided the best classification results. The bare soil class's accuracy increased from an F-score of 74.96 ± 3.21 in 2006 to 82.03 ± 2.90 in 2015 (Figure 12.4). 2018 showed deviating accuracy metrics with greater imbalance between user´s and producer´s accuracy (70.68 ± 2.96% and 87.96 ± 2.36% respectively) and a lower F-score compared to 2015 with 78.38 ± 2.05. The folded-years-accuracies closely

matched all-years-accuracies. Throughout the time series, the folded-years overall accuracy was 91.87 ± 0.6%, and the F-score of the bare soil class was estimated at 76.74 ± 1.68. The agreement between all-years-accuracy and folded-years-accuracy indicates a good model transferability to years without reference data.

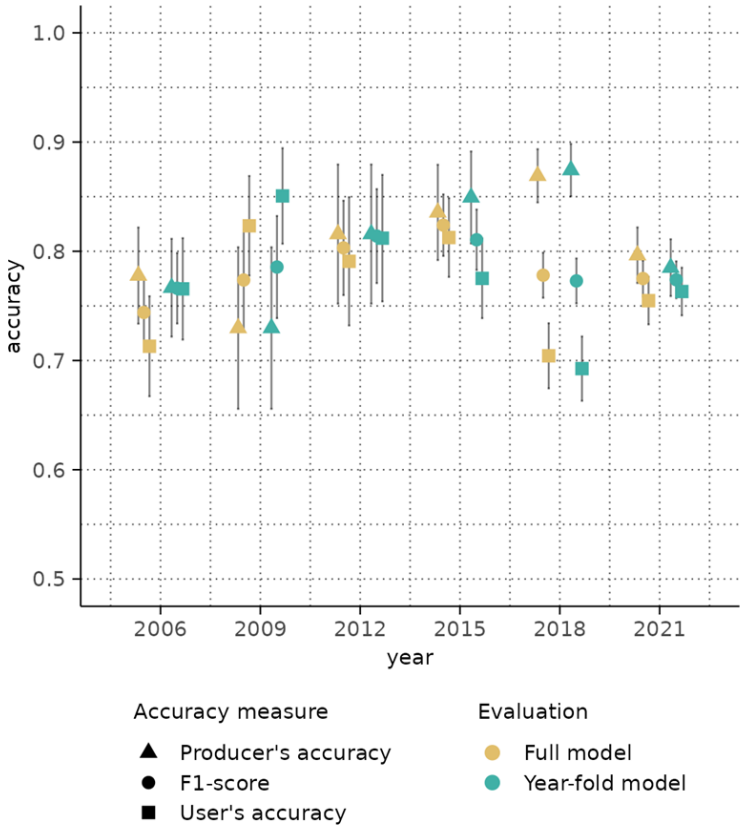

Figure 12.4: Accuracy metrics for bare soil class, for the full model and year-folded models.

4.2 Grassland age estimation

The highest accuracies of grassland age were achieved by the rule set parameters $max_{count_{BS}} = 3$, temp_window_seasons=20 and max_share_BS=30. The parametrization was selected based on the IACS-training sample, which achieved an overall accuracy of 99.00 ± 0.03% for the differentiation of persistent grassland and recent grassland, and an F-score for recent grassland of 74.03 ± 0.48. The correlation between predicted and reference grassland age showed a CCC of 0.82.

The independent validation of persistent and recent grassland, based on the IACS-validation sample, aligned with estimated accuracy during the parametriza-

Table 12.5: Accuracy metrics for detecting grassland establishment within the analysis time frame. Overall Accuracy considers persistent and recent grassland. To improve readability, user´s and producer´s accuracies for persistent grassland (98.81 ± 0.32% - 99.69 ± 0.25%) were omitted. This causes the high overall accuracies.

Federal state		Recent grassland		
	Overall accuracy [%]	Producer's accuracy [%]	User's accuracy [%]	F-score
Brandenburg	97.79 ± 0.09	62.92 ± 0.10	64.58 ± 1.85	63.74 ± 1.04
Baden-Wurttemberg	99.85 ± 0.04	83.33 ± 4.25	99.08 ± 0.06	90.91 ± 2.52
Bavaria	99.06 ± 0.06	79.86 ± 0.98	80.38 ± 2.09	80.13 ± 1.15
Hesse	99.50 ± 0.06	76.87 ± 2.36	94.53 ± 2.99	84.80 ± 1.87
Lower Saxony	99.29 ± 0.06	71.44 ± 1.47	83.42 ± 3.10	76.97 ± 1.57
North Rhine-Westphalia	99.56 ± 0.07	77.53 ± 2.56	90.20 ± 4.42	83.39 ± 2.40
Rhineland-Palatinate	98.97 ± 0.14	74.05 ± 2.28	92.24 ± 4.13	82.15 ± 2.16
Germany	99.21 ± 0.02	75.37 ± 0.66	82.82 ± 1.15	78.92 ± 0.64

tion. The overall accuracy was estimated at 99.21 ± 0.02%, and the F-score of recent grassland was estimated at 78.92 ± 0.64 (Table 12.5). User's and producer's accuracy of the recent grassland differed by about eight percentage points, with 82.82 ± 1.15% and 75.37 ± 0.66% respectively, indicating a slight overall underestimation of LU change by our approach. User's and producer's accuracy of persistent grassland were estimated above 99.51 ± 0.13% for Germany and above 98.81 ± 0.32% for each federal state.

Assessing the mapping of recent grassland by federal state revealed regional accuracy differences (Table 12.5). The most reliable areas were Baden-Wurttemberg (F-score 90.91 ± 2.52), while Brandenburg had the highest errors (F-score 63.74 ± 1.04). In Brandenburg and Bavaria, user's and producer's accuracy were more balanced, indicating less over- or underestimation of mapped recent grassland area. For Rhineland-Palatinate and Hesse, differences between user´s and producer´s accuracy were prominent, representing an overall underrepresentation of recent grassland compared to reference data.

When grassland establishment was predicted correctly within the analysis timeframe, the MAE was estimated at 1.30 years based on the IACS-validation sample. For parcels where the establishment of today's grassland was the only LU change recorded in the reference data, MAEs were between 0.84 years (Lower Saxony) and 1.46 years (Brandenburg, Figure 12.5). For parcels where multiple LU changes occurred before the establishment of the final grassland use, MAEs were between 1.13 years (North-Rhine Westphalia) and 2.12 years (Brandenburg).

Brandenburg was the only region with IACS reference data available from 2006 to 2010, and showed the largest MAE rates. However, the wider scatter around the 1:1 is observable for the entire analysis period. Predictions before 2010 still appear reliable, indicating consistent error rates throughout the time series back to 2006.

4.3 Regional and temporal patterns

From today's grassland, about 78% was established before 1990. The analysis based on the soil-climate regions showed pronounced regional differences. Regions with high shares of grassland per agricultural area were dominated by grassland established before 1990 (Figure 12.6). This was most prominent for mountain regions like the Alps and Black Forest (Figure 12.6, marker a, b), but also the Sauerland region (Figure 12.6, marker c), and the region along the North Sea coast (Figure 12.6, marker d). Most regions in Eastern Germany showed a higher share of grassland established after 1990. This was especially true for fertile soils dominated by intensive cropland use (Figure 12.6, markers e and f).

The visual inspection of mapped grassland age showed plausible patterns (Figure 12.7). Single grassland parcels contained mostly homogeneous predictions of one year and little prediction error at parcel boundaries or single-pixel noise. We mapped distinct parcel patterns in regions with high shares of grassland, from mountain and alpine regions (Figure 12.7, G, H) to low-lying and coastal regions (Figure 12.7, A, B, D). Also, in cropland-dominated regions, where grassland only occupies marginal areas between fields or along waterways (Figure 12.7, E), or regions with variable field sizes (Figure 12.7, C, F), mapped patterns appeared plausible.

The distribution of grassland age in Germany showed distinct periods during which today's grassland was established (Figure 12.8). Those distinct periods further showed regional differentiation (Figure 12.9). About 6% was established between 1990 and 1994, with a regional cluster in Eastern Germany (Figure 12.9). A second period of relatively high conversion to grassland included 2003 to 2005, with no pronounced regional differentiation. Finally, about 7% was established after 2011, with increasing shares in recent years.

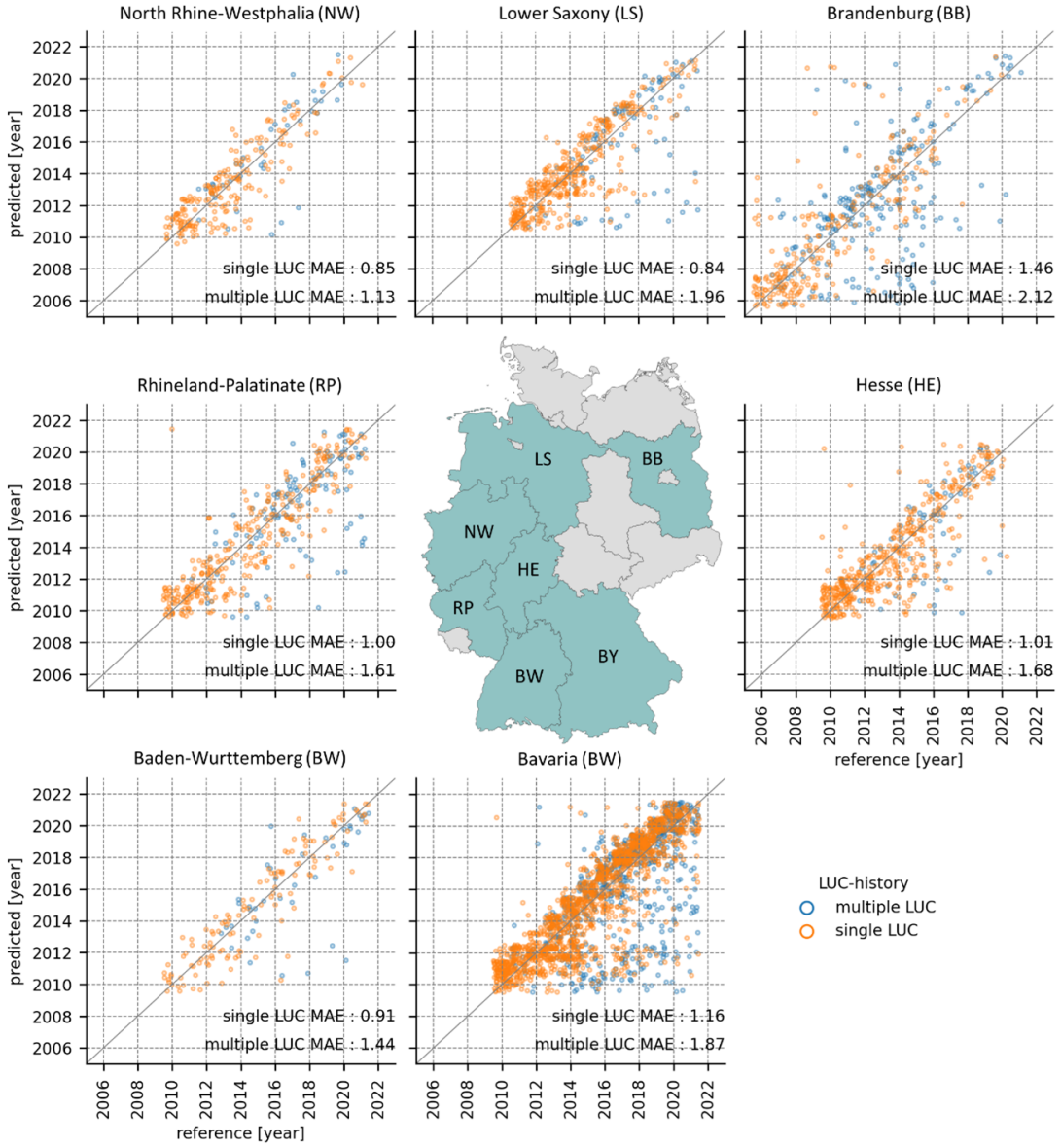

Figure 12.5: Reference and predicted grassland age for each federal state and distinct land use change (LUC) history based on the IACS-validation sample. For visualization, data is plotted with jitter of 0.5 years on both axes.

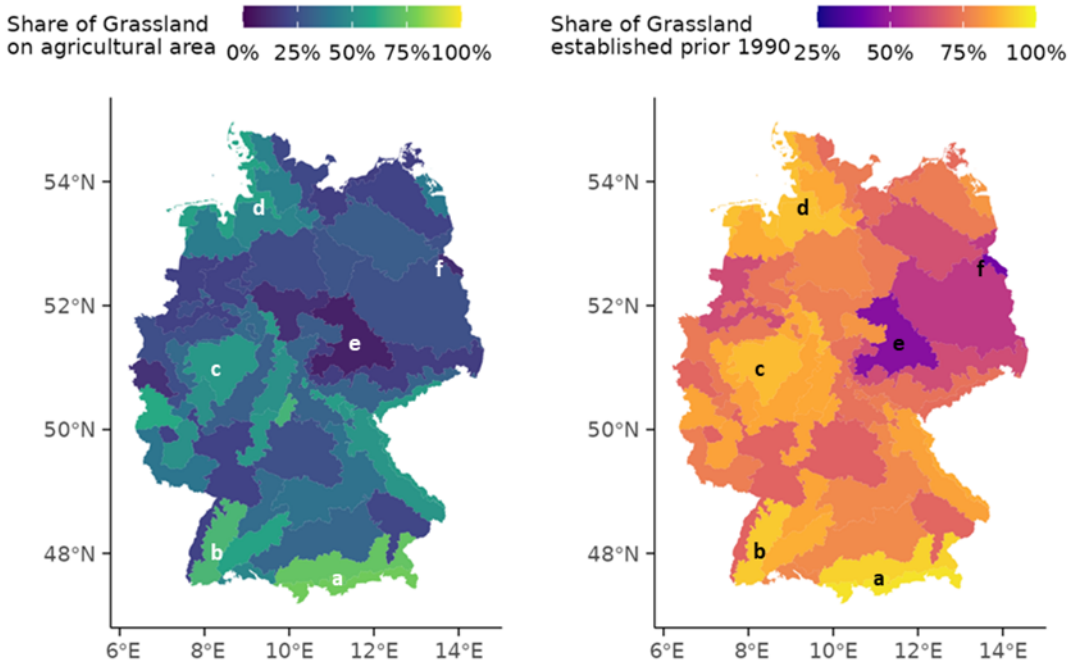

Figure 12.6: Share of grassland on agricultural area (2017) and share of grassland established before 1990 by Soil Climate Region. Total agricultural area from Schwieder et al. (2024).

5 Discussion

In this study, we demonstrated the potential of combined Sentinel-2 and Landsat time series data to estimate grassland age back to 1990. Map accuracy and spatial consistency showed that the occurrence of bare soil over time can be utilized as an indicator of agricultural LU and to derive robust area-wide estimates of grassland age. The estimated accuracy differed between federal states. The accuracy for identifying recent grassland and the user´s accuracy of grassland age were lower for Brandenburg compared to other states. Brandenburg was part of the German Democratic Republic (GDR) before 1990. GDR's agricultural policy before 1990 likely affected LU management in this region. An analysis of crop sequences in Germany by Jänicke et al. (2022) showed higher shares of leaf crops, including temporary grassland, and less cereal-dominated crop sequences in former GDR regions. This might have influenced our results, as higher shares of temporary grassland can reduce the frequency of bare soil occurrences and hamper the grassland age estimation. While not quantifiable due to missing reference data, similar effects might also be present in the other four former

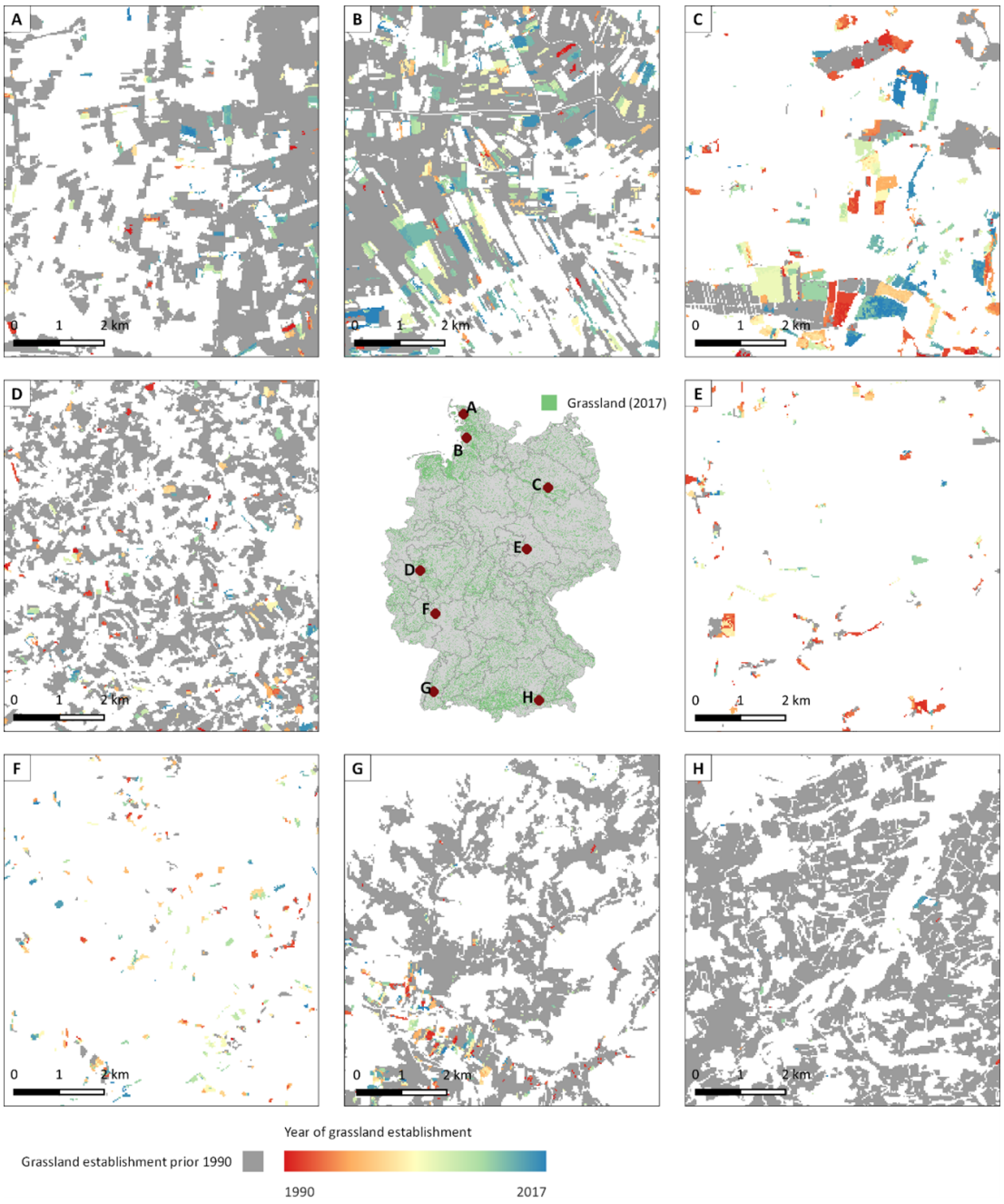

Figure 12.7: Grassland age predictions for selected regions.

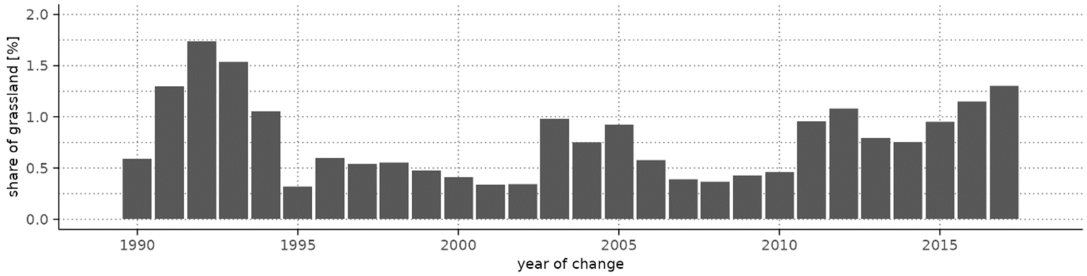

Figure 12.8: Recent grassland established in each year of the analysis time period for Germany as share of total grassland area.

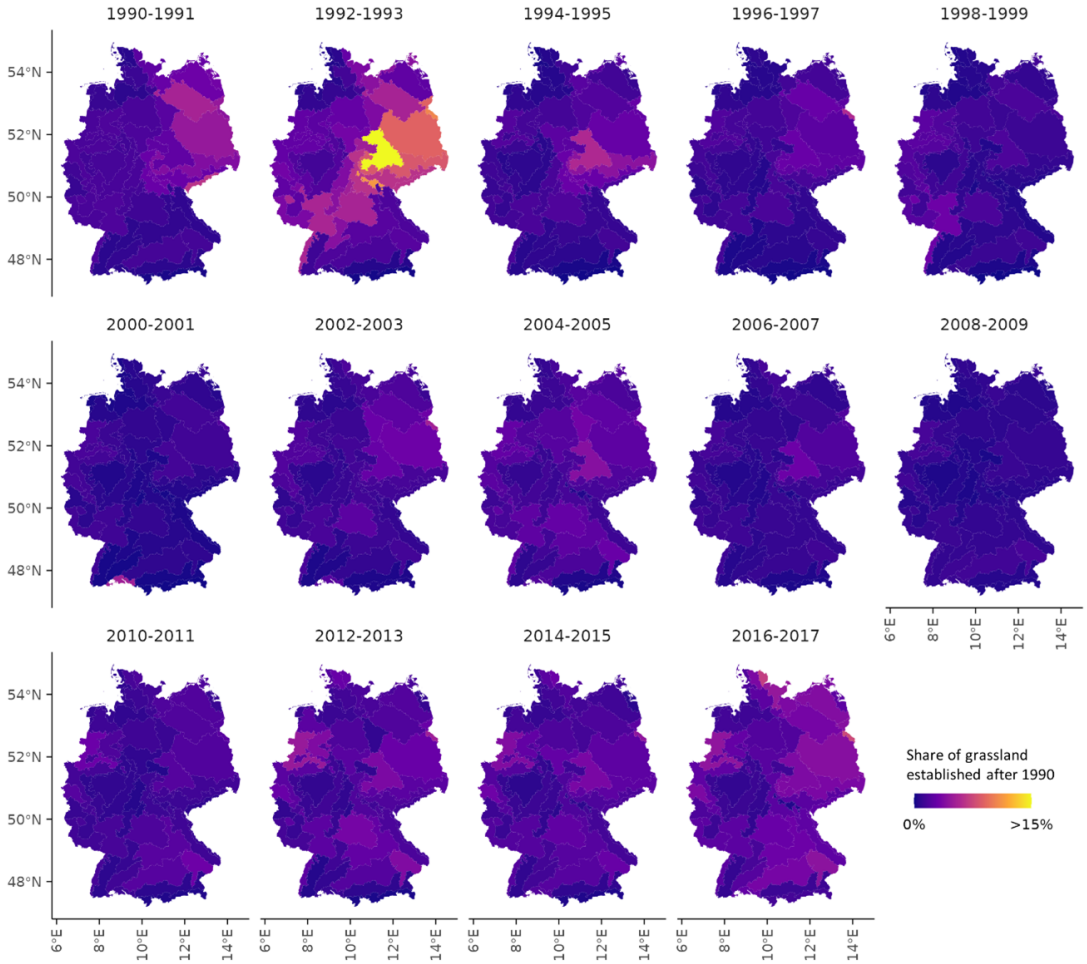

Figure 12.9: Share of grassland established within respective bi-annual intervals per Soil Climate Region.

GDR states. In addition, Brandenburg was the only state with reference data coverage back to 2006. The IACS data is not error-free either, especially in early years, which might have contributed to uncertainty in accuracy estimates. Known issues may relate to the topology of parcel boundaries (Jänicke et al., 2022) or the under-reporting of specific agricultural LUs (Fendrich et al., 2023). Despite those challenges, mapped years per parcel appeared spatially consistent (Figure 12.7, focus C).

The accuracy of grassland age estimates depends on clear-sky observations before LU change. Sparse observations increase the risk of missing the last bare soil condition before conversion, which might have led to overestimated grassland age by a few seasons or years. Further, in the case of multiple LU change events, bare soil observation free grassland periods might only be separated by short periods of cropland use. A missed bare soil condition might have shifted grassland age prediction to an earlier tillage-free LU period. However, within each of the 3-month seasonal binning periods, before Sentinel-2, about 75% of the area can be expected to have usable data (Lewińska et al., 2024). An exception is 2012, when only Landsat 7 was operational with limited imaging capacity. While it is likely that some bare soil occurrences were missed due to missing clear-sky observations, the scatter plot densities and MAE rates (Figure 12.5) suggest that this did not significantly influence the reliability of our approach.

Accurately differentiating bare soil and vegetation land cover is crucial to our approach, as classification errors can propagate into grassland age estimations. The highest error rates were expected in areas dominated by non-photosynthetically active vegetation (NPV) and dry soil conditions, as spectral differences between those classes are small (Yue et al., 2019). This becomes increasingly challenging for drought years, when grassland can appear entirely as NPV. This likely explains the reduced user's accuracy of bare soil in 2018 (Figure 12.4) and might have contributed to higher shares of predicted grassland establishment around 2003 (Figure 12.8). In 2003 and 2018, Europe experienced exceptional drought periods, negatively impacting grassland vitality (Kowalski et al., 2024). While a shift of grassland age from the drought-preceding to the drought year might occur, the rule set's noise filter and moving window should minimize effects of bare soil classification errors. Further, the high clear-sky observation densities in most drought-affected regions should also counteract effects of bare soil classification errors. However, further testing and adaptations

might be necessary when transferring the methodology to regions outside the temperate zone, such as the Mediterranean, with higher NPV shares due to drier climate conditions.

Most regional and temporal patterns seem related to political or agricultural policy changes. The high share of grassland established in the early 1990s in Eastern Germany (Figure 12.9) is linked to higher rates of land abandonment and decline in harvested areas after the collapse of the GDR (Niedertscheider et al., 2014). Other patterns primarily relate to the European CAP. After the CAP McSharry-Reform in 1992, it became a conditional requirement for large farmers receiving direct payments to exclude up to 15% of arable land from intensive production (European Commission, 1988). This policy likely contributed to the increased share of recent grassland from 1992 (Figure 12.8; Figure 12.9), especially in regions with intensive cropping systems and low grassland shares. Finally, the 2013 CAP reform introduced measures to limit the conversion of grassland to arable land (Lakner, 2025). This likely led to higher shares of grassland established after 2013, which persist until today.

6 Conclusion and outlook

We present an approach to estimate grassland age at national scale for areas under grassland use today. By analyzing multidecadal Landsat 5–9 and Sentinel-2 time series, grassland establishment can be detected back to 1990. For Germany, we found that grassland age is regionally variable, with mountainous, grassland-dominated regions having the highest shares of grassland established before 1990. Grassland age patterns after 1990 align with agricultural policy-based expectations and underline the plausibility of our results. As such, the proposed approach provides a valuable source of information for the environmental impact assessment of the CAP. It may further provide data for a more detailed representation of the grassland domain in biodiversity, and ecosystem service monitoring and modelling. More specifically, data generated through the presented approach could potentially be used for carbon modelling and climate reporting, to inform LU-dependent changes of carbon stocks over time (Emde et al., 2024; Poeplau et al., 2020b).

The proposed approach has additional potential to offer insights into grassland management, particularly soil tillage and grass sod renewal, which could be

investigated in future research. Temporally isolated bare soil observations within grasslands, masked by the current rule set, likely indicate intensive management interventions that influence soil carbon dynamics (Rutledge et al., 2014). These bare soil observations' frequency and spatial distribution could serve as a proxy for soil management intensity, providing further input for soil organic carbon modeling and carbon stock reporting frameworks. Comparable management intensity indicators were derived for cropland (Azzari et al., 2019), but they have not been explored for the grassland domain.

Acknowledgements

This research was supported by the Federal Ministry of Agriculture, Food and Regional Identity under the German Climate Protection Program 2022 through the project "KlimaFern".

Acknowledgments

Closing this very exciting chapter of my life, I'm extremely grateful to everyone who supported me throughout this journey and gave me the opportunity not only to work on this dissertation, but also to keep learning and developing as a person.

Thanks to Prof. Dr. Thomas Scholten, who saw my potential from the day I walked into his office looking for a topic for my master's thesis, and who always supported me throughout my studies.

I am grateful to Prof. Dr. Volker Hochschild for the support in evaluating this thesis.

Special thanks to Stefan Erasmi for offering me the opportunity to join the Thünen Institute and for giving me the trust, courage, and guidance to think outside the box and follow my goals.

Thanks to the amazing Thünen Earth Observation group: Marcel Schwieder, Alexander Gocht, Felix Lobert, Lukas Blickensdörfer, Jan Hemmerling, Ariane Tepaß, Javier Muro, Gideon Tetteh, Anna Köber, and Aida Taghavi, from whom I learned so much and received nothing but kindness and support from day one.

Thanks to Prof. Dr. Axel Don for always having an open ear, never refusing a request for help, and for sharing your inspiring ideas.

I'm grateful to all my collaborators from the Thünen Institute, the University of Tübingen, and the other institutions, who shared their ideas and feedback to support the manuscripts and made this work richer.

Special thanks to the Thünen Institute of Climate-Smart Agriculture, the State Research Center for Agriculture in Bavaria, and the other institutions that were involved in the collection and curation of soil samples. Without your amazing fieldwork, I would have nothing to write about.

I'm very grateful to my family and friends for always believing in me and giving me the trust and support to navigate through the highs and lows.

Lastly, I want to particularly thank my special person, Marcia, for her never-ending love and patience, each and every day. I am speaking from my heart: this would not have been possible without your support.

References

Adams, J.B., Smith, M.O., Johnson, P.E., 1986. Spectral mixture modeling: A new analysis of rock and soil types at the Viking Lander 1 Site. Journal of Geophysical Research: Solid Earth 91, 8098–8112. doi:10.1029/JB091iB08p08098

Ad-hoc-Arbeitsgruppe Boden, 2006. Bodenkundliche kartieranleitung KA5. Schweizerbart Science Publishers, Stuttgart, Germany.

AdV, 2023. Amtliches Topographisch-Kartographisches Informationssystem (ATKIS).

Akoglu, H., 2018. User's guide to correlation coefficients. Turkish journal of emergency medicine 18, 91–93. doi:10.1016/j.tjem.2018.08.001

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. Nature Communications 11, 5427. doi:10.1038/s41467-020-18887-7

Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., Bochtis, D., 2019. Remote Sensing Techniques for Soil Organic Carbon Estimation: A Review. Remote Sensing 11, 676. doi:10.3390/rs11060676

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 Regional 20, 00255. doi:10.1016/j.geodrs.2020.e00255

Atkinson, P.M., Cutler, M.E.J., Lewis, H., 1997. Mapping sub-pixel proportional land cover with AVHRR imagery. International Journal of Remote Sensing 18, 917–935. doi:10.1080/014311697218836

Azzari, G., Grassini, P., Edreira, J.I.R., Conley, S., Mourtzinis, S., Lobell, D.B., 2019. Satellite mapping of tillage practices in the North Central US region from 2005 to 2016. Remote Sensing of Environment 221, 417–429. doi:10.1016/j.rse.2018.11.010

Bai, X., Huang, Y., Ren, W., Coyne, M., Jacinthe, P.-A., Tao, B., Hui, D., Yang, J., Matocha, C., 2019. Responses of soil carbon sequestration to climate-smart agriculture practices: A meta-analysis. Global Change Biology 25, 2591–2606. doi:10.1111/gcb.14658

Baker, J.M., Ochsner, T.E., Venterea, R.T., Griffis, T.J., 2007. Tillage and soil carbon sequestration—What do we really know? Agriculture, Ecosystems & Environment 118, 1–5. doi:10.1016/j.agee.2006.05.014

Barnes, M.L., Yoder, L., Khodaee, M., 2021. Detecting Winter Cover Crops and Crop Residues in the Midwest US Using Machine Learning Classification of Thermal and Optical Imagery. Remote Sensing 13, 1998. doi:10.3390/rs13101998

Bartholomeus, H., Epema, G., Schaepman, M., 2007. Determining iron content in Mediterranean soils in partly vegetated areas, using spectral reflectance and imaging spectroscopy. International Journal of Applied Earth Observation and Geoinformation 9, 194–203. doi:10.1016/j.jag.2006.09.001

Bartholomeus, H., Kooistra, L., Stevens, A., Van Leeuwen, M., Van Wesemael, B., Ben-Dor, E., Tychon, B., 2011. Soil Organic Carbon mapping of partially vegetated agricultural fields with imaging spectroscopy. International Journal of Applied Earth Observation and Geoinformation 13, 81–88. doi:10.1016/j.jag.2010.06.009

Becker, M., Schratz, P., 2024. Mlr3spatial: Support for Spatial Objects Within the 'mlr3' Ecosystem.

Behrens, T., Förster, H., Scholten, T., Steinrücken, U., Spies, E.-D., Goldschmitt, M., 2005. Digital soil mapping using artificial neural networks. Journal of Plant Nutrition and Soil Science 168, 21–33. doi:10.1002/jpln.200421414

Behrens, T., Schmidt, K., MacMillan, R.A., Viscarra Rossel, R.A., 2018a. Multi-scale contextual spatial modelling with the Gaussian scale space. Geoderma 310, 128–137. doi:10.1016/j.geoderma.2017.09.015

Behrens, T., Schmidt, K., Viscarra Rossel, R.A., Gries, P., Scholten, T., MacMillan, R.A., 2018b. Spatial modelling with Euclidean distance fields and machine learning. European Journal of Soil Science 69, 757–770. doi:10.1111/ejss.12687

Behrens, T., Scholten, T., 2006b. Chapter 25 A Comparison of Data-Mining Techniques in Predictive Soil Mapping, in: Developments in Soil Science. Elsevier, pp. 353–617. doi:10.1016/S0166-2481(06)31025-2

Behrens, T., Scholten, T., 2006a. Digital soil mapping in Germany—a review. Journal of Plant Nutrition and Soil Science 169, 434–443. doi:10.1002/jpln.200521962

Behrens, T., Zhu, A.-X., Schmidt, K., Scholten, T., 2010. Multi-scale digital

terrain analysis and feature selection for digital soil mapping. Geoderma 155, 175–185. doi:10.1016/j.geoderma.2009.07.010

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

Bellamy, P.H., Loveland, P.J., Bradley, R.I., Lark, R.M., Kirk, G.J.D., 2005. Carbon losses from all soils across England and Wales 1978–2003. Nature 437, 245–248. doi:10.1038/nature04038

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 Sensing of Environment 113, 38–55. doi:10.1016/j.rse.2008.09.019

Ben-Dor, E., Taylor, R.G., Hill, J., Demattê, J.A.M., Whiting, M.L., Chabrillat, S., Sommer, S., 2008. Imaging Spectrometry for Soil Applications, in: Advances in Agronomy. Elsevier, pp. 321–392. doi:10.1016/S0065-2113(07)00008-9

Bengtsson, J., Bullock, J.M., Egoh, B., Everson, C., Everson, T., O'Connor, T., O'Farrell, P.J., Smith, H.G., Lindborg, R., 2019. Grasslands—more important for ecosystem services than you might think. Ecosphere (Washington, D.C) 10. doi:10.1002/ecs2.2582

BGR, 2001. Geomorphographische einheiten der bundesrepublik deutschland.

BGR, 2007b. Gruppen der bodenausgangsgesteine 1:5.000.000 (BAG5000).

BGR, 2007a. Bodenarten der Böden Deutschlands.

BGR, 2008. Karte der Bodenregionen und Bodengroßlandschaften 1:5.000.000 (BGL5000).

BGR, 2020. Bodenübersichtskarte 1:200.000 (BÜK200).

BGR, 2021. Geologische Übersichtskarte der bundesrepublik deutschland 1 : 200 000 (GÜK200).

BKG, 2024. Digitales Geländemodell Gitterweite 10 m (DGM10).

Blickensdörfer, L., Schwieder, M., Pflugmacher, D., Nendel, C., Erasmi, S., Hostert, P., 2022. Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany. Remote Sensing of Environment 269, 112831. doi:10.1016/j.rse.2021.112831

Bolton, D.K., Gray, J.M., Melaas, E.K., Moon, M., Eklundh, L., Friedl, M.A., 2020. Continental-scale land surface phenology from harmonized Landsat 8 and Sentinel-2 imagery. Remote Sensing of Environment 240, 111685.

doi:10.1016/j.rse.2020.111685

Bonannella, C., Hengl, T., Heisig, J., Parente, L., Wright, M.N., Herold, M., Bruin, S., 2022. Forest tree species distribution for Europe 2000-2020: Mapping potential and realized distributions using spatiotemporal machine learning. PeerJ 10, 13728. doi:10.7717/peerj.13728

Bouroubi, Y., Tremblay, N., Vigneault, P., Benoit, M., 2014. Linear spectral unmixing for crop and soil information extraction from a single worldview-2 image, in: 2014 IEEE Geoscience and Remote Sensing Symposium. Presented at the 2014 IEEE Geoscience and Remote Sensing Symposium, pp. 5103–5106. doi:10.1109/IGARSS.2014.6947645

Breiman, L., 2001. Random Forests. Machine Learning 45, 5–32. doi: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 Sensing 15, 876. doi:10.3390/rs15040876

Broeg, T., Don, A., Gocht, A., Scholten, T., Taghizadeh-Mehrjardi, R., Erasmi, S., 2024a. Using local ensemble models and Landsat bare soil composites for large-scale soil organic carbon maps in cropland. Geoderma 444, 116850. doi:10.1016/j.geoderma.2024.116850

Broeg, T., Don, A., Scholten, T., Erasmi, S., 2026. Reducing bias in cropland soil organic carbon and clay predictions using Sentinel-2 composites and data balancing. Remote Sensing of Environment 333, 115109. doi:10.1016/j.rse.2025.115109

Broeg, T., Don, A., Wiesmeier, M., Scholten, T., Erasmi, S., 2024b. Spatiotemporal Monitoring of Cropland Soil Organic Carbon Changes From Space. Global Change Biology 30, e17608. doi:10.1111/gcb.17608

Broeg, T., Don, A., Wiesmeier, M., Scholten, T., Erasmi, S., 2024c. Dataset and code for "Spatiotemporal monitoring of cropland soil organic carbon changes from space". Zenodo Version v1. doi:10.5281/zenodo.14191435

Buchner, J., Yin, H., Frantz, D., Kuemmerle, T., Askerov, E., Bakuradze, T., Bleyhl, B., Elizbarashvili, N., Komarova, A., Lewińska, K.E., Rizayeva, A., Sayadyan, H., Tan, B., Tepanosyan, G., Zazanashvili, N., Radeloff, V.C., 2020. Land-cover change in the Caucasus Mountains since 1987 based on the topographic correction of multi-temporal Landsat composites. Remote Sensing of Environment 248, 111967. doi:10.1016/j.rse.2020.111967

Bundesrepublik Deutschland, 2021. Gesetz zur einführung eines bundes-

klimaschutzgesetzes (klimaschutzgesetz - KSG).

Carpenter, G.A., Gopal, S., Macomber, S., Martens, S., Woodcock, C.E., 1999. A Neural Network Method for Mixture Estimation for Vegetation Mapping. Remote Sensing of Environment 70, 138–152. doi:10.1016/S0034-4257(99)00027-9

Castaldi, F., 2021. Sentinel-2 and Landsat-8 Multi-Temporal Series to Estimate Topsoil Properties on Croplands. Remote Sensing 13, 3345. doi:10.3390/rs13173345

Castaldi, F., Chabrillat, S., Don, A., Van Wesemael, B., 2019a. Soil Organic Carbon Mapping Using LUCAS Topsoil Database and Sentinel-2 Data: An Approach to Reduce Soil Moisture and Crop Residue Effects. Remote Sensing 11, 2121. doi:10.3390/rs11182121

Castaldi, F., Chabrillat, S., van Wesemael, B., 2019b. Sampling Strategies for Soil Property Mapping Using Multispectral Sentinel-2 and Hyperspectral EnMAP Satellite Data. Remote Sensing 11, 309. doi:10.3390/rs11030309

Castaldi, F., Halil Koparan, M., Wetterlind, J., Žydelis, R., Vinci, I., Savaş, A. Özge, 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 Journal of Photogrammetry and Remote Sensing 199, 40–60. doi:10.1016/j.isprsjprs.2023.03.016

Cerdan, O., Govers, G., Le Bissonnais, Y., Van Oost, K., Poesen, J., Saby, N., Gobin, A., Vacca, A., Quinton, J., Auerswald, K., Klik, A., Kwaad, F.J.P.M., Raclot, D., Ionita, I., Rejman, J., Rousseva, S., Muxart, T., Roxo, M.J., Dostal, T., 2010. Rates and spatial variations of soil erosion in Europe: A study based on erosion plot data. Geomorphology 122, 167–177. doi:10.1016/j.geomorph.2010.06.011

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

Chabrillat, S., Foerster, S., Segl, K., Beamish, A., Brell, M., Asadzadeh, S., Milewski, R., Ward, K.J., 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., Wocher, M., Hostert, P., Van Der Linden, S., Okujeni, A., Janz, A., Jakimow, B., Bracher, A., Soppa, M.A.,

Alvarado, L.M.A., Buddenbaum, H., Heim, B., Heiden, U., Moreno, J., Ong, C., Bohn, N., Green, R.O., Bachmann, M., Kokaly, R., Schodlok, M., Painter, T.H., Gascon, F., Buongiorno, F., Mottus, M., Brando, V.E., Feilhauer, H., Betz, M., Baur, S., Feckl, R., Schickling, A., Krieger, V., Bock, M., La Porta, L., Fischer, S., 2024. The EnMAP spaceborne imaging spectroscopy mission: Initial scientific results two years after launch. Remote Sensing of Environment 315, 114379. doi:10.1016/j.rse.2024.114379

Chabrillat, S., Goetz, A.F.H., Krosley, L., Olsen, H.W., 2002. Use of hyperspectral images in the identification and mapping of expansive clay soils and the role of spatial resolution. Remote Sensing of Environment 82, 431–445. doi:10.1016/S0034-4257(02)00060-3

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research 16, 321–357. doi:10.1613/jair.953

Chen, C., Liaw, A., Breiman, L., 2004. Using random forest to learn imbalanced data.

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. doi:10.1016/j.geoderma.2021.115567

Ciais, P., Wattenbach, M., Vuichard, N., Smith, P., Piao, S.L., Don, A., Luyssaert, S., Janssens, I.A., Bondeau, A., Dechow, R., Leip, A., Smith, Pc., Beer, C., Van Der Werf, G.R., Gervois, S., Van Oost, K., Tomelleri, E., Freibauer, A., Schulze, E.D., Team, C.S., 2010. The European carbon balance. Part 2: croplands. Global Change Biology 16, 1409–1428. doi:10.1111/j.1365-2486.2009.02055.x

Cleveland, W.S., Grosse, E., Shyu, W.M., 1992. Local Regression Models, in: Statistical Models in S. Routledge.

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

Cooper, S., Okujeni, A., Jänicke, C., Clark, M., van der Linden, S., Hostert, P., 2020. Disentangling fractional vegetation cover: Regression-based unmixing

of simulated spaceborne imaging spectroscopy data. Remote Sensing of Environment 246, 111856. doi:10.1016/j.rse.2020.111856

Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20, 273–297. doi:10.1007/BF00994018

Coulibaly, L.K., Guan, Q., Assoma, T.V., Fan, X., Coulibaly, N., 2021. Coupling linear spectral unmixing and RUSLE2 to model soil erosion in the Boubo coastal watershed, Côte d'Ivoire. Ecological Indicators 130, 108092. doi:10.1016/j.ecolind.2021.108092

Croft, H., Kuhn, N.J., Anderson, K., 2012. On the use of remote sensing techniques for monitoring spatio-temporal soil organic carbon dynamics in agricultural systems. CATENA, Soil erosion and the global carbon cycle 94, 64–74. doi:10.1016/j.catena.2012.01.001

d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H.I., Joebges, C., Lemoine, G., van der Velde, M., 2020. Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union. Scientific data 7, 352. doi:10.1038/s41597-020-00675-z

de Arruda, G.P., Demattê, J., Chagas, C., Fiorio, P.R., e Souza, A.B., Fongaro, C.T., 2016. Digital soil mapping using reference area and artificial neural networks. Scientia Agricola 73, 266–273. doi:10.1590/0103-9016-2015-0131

de Beurs, K.M., Owsley, B.C., Julian, J.P., 2016. Disturbance analyses of forests and grasslands with MODIS and landsat in new zealand. International Journal of Applied Earth Observation and Geoinformation 45, 42–54. doi:10.1016/j.jag.2015.10.009

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. European Journal of Soil Science 66, 121–134. doi:10.1111/ejss.12193

De Rosa, D., Ballabio, C., Lugato, E., Fasiolo, M., Jones, A., Panagos, P., 2024. Soil organic carbon stocks in European croplands and grasslands: How much have we lost in the past decade? Global Change Biology 30, e16992. doi:10.1111/gcb.16992

Delaney, B., Tansey, K., Whelan, M., 2025. Satellite Remote Sensing Techniques and Limitations for Identifying Bare Soil. Remote Sensing 17, 630. doi:10.3390/rs17040630

Delegido, J., Verrelst, J., Rivera, J.P., Ruiz-Verdú, A., Moreno, J., 2015. Brown and green LAI mapping through spectral indices. International Journal of Applied Earth Observation and Geoinformation 35, 350–358. doi:10.1016/j.jag.2014.10.001

Demattê, J.A.M., Alves, M.R., Terra, F. da S., Bosquilia, R.W.D., Fongaro, C.T., Barros, P.P. da S., 2016. Is It Possible to Classify Topsoil Texture Using a Sensor Located 800 km Away from the Surface? Revista Brasileira de Ciência do Solo 40. doi:10.1590/18069657rbcs20150335

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 Sensing of Environment 212, 161–175. doi:10.1016/j.rse.2018.04.047

Demattê, J.A.M., R. Huete, A., Guimaraes Ferreira Jr, L., Rafael Nanni, M., Cardoso Alves, M., Ricardo Fiorio, P., 2009. Methodology for Bare Soil Detection and Discrimination by Landsat TM Image. The Open Remote Sensing Journal 2, 24–35. doi:10.2174/1875413901002010024

Demattê, J.A.M., Rizzo, R., Rosin, N.A., Poppiel, R.R., Novais, J.J.M., Amorim, M.T.A., Rodriguez-Albarracín, H.S., Rosas, J.T.F., Bartsch, B. dos A., Vogel, L.G., Minasny, B., Grunwald, S., Ge, Y., Ben-Dor, E., Gholizadeh, A., Gomez, C., Chabrillat, S., Francos, N., Fiantis, D., Belal, A., Tsakiridis, N., Kalopesa, E., Naimi, S., Ayoubi, S., Tziolas, N., Das, B.S., Zalidis, G., Francelino, M.R., Mello, D.C. de, Hafshejani, N.A., Peng, Y., Ma, Y., Coblinski, J.A., Wadoux, A.M.J.-C., Savin, I., Malone, B.P., Karyotis, K., Milewski, R., Vaudour, E., Wang, C., Salama, E.S.M., Shepherd, K.D., 2025. A global soil spectral grid based on space sensing. Science of The Total Environment 968, 178791. doi:10.1016/j.scitotenv.2025.178791

Demattê, J.A.M., Safanelli, J.L., Poppiel, R.R., Rizzo, R., Silvero, N.E.Q., Mendes, W. de S., Bonfatti, B.R., Dotto, A.C., Salazar, D.F.U., Mello, Fellipe Alcântara de Oliveira, Da Paiva, A.F.S., Souza, A.B., Santos, N.V. dos, Maria Nascimento, C., Mello, D.C., Bellinaso, H., Gonzaga Neto, L., Amorim, M.T.A., Resende, M.E.B., Da Vieira, J.S., Queiroz, L.G., Gallo, B.C., Sayão, V.M., Da Lisboa, C.J.S., 2020. Bare Earth's Surface Spectra as a Proxy for Soil Resource Monitoring. Scientific Reports 10, 4461. doi:10.1038/s41598-020-61408-1

Dennison, P.E., Qi, Y., Meerdink, S.K., Kokaly, R.F., Thompson, D.R., Daugh-

try, C.S.T., Quemada, M., Roberts, D.A., Gader, P.D., Wetherley, E.B., Numata, I., Roth, K.L., 2019. Comparison of Methods for Modeling Fractional Cover Using Simulated Satellite Hyperspectral Imager Spectra. Remote Sensing 11, 2072. doi:10.3390/rs11182072

Destatis, 2022. Land- und Forstwirtschaft, Fischerei - Bodenfläche nach Art der tatsächlichen Nutzung.

Deutscher Wetterdienst, 2022. Open-data-server des deutschen wetterdienstes (DWD).

Deutscher Wetterdienst, 2024. German Climate Atlas.

Diek, S., Fornallaz, F., Schaepman, M.E., 2017. Barest Pixel Composite for Agricultural Areas Using Landsat Time Series. Remote Sensing 9, 1245. doi:10.3390/rs9121245

Diek, S., Schaepman, M., De Jong, R., 2016. Creating Multi-Temporal Composites of Airborne Imaging Spectroscopy Data in Support of Digital Soil Mapping. Remote Sensing 8, 906. doi:10.3390/rs8110906

Don, A., Drexler, S., Poeplau, C., 2024a. Indikatoren zur Bewertung von Humusgehalten für die Bodengesundheit. Bodenschutz 4. doi:10.37307/j.1868-7741.2024.01.04

Don, A., Schumacher, J., Scherer-Lorenzen, M., Scholten, T., Schulze, E.-D., 2007. Spatial and vertical variation of soil carbon at two grassland sites — Implications for measuring soil carbon stocks. Geoderma 141, 272–282. doi:10.1016/j.geoderma.2007.06.003

Don, A., Seidel, F., Leifeld, J., Kätterer, T., Martin, M., Pellerin, S., Emde, D., Seitz, D., Chenu, C., 2024b. Carbon sequestration in soils and climate change mitigation—Definitions and pitfalls. Global Change Biology 30, e16983. doi:10.1111/gcb.16983

Dudley, K.L., Dennison, P.E., Roth, K.L., Roberts, D.A., Coates, A.R., 2015. A multi-temporal spectral library approach for mapping vegetation species across spatial and temporal phenological gradients. Remote Sensing of Environment, Special Issue on the Hyperspectral Infrared Imager (HyspIRI) 167, 121–134. doi:10.1016/j.rse.2015.05.004

Dudley, N., Alexander, S., 2017. Agriculture and biodiversity: A review. Biodiversity 18, 45–49. doi:10.1080/14888386.2017.1351892

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 com-

posite by assessing surface conditions and uncertainties. Geoderma 116128. doi:10.1016/j.geoderma.2022.116128

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, 116128. doi:10.1016/j.geoderma.2022.116128

Dvorakova, K., Heiden, U., van Wesemael, B., 2021. Sentinel-2 Exposed Soil Composite for Soil Organic Carbon Prediction. Remote Sensing 13, 1791. doi:10.3390/rs13091791

Dvorakova, K., Shi, P., Limbourg, Q., van Wesemael, B., 2020. Soil Organic Carbon Mapping from Remote Sensing: The Effect of Crop Residues. Remote Sensing 12, 1913. doi:10.3390/rs12121913

Elith, J., Phillips, S.J., Hastie, T., Dud\'\ik, M., Chee, Y.E., Yates, C.J., 2011. A statistical explanation of MaxEnt for ecologists. Diversity and distributions 17, 43–57.

Emadi, M., Taghizadeh-Mehrjardi, R., Cherati, A., Danesh, M., Mosavi, A., Scholten, T., 2020. Predicting and mapping of soil organic carbon using machine learning algorithms in northern iran. Remote Sensing 12, 2234. doi:10.3390/rs12142234

Emde, D., Poeplau, C., Don, A., Heilek, S., Schneider, F., 2024. The centennial legacy of land-use change on organic carbon stocks of German agricultural soils. Global Change Biology 30, e17444. doi:10.1111/gcb.17444

European Commission, 1988. Commission Regulation (EEC) No 1272/88 of 29 April 1988 laying down detailed rules for applying the set-aside incentive scheme for arable land.

European Commission, 2020. Directorate-general for agriculture and rural development and alliance environnement: Evaluation of the impact of the CAP on habitats, landscapes, biodiversity – final report. doi:10.2762/818843

European Commission, 2022. Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL establishing a Union certification framework for carbon removals (No. COM/2022/672).

European Commission, 2023. Proposal for a Directive of the European Parliament and of the Council on Soil Monitoring and Resilience (Soil Monitoring Law).

European Commission, Joint Research Centre, 2020. LUCAS soil, the land

use/cover area frame survey: Soil data 2009–2018.

European Union, 2024. Regulation (EU) 2024/3012 of the European Parliament and of the Council of 27 November 2024 establishing a Union certification framework for permanent carbon removals, carbon farming and carbon storage in products.

Eurostat, 2024. Land cover statistics.

Farr, T.G., Kobrick, M., 2000. Shuttle radar topography mission produces a wealth of data. Eos, Transactions American Geophysical Union 81, 583–585. doi:10.1029/EO081i048p00583

Fathizad, H., Taghizadeh-Mehrjardi, R., Hakimzadeh Ardakani, M.A., Zeraatpisheh, M., Heung, B., Scholten, T., 2022. Spatiotemporal Assessment of Soil Organic Carbon Change Using Machine-Learning in Arid Regions. Agronomy 12, 628. doi:10.3390/agronomy12030628

Fathololoumi, S., Vaezi, A.R., Alavipanah, S.K., Ghorbani, A., Saurette, D., Biswas, A., 2020. Improved digital soil mapping with multitemporal remotely sensed satellite data fusion: A case study in Iran. The Science of the total environment 721, 137703. doi:10.1016/j.scitotenv.2020.137703

Federal Statistical Office, 2024. Field crops and grassland.

Feeney, C.J., Cosby, B.J., Robinson, D.A., Thomas, A., Emmett, B.A., Henrys, P., 2022. Multiple soil map comparison highlights challenges for predicting topsoil organic carbon concentration at national scale. Scientific Reports 12, 1379. doi:10.1038/s41598-022-05476-5

Fendrich, A.N., Matthews, F., Van Eynde, E., Carozzi, M., Li, Z., d' Andrimont, R., Lugato, E., Martin, P., Ciais, P., Panagos, P., 2023. From regional to parcel scale: A high-resolution map of cover crops across Europe combining satellite data with statistical surveys. Science of The Total Environment 873, 162300. doi:10.1016/j.scitotenv.2023.162300

Fiorio, P.R., Demattê, J.A.M., 2009. Orbital and laboratory spectral data to optimize soil analysis. Scientia Agricola 66, 250–257. doi:10.1590/S0103-90162009000200015

Fisher, A., Rudin, C., Dominici, F., 2019. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S., Coe, M.T., Daily, G.C., Gibbs, H.K., Helkowski, J.H., Holloway,

T., Howard, E.A., Kucharik, C.J., Monfreda, C., Patz, J.A., Prentice, I.C., Ramankutty, N., Snyder, P.K., 2005. Global Consequences of Land Use. Science 309, 570–574. doi:10.1126/science.1111772

Frantz, D., 2019. FORCE—Landsat + Sentinel-2 Analysis Ready Data and Beyond. Remote Sensing 11, 1124. doi:10.3390/rs11091124

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 Sensing of Environment 215, 471–481. doi:10.1016/j.rse.2018.04.046

Frantz, D., Roder, A., Udelhoven, T., Schmidt, M., 2015. Enhancing the Detectability of Clouds and Their Shadows in Multitemporal Dryland Landsat Imagery: Extending Fmask. IEEE Geoscience and Remote Sensing Letters 12, 1242–1246. doi:10.1109/LGRS.2015.2390673

Frantz, D., Röder, A., Stellmes, M., Hill, J., 2016a. An operational radiometric landsat preprocessing framework for large-area time series applications. IEEE Transactions on Geoscience and Remote Sensing 54, 3928–3943. doi:10.1109/TGRS.2016.2530856

Frantz, D., Rufin, P., Janz, A., Ernst, S., Pflugmacher, D., Schug, F., Hostert, P., 2023. Understanding the robustness of spectral-temporal metrics across the global Landsat archive from 1984 to 2019 – a quantitative evaluation. Remote Sensing of Environment 298, 113823. doi:10.1016/j.rse.2023.113823

Frantz, D., Stellmes, M., Roder, A., Udelhoven, T., Mader, S., Hill, J., 2016b. Improving the Spatial Resolution of Land Surface Phenology by Fusing Medium- and Coarse-Resolution Inputs. IEEE Transactions on Geoscience and Remote Sensing 54, 4153–4164. doi:10.1109/TGRS.2016.2537929

Fuß, R., 2025. Hohe Treibhausgas-Emissionen aus Landnutzung.

Gallego, J., Bamps, C., 2008. Using CORINE land cover and the point survey LUCAS for area estimation. International Journal of Applied Earth Observation and Geoinformation 10, 467–475. doi:10.1016/j.jag.2007.11.001

GDAL/OGR contributors, 2022. GDAL/OGR geospatial data abstraction software library. doi:10.5281/zenodo.5884351

Gehl, R.J., Rice, C.W., 2007. Emerging technologies for in situ measurement of soil carbon. Climatic Change 80, 43–54. doi:10.1007/s10584-006-9150-2

Gensior, A., Drexler, S., Fuß, R., Stümer, W., Rüter, S., 2025. Treibhausgasemissionen durch Landnutzung, Landnutzungsänderung und Forstwirtschaft

(LULUCF).

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 International 36, 121–136. doi:10.1080/10106049.2019.1595177

Gerighausen, H., Menz, G., Kaufmann, H., 2012. Spatially Explicit Estimation of Clay and Organic Carbon Content in Agricultural Soils Using Multi-Annual Imaging Spectroscopy Data. Applied and Environmental Soil Science 2012, 1–23. doi:10.1155/2012/868090

Gholizadeh, A., Neumann, C., Chabrillat, S., van Wesemael, B., Castaldi, F., Borůvka, L., Sanderman, J., Klement, A., Hohmann, C., 2021. Soil organic carbon estimation using VNIR–SWIR spectroscopy: The effect of multiple sensors and scanning conditions. Soil and Tillage Research 211. doi:10.1016/j.still.2021.105017

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 Sensing of Environment 218, 89–103. doi: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 Sensing of Environment 22, 343–365. doi:10.1016/0034-4257(87)90088-5

Gitelson, A., Merzlyak, M.N., 1994. Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. And Acer platanoides L. Leaves. Spectral Features and Relation to Chlorophyll Estimation. Journal of Plant Physiology 143, 286–292. doi:10.1016/S0176-1617(11)81633-0

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. doi:10.1016/j.geoderma.2022.115959

Gräler, B., Pebesma, E., Heuvelink, G., 2016. Spatio-Temporal Interpolation using gstat. The R Journal 8, 204–218.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., Chabrillat, S., Kuester, T., Hollstein, A., Rossner, G., Chlebek, C., Straif, C., Fischer, S.,

Schrader, S., Storch, T., Heiden, U., Mueller, A., Bachmann, M., Mühle, H., Müller, R., Habermeyer, M., Ohndorf, A., Hill, J., Buddenbaum, H., Hostert, P., Van Der Linden, S., Leitão, P., Rabe, A., Doerffer, R., Krasemann, H., Xi, H., Mauser, W., Hank, T., Locherer, M., Rast, M., Staenz, K., Sang, B., 2015. The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. Remote Sensing 7, 8830–8857. doi:10.3390/rs70708830

Guerschman, J.P., Hill, M.J., Leys, J., Heidenreich, S., 2020. Vegetation cover dependence on accumulated antecedent precipitation in Australia: Relationships with photosynthetic and non-photosynthetic vegetation fractions. Remote Sensing of Environment 240, 111670. doi:10.1016/j.rse.2020.111670

Guerschman, J.P., Hill, M.J., Renzullo, L.J., Barrett, D.J., Marks, A.S., Botha, E.J., 2009. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. Remote Sensing of Environment 113, 928–945. doi:10.1016/j.rse.2009.01.006

Guerschman, J.P., Scarth, P.F., McVicar, T.R., Renzullo, L.J., Malthus, T.J., Stewart, J.B., Rickards, J.E., Trevithick, R., 2015. Assessing the effects of site heterogeneity and soil properties when unmixing photosynthetic vegetation, non-photosynthetic vegetation and bare soil fractions from Landsat and MODIS data. Remote Sensing of Environment 161, 12–26. doi:10.1016/j.rse.2015.01.021

Guo, L., Sun, X., Fu, P., Shi, T., Dang, L., Chen, Y., Linderman, M., Zhang, G., Zhang, Y., Jiang, Q., Zhang, H., Zeng, C., 2021. Mapping soil organic carbon stock by hyperspectral and time-series multispectral remote sensing images in low-relief agricultural areas. Geoderma 398. doi:10.1016/j.geoderma.2021.115118

Haddaway, N.R., Hedlund, K., Jackson, L.E., Kätterer, T., Lugato, E., Thomsen, I.K., Jørgensen, H.B., Isberg, P.-E., 2017. How does tillage intensity affect soil organic carbon? A systematic review. Environmental Evidence 6, 30. doi:10.1186/s13750-017-0108-9

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 Sensing 14, 4526. doi:10.3390/rs14184526

Heikkinen, J., Ketoja, E., Nuutinen, V., Regina, K., 2013. Declining trend of

carbon in Finnish cropland soils in 1974–2009. Global Change Biology 19, 1456–1469. doi:10.1111/gcb.12137

Helfenstein, A., Mulder, V.L., Heuvelink, G.B.M., Hack-ten Broeke, M.J.D., 2024. Three-dimensional space and time mapping reveals soil organic matter decreases across anthropogenic landscapes in the Netherlands. Communications Earth & Environment 5, 1–16. doi:10.1038/s43247-024-01293-y

Hengl, T., 2020. GSIF: Global soil information facilities.

Hengl, T., 2021. Extrapolation is tough for trees (tree-based learners), combining learners of different type makes it less tough.

Hengl, T., Consoli, D., Tian, X., Nauman, T.W., Nussbaum, M., Isik, M.S., Parente, L., Ho, Y.-F., Simoes, R., Gupta, S., Samuel-Rosa, A., Zborowski Horst, T., Safanelli, J.L., Harris, N., 2025. OpenLandMap-soildb: Global soil information at 30 m spatial resolution for 2000–2022+ based on spatiotemporal Machine Learning and harmonized legacy soil samples and observations. doi:10.5194/essd-2025-336

Hengl, T., MacMillan, R.A., 2019. Predictive soil mapping with R. Zenodo. doi:10.5281/zenodo.2567061

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

Hengl, T., Miller, M.A.E., Križan, J., Shepherd, K.D., Sila, A., Kilibarda, M., Antonijević, O., Glušica, L., Dobermann, A., Haefele, S.M., McGrath, S.P., Acquah, G.E., Collinson, J., Parente, L., Sheykhmousa, M., Saito, K., Johnson, J.-M., Chamberlin, J., Silatsa, F.B.T., Yemefack, M., Wendt, J., MacMillan, R.A., Wheeler, I., Crouch, J., 2021. African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. Scientific Reports 11, 6130. doi:10.1038/s41598-021-85639-y

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. doi:10.7717/peerj.5518

Hengl, T., Parente, L., Bouasria, A., Wheeler, I., 2022. Spatial sampling and resampling for machine learning. doi:10.5281/zenodo.5886678

Heuvelink, G.B.M., Angelini, M.E., Poggio, L., Bai, Z., Batjes, N.H., Bosch, R., Bossio, D., Estella, S., Lehmann, J., Olmedo, G.F., Sanderman, J., 2021. Machine learning in space and time for modelling soil organic carbon change. European Journal of Soil Science 72, 1607–1623. doi:10.1111/ejss.12998

Heuvelink, G.B.M., Webster, R., 2001. Modelling soil variation: Past, present, and future. Geoderma 100, 269–301. doi:10.1016/s0016-7061(01)00025-8

Hijmans, R.J., 2025. Terra: Spatial Data Analysis.

Hively, W.D., Lee, S., Sadeghi, A.M., McCarty, G.W., Lamb, B.T., Soroka, A., Keppler, J., Yeo, I.-Y., Moglen, G.E., 2020. Estimating the effect of winter cover crops on nitrogen leaching using cost-share enrollment data, satellite remote sensing, and Soil and Water Assessment Tool (SWAT) modeling. Journal of Soil and Water Conservation 75, 362–375. doi:10.2489/jswc.75.3.362

Hong, Y., Chen, S., Chen, Y., Linderman, M., Mouazen, A.M., Liu, Y., Guo, L., Yu, L., Liu, Y., Cheng, H., Liu, Y., 2020. Comparing laboratory and airborne hyperspectral data for the estimation and mapping of topsoil organic carbon: Feature selection coupled with random forest. Soil and Tillage Research 199. doi:10.1016/j.still.2020.104589

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment 25, 295–309. doi:10.1016/0034-4257(88)90106-X

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. Computers and Electronics in Agriculture 32, 149–165. doi:10.1016/S0168-1699(01)00163-6

IPCC, 2022. Climate Change and Land: IPCC Special Report on Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems, 1st ed. Cambridge University Press. doi:10.1017/9781009157988

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

Jaconi, A., Vos, C., Don, A., 2019. Near infrared spectroscopy as an easy and precise method to estimate soil texture. Geoderma 337, 906–913. doi:10.1016/j.geoderma.2018.10.038

Jänicke, C., Goddard, A., Stein, S., Steinmann, H.-H., Lakes, T., Nendel, C., Müller, D., 2022. Field-level land-use data reveal heterogeneous crop se-

quences with distinct regional differences in Germany. European Journal of Agronomy 141, 126632. doi:10.1016/j.eja.2022.126632

Janzen, H.H., 2006. The soil carbon dilemma: Shall we hoard it or use it? Soil Biology and Biochemistry 38, 419–424. doi:10.1016/j.soilbio.2005.10.008

JAXA EORC, 2021. ALOS global digital surface model (DSM) 1–21.

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. 1–10. doi:10.1145/3139958.3140044

Johnson, J.M.F., Franzluebbers, A.J., Weyers, S.L., Reicosky, D.C., 2007. Agricultural opportunities to mitigate greenhouse gas emissions. Environmental Pollution 150, 107–124. doi:10.1016/j.envpol.2007.06.030

Johnston, A.E., Poulton, P.R., Coleman, K., 2009. Chapter 1 Soil Organic Matter, in: Advances in Agronomy. Elsevier, pp. 1–57. doi:10.1016/S0065-2113(08)00801-8

Jones, A., Fernández-Ugalde, O., Scarpa, S., 2020. LUCAS 2015 topsoil survey: Presentation of dataset and results, EUR. Publications Office of the European Union, Luxembourg.

Kassambara, A., 2020. Ggpubr: 'ggplot2' based publication ready plots.

Kidd, D., Searle, R., Grundy, M., McBratney, A., Robinson, N., O'Brien, L., Zund, P., Arrouays, D., Thomas, M., Padarian, J., Jones, E., Bennett, J.M., Minasny, B., Holmes, K., Malone, B.P., Liddicoat, C., Meier, E.A., Stockmann, U., Wilson, P., Wilford, J., Payne, J., Ringrose-Voase, A., Slater, B., Odgers, N., Gray, J., Van Gool, D., Andrews, K., Harms, B., Stower, L., Triantafilis, J., 2020. Operationalising digital soil mapping – Lessons from Australia. Geoderma Regional 23, e00335. doi:10.1016/j.geodrs.2020.e00335

Kobayashi, S., Sanga-Ngoie, K., 2008. The integrated radiometric correction of optical remote sensing imageries. International Journal of Remote Sensing 29, 5957–5985. doi:10.1080/01431160701881889

Kowalski, K., Okujeni, A., Brell, M., Hostert, P., 2022. Quantifying drought effects in Central European grasslands through regression-based unmixing of intra-annual Sentinel-2 time series. Remote Sensing of Environment 268, 112781. doi:10.1016/j.rse.2021.112781

Kowalski, K., Okujeni, A., Hostert, P., 2023. A generalized framework for

drought monitoring across Central European grassland gradients with Sentinel-2 time series. Remote Sensing of Environment 286, 113449. doi:10.1016/j.rse.2022.113449

Kowalski, K., Senf, C., Okujeni, A., Hostert, P., 2024. Large-scale remote sensing analysis reveals an increasing coupling of grassland vitality to atmospheric water demand. Global change biology 30, e17315. doi:10.1111/gcb.17315

Kross, A., Kaur, G., Jaeger, J.A.G., 2022. A geospatial framework for the assessment and monitoring of environmental impacts of agriculture. Environmental Impact Assessment Review 97, 106851–106851. doi:10.1016/j.eiar.2022.106851

Laamrani, A., Joosse, P., McNairn, H., Berg, A.A., Hagerman, J., Powell, K., Berry, M., 2020. Assessing Soil Cover Levels during the Non-Growing Season Using Multitemporal Satellite Imagery and Spectral Unmixing Techniques. Remote Sensing 12, 1397. doi:10.3390/rs12091397

Lagacherie, P., Baret, F., Feret, J.-B., Madeira Netto, J., Robbez-Masson, J.M., 2008. Estimation of soil clay and calcium carbonate using laboratory, field and airborne hyperspectral measurements. Remote Sensing of Environment 112, 825–835. doi:10.1016/j.rse.2007.06.014

Lakner, S., 2025. The greening of agricultural policies in Germany. FAO. doi:10.4060/cd4950en

Lal, R., 2003. Global Potential of Soil Carbon Sequestration to Mitigate the Greenhouse Effect. Critical Reviews in Plant Sciences 22, 151–184. doi:10.1080/713610854

Lal, R., 2009. Challenges and opportunities in soil organic matter research. European Journal of Soil Science 60, 158–169. doi:10.1111/j.1365-2389.2008.01114.x

Lal, R., 2016. Soil health and carbon management. Food and Energy Security 5, 212–222. doi:10.1002/fes3.96

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. Journal of Open Source Software. doi:10.21105/joss.01903

LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., Kraljevic, T., Nykodym, T., Aboyoun, P., Kurka, M., Malohlava, M., 2021. H2o: R interface for the 'H2O' scalable machine learning platform.

Lehmann, J., Bossio, D.A., Kögel-Knabner, I., Rillig, M.C., 2020. The concept and future prospects of soil health. Nature Reviews Earth & Environment 1, 544–553. doi:10.1038/s43017-020-0080-8

Lettens, S., Van Orshoven, J., Van Wesemael, B., Muys, B., Perrin, D., 2005. Soil organic carbon changes in landscape units of Belgium between 1960 and 2000 with reference to 1990. Global Change Biology 11, 2128–2140. doi:10.1111/j.1365-2486.2005.001074.x

Lewińska, K.E., Frantz, D., Leser, U., Hostert, P., 2024. Usable observations over Europe: Evaluation of compositing windows for Landsat and Sentinel-2 time series. European Journal of Remote Sensing 57, 2372855. doi:10.1080/22797254.2024.2372855

Lewińska, K.E., Hostert, P., Buchner, J., Bleyhl, B., Radeloff, V.C., 2020. Short-term vegetation loss versus decadal degradation of grasslands in the Caucasus based on Cumulative Endmember Fractions. Remote Sensing of Environment 248, 111969. doi:10.1016/j.rse.2020.111969

LfL (Ed.), 2022. 35 Jahre Boden-Dauerbeobachtung landwirtschaftlich genutzter Flächen in Bayern - Band 1: Methoden, Standorte, Bewirtschaftung, ISSN 1611-4159. Freising-Weihenstephan.

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A., Talwalkar, A., 2017. Hyperband: A novel bandit-based approach to hyperparameter optimization. J. Mach. Learn. Res. 18, 6765–6816.

Li, X., Zheng, G., Wang, J., Ji, C., Sun, B., Gao, Z., 2016. Comparison of Methods for Estimating Fractional Cover of Photosynthetic and Non-Photosynthetic Vegetation in the Otindag Sandy Land Using GF-1 Wide-Field View Data. Remote Sensing 8, 800. doi:10.3390/rs8100800

Lin, L.I.-K., 1989. A Concordance Correlation Coefficient to Evaluate Reproducibility. Biometrics 45, 255. doi:10.2307/2532051

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. doi:10.1145/3236386.3241340

Lloyd, S., 1982. Least squares quantization in PCM. IEEE Transactions on Information Theory 28, 129–137. doi:10.1109/TIT.1982.1056489

Lobert, F., Löw, J., Schwieder, M., Gocht, A., Schlund, M., Hostert, P., Erasmi, S., 2023. A deep learning approach for deriving winter wheat phenology from optical and SAR time series at field level. Remote Sensing of Environment

298, 113800. doi:10.1016/j.rse.2023.113800

Lobert, F., Schwieder, M., Alsleben, J., Broeg, T., Kowalski, K., Okujeni, A., Hostert, P., Erasmi, S., 2025a. Unveiling year-round cropland cover by soil-specific spectral unmixing of Landsat and Sentinel-2 time series. Remote Sensing of Environment 318, 114594. doi:10.1016/j.rse.2024.114594

Lobert, F., Schwieder, M., Hostert, P., Gocht, A., Erasmi, S., 2025b. Characterizing spatio-temporal patterns of winter cropland cover in Germany based on Landsat and Sentinel-2 time series. International Journal of Applied Earth Observation and Geoinformation 142, 104728. doi:10.1016/j.jag.2025.104728

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

Lyons, M.B., Keith, D.A., Phinn, S.R., Mason, T.J., Elith, J., 2018. A comparison of resampling methods for remote sensing classification and accuracy assessment. Remote Sensing of Environment 208, 145–153. doi:10.1016/j.rse.2018.02.026

Ma, T., Brus, D.J., Zhu, A.-X., Zhang, L., Scholten, T., 2020. Comparison of conditioned Latin hypercube and feature space coverage sampling for predicting soil classes using simulation from soil maps. Geoderma 370, 114366. doi:10.1016/j.geoderma.2020.114366

Machado, D.F.T., Silva, S.H.G., Curi, N., de Menezes, M.D., 2019. Soil type spatial prediction from random forest: Different training datasets, transferability, accuracy and uncertainty assessment. Scientia Agricola 76, 243–254. doi:10.1590/1678-992x-2017-0300

Maleki, S., Khormali, F., Mohammadi, J., Bogaert, P., Bagheri Bodaghabadi, M., 2020. Effect of the accuracy of topographic data on improving digital soil mapping predictions with limited soil data: An application to the Iranian loess plateau. CATENA 195. doi:10.1016/j.catena.2020.104810

Malone, B.P., Jha, S.K., Minasny, B., McBratney, A.B., 2016. Comparing regression-based digital soil mapping and multiple-point geostatistics for the spatial extrapolation of soil data. Geoderma 262, 243–253. doi:10.1016/j.geoderma.2015.08.037

Mandl, L., Viana-Soto, A., Seidl, R., Stritih, A., Senf, C., 2024. Unmixing-based forest recovery indicators for predicting long-term recovery success. Remote Sensing of Environment 308, 114194. doi:10.1016/j.rse.2024.114194

Mcbratney, A., Mendonça Santos, M., Minasny, B., 2003. On Digital Soil Map-

ping. Geoderma 117, 3–52. doi:10.1016/S0016-7061(03)00223-4

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

McClelland, S.C., Paustian, K., Schipanski, M.E., 2021. Management of cover crops in temperate climates influences soil organic carbon stocks: A meta-analysis. Ecological Applications 31, 1–19. doi:10.1002/eap.2278

McKay, J., Grunwald, S., Shi, X., Long, R.F., 2010. Evaluation of the transferability of a knowledge-based soil-landscape model, in: Digital Soil Mapping. Springer Netherlands, pp. 165–178. doi:10.1007/978-90-481-8863-5{\textunderscore }14

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 Sensing of Environment 300, 113911. doi:10.1016/j.rse.2023.113911

Meng, X., Bao, Y., Wang, Y., Zhang, X., Liu, H., 2022. An advanced soil organic carbon content prediction model via fused temporal-spatial-spectral (TSS) information based on machine learning and deep learning algorithms. Remote Sensing of Environment 280, 113166. doi:10.1016/j.rse.2022.113166

Meurer, K.H.E., Haddaway, N.R., Bolinder, M.A., Kätterer, T., 2018. Tillage intensity affects total SOC stocks in boreo-temperate regions only in the topsoil—A systematic review using an ESM approach. Earth-Science Reviews 177, 613–622. doi:10.1016/j.earscirev.2017.12.015

Meusburger, K., Konz, N., Schaub, M., Alewell, C., 2010. Soil erosion modelled with USLE and PESERA using QuickBird derived vegetation parameters in an alpine catchment. International Journal of Applied Earth Observation and Geoinformation 12, 208–215. doi:10.1016/j.jag.2010.02.004

Meyer, G.E., Hindman, T.W., Laksmi, K., 1999. Machine vision detection parameters for plant species identification, in: Meyer, G.E., DeShazer, J.A. (Eds.), Photonics East (ISAM, VVDC, IEMB). Boston, MA, pp. 327–335. doi:10.1117/12.336896

Meyer, H., Pebesma, E., 2021. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods in Ecology and Evolution 12, 1620–1633. doi:10.1111/2041-210X.13650

Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., Nauss, T., 2018. Improving performance of spatio-temporal machine learning models using forward feature

selection and target-oriented validation. Environmental Modelling & Software 101, 1–9. doi:10.1016/j.envsoft.2017.12.001

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. Ecological Modelling 411, 108815. doi:10.1016/j.ecolmodel.2019.108815

Miller, L., Pelletier, C., Webb, G.I., 2024. Deep Learning for Satellite Image Time-Series Analysis: A review. IEEE Geoscience and Remote Sensing Magazine 12, 81–124. doi:10.1109/mgrs.2024.3393010

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

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. doi:10.1016/B978-0-12-405942-9.00001-3

Moinet, G.Y.K., Hijbeek, R., Van Vuuren, D.P., Giller, K.E., 2023. Carbon for soils, not soils for carbon. Global Change Biology 29, 2384–2398. doi:10.1111/gcb.16570

Möller, M., Zepp, S., Wiesmeier, M., Gerighausen, H., Heiden, U., 2022. Scale-Specific Prediction of Topsoil Organic Carbon Contents Using Terrain Attributes and SCMaP Soil Reflectance Composites. Remote Sensing 14, 2295. doi:10.3390/rs14102295

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. doi:10.5194/soil-6-269-2020

Montanarella, L., Panagos, P., 2021. The relevance of sustainable soil management within the European Green Deal. Land Use Policy 100, 104950. doi:10.1016/j.landusepol.2020.104950

Montandon, L.M., Small, E.E., 2008. The impact of soil reflectance on the quan-

tification of the green vegetation fraction from NDVI. Remote Sensing of Environment, Remote Sensing Data Assimilation Special Issue 112, 1835–1845. doi:10.1016/j.rse.2007.09.007

Mulder, V.L., De Bruin, S., Schaepman, M.E., Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — A review. Geoderma 162, 1–19. doi:10.1016/j.geoderma.2010.12.018

Murphy, B.W., 2015. Impact of soil organic matter on soil properties—a review with emphasis on Australian soils. Soil Research 53, 605–635. doi:10.1071/SR14246

Mzid, N., Pignatti, S., Huang, W., Casa, R., 2021. An Analysis of Bare Soil Occurrence in Arable Croplands for Remote Sensing Topsoil Applications. Remote Sensing 13, 474. doi:10.3390/rs13030474

Neyestani, M., Sarmadian, F., Jafari, A., Keshavarzi, A., Sharififar, A., 2021. Digital mapping of soil classes using spatial extrapolation with imbalanced data. Geoderma Regional 26. doi:10.1016/j.geodrs.2021.e00422

Ng, W., Minasny, B., Mendes, W.D.S., Demattê, J.A.M., 2020. The influence of training sample size on the accuracy of deep learning models for the prediction of soil properties with near-infrared spectroscopy data. SOIL 6, 565–578. doi:10.5194/soil-6-565-2020

Nguyen, C.T., Chidthaisong, A., Kieu Diem, P., Huo, L.-Z., 2021. A Modified Bare Soil Index to Identify Bare Land Features during Agricultural Fallow-Period in Southeast Asia Using Landsat 8. Land 10, 231. doi:10.3390/land10030231

Niedertscheider, M., Kuemmerle, T., Müller, D., Erb, K.-H., 2014. Exploring the effects of drastic institutional and socio-economic changes on land system dynamics in Germany between 1883 and 2007. Global environmental change : human and policy dimensions 28, 98–108. doi:10.1016/j.gloenvcha.2014.06.006

Nill, L., Grünberg, I., Ullmann, T., Gessner, M., Boike, J., Hostert, P., 2022. Arctic shrub expansion revealed by Landsat-derived multitemporal vegetation cover fractions in the Western Canadian Arctic. Remote Sensing of Environment 281, 113228. doi:10.1016/j.rse.2022.113228

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. doi:10.1016/j.geoderma.2012.07.020

Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M.E., Papritz, A., 2018. Evaluation of digital soil mapping approaches with large sets of environmental covariates. SOIL 4, 1–22. doi:10.5194/soil-4-1-2018

Ogle, S.M., Alsaker, C., Baldock, J., Bernoux, M., Breidt, F.J., McConkey, B., Regina, K., Vazquez-Amabile, G.G., 2019. Climate and Soil Characteristics Determine Where No-Till Management Can Store Carbon in Soils and Mitigate Greenhouse Gas Emissions. Scientific Reports 9, 11665. doi:10.1038/s41598-019-47861-7

Okin, G.S., Roberts, D.A., Murray, B., Okin, W.J., 2001. Practical limits on hyperspectral vegetation discrimination in arid and semiarid environments. Remote Sensing of Environment 77, 212–225. doi:10.1016/S0034-4257(01)00207-3

Okujeni, A., Jänicke, C., Cooper, S., Frantz, D., Hostert, P., Clark, M., Segl, K., van der Linden, S., 2021. Multi-season unmixing of vegetation class fractions across diverse Californian ecoregions using simulated spaceborne imaging spectroscopy data. Remote Sensing of Environment 264, 112558. doi:10.1016/j.rse.2021.112558

Okujeni, A., Kowalski, K., Lewińska, K.E., Schneidereit, S., Hostert, P., 2024. Multidecadal grassland fractional cover time series retrieval for Germany from the Landsat and Sentinel-2 archives. Remote Sensing of Environment 302, 113980. doi:10.1016/j.rse.2023.113980

Okujeni, A., Van Der Linden, S., Suess, S., Hostert, P., 2017. Ensemble Learning From Synthetically Mixed Training Data for Quantifying Urban Land Cover With Support Vector Regression. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10, 1640–1650. doi:10.1109/JSTARS.2016.2634859

Okujeni, A., van der Linden, S., Tits, L., Somers, B., Hostert, P., 2013. Support vector regression and synthetically mixed training data for quantifying urban land cover. Remote Sensing of Environment 137, 184–197. doi:10.1016/j.rse.2013.06.007

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment 148, 42–57. doi:10.1016/j.rse.2014.02.015

Öttl, L.K., Wilken, F., Juřicová, A., Batista, P.V.G., Fiener, P., 2024. A millennium of arable land use – the long-term impact of tillage and water erosion on landscape-scale carbon dynamics. SOIL 10, 281–305. doi:10.5194/soil-10-281-2024

Pacheco, A., McNairn, H., 2010. Evaluating multispectral remote sensing and spectral unmixing analysis for crop residue mapping. Remote Sensing of Environment 114, 2219–2228. doi:10.1016/j.rse.2010.04.024

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

Panagos, P., Borrelli, P., Meusburger, K., Alewell, C., Lugato, E., Montanarella, L., 2015. Estimating the soil erosion cover-management factor at the European scale. Land Use Policy 48, 38–50. doi:10.1016/j.landusepol.2015.05.021

Paul, C., Bartkowski, B., Dönmez, C., Don, A., Mayer, S., Steffens, M., Weigl, S., Wiesmeier, M., Wolf, A., Helming, K., 2023. Carbon farming: Are soil carbon certificates a suitable tool for climate change mitigation? Journal of Environmental Management 330, 117142. doi:10.1016/j.jenvman.2022.117142

Paul, S.S., Dowell, L., Coops, N.C., Johnson, M.S., Krzic, M., Geesing, D., Smukler, S.M., 2020. Tracking changes in soil organic carbon across the heterogeneous agricultural landscape of the Lower Fraser Valley of British Columbia. Science of The Total Environment 732, 138994. doi:10.1016/j.scitotenv.2020.138994

Paustian, K., Collier, S., Baldock, J., Burgess, R., Creque, J., DeLonge, M., Dungait, J., Ellert, B., Frank, S., Goddard, T., Govaerts, B., Grundy, M., Henning, M., Izaurralde, R.C., Madaras, M., McConkey, B., Porzig, E., Rice, C., Searle, R., Seavy, N., Skalsky, R., Mulhern, W., Jahn, M., 2019. Quantifying carbon for agricultural soil management: From the current status toward a global soil information system. Carbon Management 10, 567–587. doi:10.1080/17583004.2019.1633231

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

Pazúr, R., Lieskovský, J., Bürgi, M., Müller, D., Lieskovský, T., Zhang, Z., Prishchepov, A., 2020. Abandonment and recultivation of agricultural lands in slovakia—patterns and determinants from the past to the future. Land 9, 316. doi:10.3390/LAND9090316

Petermann, J.S., Buzhdygan, O.Y., 2021. Grassland biodiversity. Current Biology 31, R1195–R1201. doi:10.1016/j.cub.2021.06.060

Pham, V.-D., Tetteh, G., Thiel, F., Erasmi, S., Schwieder, M., Frantz, D., Van Der Linden, S., 2024. Temporally transferable crop mapping with temporal encoding and deep learning augmentations. International Journal of Applied Earth Observation and Geoinformation 129, 103867. doi:10.1016/j.jag.2024.103867

Poeplau, C., Don, A., 2013. Sensitivity of soil organic carbon stocks and fractions to different land-use changes across Europe. Geoderma 192, 189–201. doi:10.1016/j.geoderma.2012.08.003

Poeplau, C., Don, A., 2015. Carbon sequestration in agricultural soils via cultivation of cover crops – A meta-analysis. Agriculture, Ecosystems & Environment 200, 33–41. doi:10.1016/j.agee.2014.10.024

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

Poeplau, C., Gregorich, E., 2023. Advances in measuring soil organic carbon stocks and dynamics at the profile scale, in: Rumpel, C. (Ed.), Understanding and Fostering Soil Carbon Sequestration, Burleigh Dodds Series in Agricultural Science. Burleigh Dodds Science Publishing, Cambridge, pp. 323–350. doi:10.19103/AS.2022.0106.10

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. Journal of Plant Nutrition and Soil Science 183, 665–681. doi:10.1002/jpln.202000113

Poeplau, C., Prietz, R., Don, A., 2022. Plot-scale variability of organic carbon in temperate agricultural soils—Implications for soil monitoring. Journal of Plant Nutrition and Soil Science 185, 403–416. doi:10.1002/jpln.202100393

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

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, 217–240.

doi:10.5194/soil-7-217-2021

Pouladi, N., Gholizadeh, A., Khosravi, V., Borůvka, L., 2023. Digital mapping of soil organic carbon using remote sensing data: A systematic review. CATENA 232, 107409. doi:10.1016/j.catena.2023.107409

Prechtel, A., Von Lützow, M., Uwe Schneider, B., Bens, O., Bannick, C.G., Kögel-Knabner, I., Hüttl, R.F., 2009. Organic carbon in soils of Germany: *Status Quo* and the need for new data to evaluate potentials and trends of soil carbon sequestration. Journal of Plant Nutrition and Soil Science 172, 601–614. doi:10.1002/jpln.200900034

Prishchepov, A.V., Müller, D., Baumann, M., Kuemmerle, T., Alcantara, C., Radeloff, V.C., 2017. Underlying drivers and spatial determinants of post-Soviet agricultural land abandonment in temperate eastern europe, in: Gutman, G., Radeloff, V. (Eds.), Land-Cover and Land-Use Changes in Eastern Europe After the Collapse of the Soviet Union in 1991. Springer International Publishing, Cham, pp. 91–117. doi:10.1007/978-3-319-42638-9{\textunderscore }5

Prout, J.M., Shepherd, K.D., McGrath, S.P., Kirk, G.J.D., Haefele, S.M., 2021. What is a good level of soil organic matter? An index based on organic carbon to clay ratio. European Journal of Soil Science 72, 2493–2503. doi:10.1111/ejss.13012

Qiu, S., Zhu, Z., Shang, R., Crawford, C.J., 2021. Can Landsat 7 preserve its science capability with a drifting orbit? Science of Remote Sensing 4, 100026. doi:10.1016/j.srs.2021.100026

R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Rasmussen, C., Heckman, K., Wieder, W.R., Keiluweit, M., Lawrence, C.R., Berhe, A.A., Blankinship, J.C., Crow, S.E., Druhan, J.L., Hicks Pries, C.E., Marin-Spiotta, E., Plante, A.F., Schädel, C., Schimel, J.P., Sierra, C.A., Thompson, A., Wagai, R., 2018. Beyond clay: Towards an improved set of variables for predicting soil organic matter content. Biogeochemistry 137, 297–306. doi:10.1007/s10533-018-0424-3

Rast, M., Nieke, J., Adams, J., Isola, C., Gascon, F., 2021. Copernicus Hyperspectral Imaging Mission for the Environment (Chime), in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. IEEE, Brussels, Belgium, pp. 108–111. doi:10.1109/igarss47720.2021.9553319

Rau, K., Eggensperger, K., Schneider, F., Hennig, P., Scholten, T., 2024. How can we quantify, explain, and apply the uncertainty of complex soil maps predicted with neural networks? Science of The Total Environment 944, 173720. doi:10.1016/j.scitotenv.2024.173720

Rawlins, B.G., Marchant, B.P., Smyth, D., Scheib, C., Lark, R.M., Jordan, C., 2009. Airborne radiometric survey data and a DTM as covariates for regional scale mapping of soil organic carbon across Northern Ireland. European Journal of Soil Science 60, 44–54. doi:10.1111/j.1365-2389.2008.01092.x

Rentschler, T., Bartelheim, M., Behrens, T., Díaz-Zorita Bonilla, M., Teuber, S., Scholten, T., Schmidt, K., 2022. Contextual spatial modelling in the horizontal and vertical domains. Scientific Reports 12, 1–11. doi:10.1038/s41598-022-13514-5

Richer-de-Forges, A.C., Chen, Q., Baghdadi, N., Chen, S., Gomez, C., Jacquemoud, S., Martelet, G., Mulder, V.L., Urbina-Salazar, D., Vaudour, E., Weiss, M., Wigneron, J.-P., Arrouays, D., 2023. Remote Sensing Data for Digital Soil Mapping in French Research—A Review. Remote Sensing 15, 3070. doi:10.3390/rs15123070

Richter, N., Jarmer, T., Chabrillat, S., Oyonarte, C., Hostert, P., Kaufmann, H., 2009. Free Iron Oxide Determination in Mediterranean Soils using Diffuse Reflectance Spectroscopy. Soil Science Society of America Journal 73, 72–81. doi:10.2136/sssaj2008.0025

Roberts, D.A., Gardner, M., Church, R., Ustin, S., Scheer, G., Green, R.O., 1998. Mapping Chaparral in the Santa Monica Mountains Using Multiple Endmember Spectral Mixture Models. Remote Sensing of Environment 65, 267–279. doi:10.1016/S0034-4257(98)00037-6

Roberts, D., Wilford, J., Ghattas, O., 2019. Exposed soil and mineral map of the Australian continent revealing the land at its barest. Nature Communications 10, 5297. doi: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. Nature Climate Change 9, 817–828. doi: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 Sensing of Environment 205, 1–17. doi:10.1016/j.rse.2017.11.004

Ross, B.C., 2014. Mutual information between discrete and continuous data sets. PloS one 9, e87357. doi:10.1371/journal.pone.0087357

Roßberg, D., Michel, V., Graf, R., Neukampf, R., 2007. Definition von Boden-Klima-Räumen für die Bundesrepublik Deutschland. Nachrichtenbl. Deut. Pflanzenschutzd. 59, 155–161.

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

Roudier, P., Odgers, N., Carrick, S., Eger, A., Hainsworth, S., Beaudette, D., 2022. Soilscapes of New Zealand: Pedologic diversity as organised along environmental gradients. Geoderma 409, 115637. doi:10.1016/j.geoderma.2021.115637

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 Sensing of Environment 176, 255–271. doi: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 Sensing of Environment 24, 423–446. doi:10.1016/0034-4257(88)90017-X

Rufin, P., Frantz, D., Yan, L., Hostert, P., 2021. Operational Coregistration of the Sentinel-2A/B Image Archive Using Multitemporal Landsat Spectral Averages. IEEE Geoscience and Remote Sensing Letters 18, 712–716. doi:10.1109/LGRS.2020.2982245

Rumpel, C. (Ed.), 2023. Understanding and fostering soil carbon sequestration, Burleigh Dodds Series in Agricultural Science. Burleigh Dodds Science Publishing, Cambridge. doi:10.19103/AS.2022.0106

Rutledge, S., Mudge, P.L., Wallace, D.F., Campbell, D.I., Woodward, S.L., Wall, A.M., Schipper, L.A., 2014. CO2 emissions following cultivation of a temperate permanent pasture. Agriculture, Ecosystems & Environment 184, 21–33. doi:10.1016/j.agee.2013.11.005

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, 1369. doi:10.3390/rs12091369

Sagi, O., Rokach, L., 2018. Ensemble learning: A survey. WIREs Data Mining and Knowledge Discovery 8, e1249. doi:10.1002/widm.1249

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. doi:10.5194/soil-8-587-2022

Schug, F., Pfoch, K.A., Pham, V.-D., Van Der Linden, S., Okujeni, A., Frantz, D., Radeloff, V.C., 2024. Land cover fraction mapping across global biomes with Landsat data, spatially generalized regression models and spectral-temporal metrics. Remote Sensing of Environment 311, 114260. doi:10.1016/j.rse.2024.114260

Schulz, C., Holtgrave, A., Kleinschmit, B., 2021. Large-scale winter catch crop monitoring with Sentinel-2 time series and machine learning – An alternative to on-Site controls? Computers and Electronics in Agriculture 186, 106173. doi:10.1016/j.compag.2021.106173

Schwieder, M., Tetteh, G.O., Blickensdörfer, L., Gocht, A., Erasmi, S., 2024. Agricultural land use (raster) : National-scale crop type maps for Germany from combined time series of Sentinel-1, Sentinel-2 and Landsat data (2017 to 2021). doi:10.5281/zenodo.10640528

Schwieder, M., Wesemeyer, M., Frantz, D., Pfoch, K., Erasmi, S., Pickert, J., Nendel, C., Hostert, P., 2022. Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series. Remote Sensing of Environment 269, 112795. doi:10.1016/j.rse.2021.112795

Seitz, D., Fischer, L.M., Dechow, R., Wiesmeier, M., Don, A., 2023. The potential of cover crops to increase soil organic carbon storage in German croplands. Plant and Soil 488, 157–173. doi:10.1007/s11104-022-05438-w

Sekulić, A., Kilibarda, M., Heuvelink, G.B.M., Nikolić, M., Bajat, B., 2020. Random Forest Spatial Interpolation. Remote Sensing 12, 1687. doi:10.3390/rs12101687

Shabou, M., Mougenot, B., Chabaane, Z.L., Walter, C., Boulet, G., Aissa, N.B., Zribi, M., 2015. Soil Clay Content Mapping Using a Time Series of Landsat TM Data in Semi-Arid Lands. Remote Sensing 7, 6059–6078. doi:10.3390/rs70506059

Sharififar, A., Sarmadian, F., Malone, B.P., Minasny, B., 2019. Addressing the issue of digital mapping of soil classes with imbalanced class observations. Geoderma 350, 84–92. doi:10.1016/j.geoderma.2019.05.016

Shi, P., Six, J., Sila, A., Vanlauwe, B., Van Oost, K., 2022. Towards spatially continuous mapping of soil organic carbon in croplands using multitemporal

Sentinel-2 remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 193, 187–199. doi:10.1016/j.isprsjprs.2022.09.013

Silvero, N.E.Q., Demattê, J.A.M., Amorim, M.T.A., Santos, N.V. dos, Rizzo, R., Safanelli, J.L., Poppiel, R.R., Mendes, W. de S., Bonfatti, B.R., 2021. Soil variability and quantification based on Sentinel-2 and Landsat-8 bare soil images: A comparison. Remote Sensing of Environment 252, 112117. doi:10.1016/j.rse.2020.112117

Smit, H.J., Metzger, M.J., Ewert, F., 2008. Spatial distribution of grassland productivity and land use in Europe. Agricultural Systems 98, 208–219. doi:10.1016/j.agsy.2008.07.004

Smith, P., 2004. How long before a change in soil organic carbon can be detected? Global Change Biology 10, 1878–1883. doi:10.1111/j.1365-2486.2004.00854.x

Smith, P., Falloon, P., Kutsch, W.L., 2010. The role of soils in the Kyoto Protocol, in: Soil Carbon Dynamics: An Integrated Methodology. Cambridge University Press, pp. 245–256. doi:10.1017/CBO9780511711794.014

Smith, P., Martino, D., Cai, Z., Gwary, D., Janzen, H., Kumar, P., McCarl, B., Ogle, S., O'Mara, F., Rice, C., Scholes, B., Sirotenko, O., Howden, M., McAllister, T., Pan, G., Romanenkov, V., Schneider, U., Towprayoon, S., Wattenbach, M., Smith, J., 2008. Greenhouse gas mitigation in agriculture. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 789–813. doi:10.1098/rstb.2007.2184

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

Smith, R.J., 2009. Use and misuse of the reduced major axis for line-fitting. American Journal of Physical Anthropology 140, 476–486. doi:10.1002/ajpa.21090

Smith, W.K., Dannenberg, M.P., Yan, D., Herrmann, S., Barnes, M.L., Barron-Gafford, G.A., Biederman, J.A., Ferrenberg, S., Fox, A.M., Hudson, A., Knowles, J.F., MacBean, N., Moore, D.J.P., Nagler, P.L., Reed, S.C., Rutherford, W.A., Scott, R.L., Wang, X., Yang, J., 2019. Remote sensing of dryland

ecosystem structure and function: Progress, challenges, and opportunities. Remote Sensing of Environment 233, 111401. doi:10.1016/j.rse.2019.111401

Sokolova, M., Lapalme, G., 2009. A systematic analysis of performance measures for classification tasks. Information Processing and Management 45, 427–437. doi:10.1016/j.ipm.2009.03.002

Song, Y., Shen, Z., Wu, P., Viscarra Rossel, R.A., 2021. Wavelet geographically weighted regression for spectroscopic modelling of soil properties. Scientific Reports 11, 17503. doi:10.1038/s41598-021-96772-z

Sørensen, L.K., Dalsgaard, S., 2005. Determination of Clay and Other Soil Properties by Near Infrared Spectroscopy. Soil Science Society of America Journal 69, 159. doi:10.2136/sssaj2005.0159

Sorenson, P.T., Shirtliffe, S.J., Bedard-Haughn, A.K., 2021. Predictive soil mapping using historic bare soil composite imagery and legacy soil survey data. Geoderma 401, 115316. doi:10.1016/j.geoderma.2021.115316

Stehman, S.V., 2014. Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. International Journal of Remote Sensing 35, 4923–4939. doi:10.1080/01431161.2014.930207

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, Diffuse reflectance spectroscopy in soil science and land resource assessment 158, 15–22. doi: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. doi:10.1016/S0065-2113(10)07005-7

Stevens, A., Van Wesemael, B., Bartholomeus, H., Rosillon, D., Tychon, B., Ben-Dor, E., 2008. Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils. Geoderma 144, 395–404. doi:10.1016/j.geoderma.2007.12.009

Stumpf, F., Behrens, T., Schmidt, K., Keller, A., 2024. Exploiting Soil and Remote Sensing Data Archives for 3D Mapping of Multiple Soil Properties at the Swiss National Scale. Remote Sensing 16, 2712. doi:10.3390/rs16152712

Stumpf, F., Schmidt, K., Behrens, T., Schönbrodt–Stitt, S., Buzzo, G., Dumperth, C., Wadoux, A., Xiang, W., Scholten, T., 2016. Incorporating limited field operability and legacy soil samples in a hypercube sampling

design for digital soil mapping. Journal of Plant Nutrition and Soil Science 179, 499–509. doi:10.1002/jpln.201500313

Sulik, J.J., Long, D.S., 2016. Spectral considerations for modeling yield of canola. Remote Sensing of Environment 184, 161–174. doi:10.1016/j.rse.2016.06.016

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. doi:10.1016/j.geoderma.2021.115108

Taghizadeh-Mehrjardi, R., Khademi, H., Khayamim, F., Zeraatpisheh, M., Heung, B., Scholten, T., 2022a. A Comparison of Model Averaging Techniques to Predict the Spatial Distribution of Soil Properties. Remote Sensing 14, 472. doi:10.3390/rs14030472

Taghizadeh-Mehrjardi, R., Schmidt, K., Amirian-Chakan, A., Rentschler, T., Zeraatpisheh, M., Sarmadian, F., Valavi, R., Davatgar, N., Behrens, T., Scholten, T., 2020. Improving the spatial prediction of soil organic carbon content in two contrasting climatic regions by stacking machine learning models and rescanning covariate space. Remote Sensing 12. doi:10.3390/rs12071095

Taghizadeh-Mehrjardi, R., Sheikhpour, R., Zeraatpisheh, M., Amirian-Chakan, A., Toomanian, N., Kerry, R., Scholten, T., 2022b. Semi-supervised learning for the spatial extrapolation of soil information. Geoderma 426, 116094. doi:10.1016/j.geoderma.2022.116094

Tarasiou, M., Chavez, E., Zafeiriou, S., 2023. Vits for sits: Vision transformers for satellite image time series, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10418–10428.

Tetteh, G.O., Schwieder, M., Erasmi, S., Conrad, C., Gocht, A., 2023. Comparison of an Optimised Multiresolution Segmentation Approach with Deep Neural Networks for Delineating Agricultural Fields from Sentinel-2 Images. PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science 91, 295–312. doi:10.1007/s41064-023-00247-x

Thünen-Institut, 2022. Forstliche Großlandschaften (2011).

Tian, X., Simoes, R., Isik, M.S., Minarik, R., Ho, Y., Hengl, T., 2024. Spatiotemporal prediction of soil organic carbon density (SOCD) for pan-Europe (2000-2022) in 3D+T. doi:10.5281/ZENODO.13754344

Tkachenko, M., Malyuk, M., Holmanyuk, A., Liubimov, N., 2024. Label Studio: Data labeling software: Open source software available from

https://github.com/HumanSignal/label-studio.

Tóth, K., Kučas, A., 2016. Spatial information in European agricultural data management. Requirements and interoperability supported by a domain model. Land Use Policy 57, 64–79. doi:10.1016/j.landusepol.2016.05.023

Trangmar, B.B., Yost, R.S., Uehara, G., 1986. Application of Geostatistics to Spatial Studies of Soil Properties, in: Advances in Agronomy. Elsevier, pp. 45–94. doi:10.1016/S0065-2113(08)60673-2

Tscharntke, T., Clough, Y., Wanger, T.C., Jackson, L., Motzke, I., Perfecto, I., Vandermeer, J., Whitbread, A., 2012. Global food security, biodiversity conservation and the future of agricultural intensification. Biological Conservation, ADVANCING ENVIRONMENTAL CONSERVATION: ESSAYS IN HONOR OF NAVJOT SODHI 151, 53–59. doi:10.1016/j.biocon.2012.01.068

Tseng, G., Cartuyvels, R., Zvonkov, I., Purohit, M., Rolnick, D., Kerner, H., 2023. Lightweight, Pre-trained Transformers for Remote Sensing Timeseries. doi:10.48550/ARXIV.2304.14065

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sensing of Environment 8, 127–150. doi:10.1016/0034-4257(79)90013-0

Tziolas, N., Tsakiridis, N., Chabrillat, S., Demattê, J.A.M., Ben-Dor, E., Gholizadeh, A., Zalidis, G., van Wesemael, B., 2021. Earth Observation Data-Driven Cropland Soil Monitoring: A Review. Remote Sensing 13, 4439. doi:10.3390/rs13214439

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. doi:10.1023/A:1023008322682

Umweltbundesamt, 2020. Bodendaten: Informationen zur bodenqualität in deutschland.

Umweltbundesamt (Ed.), 2023. Projektionsbericht 2023 für Deutschland. doi:10.60810/OPENUMWELT-2658

Umweltbundesamt, 2024. Indicator: Grasslands.

Umweltbundesamt, 2025. Nationale Trendtabellen für die deutsche Berichterstattung atmosphärischer Emissionen 1990–2023 (Technischer {{Bericht}}). Umweltbundesamt.

UNFCCC, 2015. Paris agreement.

Urbina-Salazar, D., Vaudour, E., Baghdadi, N., Ceschia, E., Richer-de-Forges, A.C., Lehmann, S., Arrouays, D., 2021. Using Sentinel-2 Images for Soil Organic Carbon Content Mapping in Croplands of Southwestern France. The Usefulness of Sentinel-1/2 Derived Moisture Maps and Mismatches between Sentinel Images and Sampling Dates. Remote Sensing 13, 5115. doi:10.3390/rs13245115

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 Sensing 15, 2410. doi:10.3390/rs15092410

Van der Linden, S., Rabe, A., Held, M., Jakimow, B., Leitão, P.J., Okujeni, A., Schwieder, M., Suess, S., Hostert, P., 2015. The EnMAP-Box—A Toolbox and Application Programming Interface for EnMAP Data Processing. Remote Sensing 7, 11249–11266. doi:10.3390/rs70911249

Van der Voort, T.S., Verweij, S., Fujita, Y., Ros, G.H., 2023. Enabling soil carbon farming: Presentation of a robust, affordable, and scalable method for soil carbon stock assessment. Agronomy for Sustainable Development 43, 22. doi:10.1007/s13593-022-00856-7

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. Photogrammetric Engineering and Remote Sensing 63, 87–93.

Van Wesemael, B., Abdelbaki, A., Ben-Dor, E., Chabrillat, S., d'Angelo, P., Demattê, J.A.M., Genova, G., Gholizadeh, A., Heiden, U., Karlshoefer, P., Milewski, R., Poggio, L., Sabetizade, M., Sanz, A., Schwind, P., Tsakiridis, N., Tziolas, N., Yagüe, J., Žížala, D., 2024. A European soil organic carbon monitoring system leveraging Sentinel 2 imagery and the LUCAS soil data base. Geoderma 452, 117113. doi:10.1016/j.geoderma.2024.117113

van Wesemael, B., Chabrillat, S., Wilken, F., 2021. High-spectral resolution remote sensing of soil organic carbon dynamics. Remote sensing. doi:10.3390/rs13071293

Vaserstein, L.N., 1969. Markov processes over denumerable products of spaces, describing large systems of automata. Problemy Peredachi Informatsii 64–72.

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, 2917. doi:10.3390/rs14122917

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. International Journal of Applied Earth Observation and Geoinformation 96, 102277. doi:10.1016/j.jag.2020.102277

Venter, Z.S., Hawkins, H.-J., Cramer, M.D., Mills, A.J., 2021. Mapping soil organic carbon stocks and trends with satellite-driven high resolution maps over South Africa. Science of The Total Environment 771, 145384. doi:10.1016/j.scitotenv.2021.145384

Verrelst, J., Halabuk, A., Atzberger, C., Hank, T., Steinhauser, S., Berger, K., 2023. A comprehensive survey on quantifying non-photosynthetic vegetation cover and biomass from imaging spectroscopy. Ecological Indicators 155, 110911. doi:10.1016/j.ecolind.2023.110911

Vohland, M., Ludwig, B., Seidel, M., Hutengs, C., 2022. Quantification of soil organic carbon at regional scale: Benefits of fusing vis-NIR and MIR diffuse reflectance data are greater for in situ than for laboratory-based modelling approaches. Geoderma 405, 115426. doi:10.1016/j.geoderma.2021.115426

Vrieling, A., 2006. Satellite remote sensing for water erosion assessment: A review. CATENA 65, 2–18. doi:10.1016/j.catena.2005.10.005

Wadoux, A.M.J.-C., Minasny, B., McBratney, A.B., 2020a. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. Earth-Science Reviews 210, 103359. doi: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. doi:10.1016/j.geoderma.2022.115953

Wadoux, A.M.J.-C., Samuel-Rosa, A., Poggio, L., Mulder, V.L., 2020b. A note on knowledge discovery and machine learning in digital soil mapping. European Journal of Soil Science 71, 133–136. doi:10.1111/ejss.12909

Wageningen Environmental Research, 2024. Carbon Removals Expert Group Technical Assistance: Draft for the Agricultural Land Technical Assessment Paper (Technical {{Report}}). European Commission, Directorate-General for Climate Action.

Wagg, C., Roscher, C., Weigelt, A., Vogel, A., Ebeling, A., de Luca, E., Roeder,

A., Kleinspehn, C., Temperton, V.M., Meyer, S.T., Scherer-Lorenzen, M., Buchmann, N., Fischer, M., Weisser, W.W., Eisenhauer, N., Schmid, B., 2022. Biodiversity-stability relationships strengthen over time in a long-term grassland experiment. Nature communications 13, 7752. doi:10.1038/s41467-022-35189-2

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 Research 39, 259–272. doi:10.1071/sr99114

Wang, F., Huang, J., Tang, Y., Wang, X., 2007. New Vegetation Index and Its Application in Estimating Leaf Area Index of Rice. Rice Science 14, 195–203. doi:10.1016/S1672-6308(07)60027-4

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 Sensing 12, 3451. doi:10.3390/rs12203451

Webb, N.P., Marshall, N.A., Stringer, L.C., Reed, M.S., Chappell, A., Herrick, J.E., 2017. Land degradation and climate change: Building climate resilience in agriculture. Frontiers in Ecology and the Environment 15, 450–459. doi:10.1002/fee.1530

Webster, R., Oliver, M.A., 1992. Sample adequately to estimate variograms of soil properties. Journal of Soil Science 43, 177–192. doi:10.1111/j.1365-2389.1992.tb00128.x

Wetterlind, J., Simmler, M., Castaldi, F., Borůvka, L., Gabriel, J.L., Gomes, L.C., Khosravi, V., Kıvrak, C., Koparan, M.H., Lázaro-López, A., Łopatka, A., Liebisch, F., Rodriguez, J.A., Savaş, A.Ö., Stenberg, B., Tunçay, T., Vinci, I., Volungevičius, J., Žydelis, R., Vaudour, E., 2025. Influence of Soil Texture on the Estimation of Soil Organic Carbon From Sentinel-2 Temporal Mosaics at 34 European Sites. European Journal of Soil Science 76, e70054. doi:10.1111/ejss.70054

Wiesmeier, M., Burmeister, J., Garcia-Franco, N., Sümmerer, M., 2024. Soil Organic Carbon and Nitrogen Changes in Agricultural Soils of Bavaria between 1986 and 2016 as Driven by Management and Climate Change. doi:10.2139/ssrn.4901937

Wiesmeier, M., Hübner, R., Barthold, F., Spörlein, P., Geuß, U., Hangen, E.,

Reischl, A., Schilling, B., 2013. Amount, distribution and driving factors of soil organic carbon and nitrogen in cropland and grassland soils of southeast Germany (Bavaria).

Wiesmeier, M., Poeplau, C., Sierra, C.A., Maier, H., Frühauf, C., Hübner, R., Kühnel, A., Spörlein, P., Geuß, U., Hangen, E., Schilling, B., von Lützow, M., Kögel-Knabner, I., 2016. Projected loss of soil organic carbon in temperate agricultural soils in the 21st century: Effects of climate change and carbon input trends. Scientific Reports 6, 32525. doi:10.1038/srep32525

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

Wolski, M.S., Dalmolin, R.S.D., Flores, C.A., Moura-Bueno, J.M., ten Caten, A., Kaiser, D.R., 2017. Digital soil mapping and its implications in the extrapolation of soil-landscape relationships in detailed scale. Pesquisa Agropecuaria Brasileira 52, 633–642. doi:10.1590/S0100-204X2017000800009

Wright, M.N., 2024. Feature selection, in: Bischl, B., Sonabend, R., Kotthoff, L., Lang, M. (Eds.), Applied Machine Learning Using Mlr3 in R. CRC Press.

Wright, M.N., Ziegler, A., 2017. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software 77, 1–17. doi:10.18637/jss.v077.i01

Yang, R.-M., Huang, L.-M., Zhang, X., Zhu, C.-M., Xu, L., 2023. Mapping the distribution, trends, and drivers of soil organic carbon in China from 1982 to 2019. Geoderma 429, 116232. doi:10.1016/j.geoderma.2022.116232

Yue, J., Qingjiu Tian, Tian, Q., Xinyu Dong, Dong, X., Nianxu Xu, Xu, N., 2020. Using broadband crop residue angle index to estimate the fractional cover of vegetation, crop residue, and bare soil in cropland systems. Remote Sensing of Environment 237, 111538. doi:10.1016/j.rse.2019.111538

Yue, J., Tian, Q., Tang, S., Kaijian Xu, Xu, K., Kaijian Xu, Chengquan Zhou, Zhou, C., 2019. A dynamic soil endmember spectrum selection approach for soil and crop residue linear spectral unmixing analysis. International Journal of Applied Earth Observation and Geoinformation 78, 306–317. doi:10.1016/j.jag.2019.02.001

Zepp, S., Heiden, U., Bachmann, M., Möller, M., Wiesmeier, M., Van Wesemael,

B., 2023. Optimized bare soil compositing for soil organic carbon prediction of topsoil croplands in Bavaria using Landsat. ISPRS Journal of Photogrammetry and Remote Sensing 202, 287–302. doi:10.1016/j.isprsjprs.2023.06.003

Zepp, S., Heiden, U., Bachmann, M., Wiesmeier, M., Steininger, M., van Wesemael, B., 2021a. Estimation of Soil Organic Carbon Contents in Croplands of Bavaria from SCMaP Soil Reflectance Composites. Remote Sensing 13, 3141. doi:10.3390/rs13163141

Zepp, S., Jilge, M., Metz-Marconcini, A., Heiden, U., 2021b. The influence of vegetation index thresholding on EO-based assessments of exposed soil masks in Germany between 1984 and 2019. ISPRS Journal of Photogrammetry and Remote Sensing 178, 366–381. doi:10.1016/j.isprsjprs.2021.06.015

Zeraatpisheh, M., Garosi, Y., Reza Owliaie, H., Ayoubi, S., Taghizadeh-Mehrjardi, R., Scholten, T., Xu, M., 2022. Improving the spatial prediction of soil organic carbon using environmental covariates selection: A comparison of a group of environmental covariates. CATENA 208, 105723. doi:10.1016/j.catena.2021.105723

Zhao, X., Xiong, Z., Karlshöfer, P., Tziolas, N., Wiesmeier, M., Heiden, U., Zhu, X.X., 2025. Soil organic carbon estimation using spaceborne hyperspectral composites on a large scale. International Journal of Applied Earth Observation and Geoinformation 140, 104504. doi:10.1016/j.jag.2025.104504

Zhao, Z., Yang, Q., Sun, D., Ding, X., Meng, F.-R., 2020. Extended model prediction of high-resolution soil organic matter over a large area using limited number of field samples. Computers and Electronics in Agriculture 169, 105172. doi:10.1016/j.compag.2019.105172

Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4-7, 8, and Sentinel 2 images. Remote Sensing of Environment 159, 269–277. doi:10.1016/j.rse.2014.12.014

Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. Remote Sensing of Environment 118, 83–94. doi:10.1016/j.rse.2011.10.028

Zhu, Z., Woodcock, C.E., 2014. Continuous change detection and classification of land cover using all available Landsat data. Remote Sensing of Environment 144, 152–171. doi:10.1016/j.rse.2014.01.011

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. doi:10.1016/j.catena.2022.106024

Žížala, D., Minařík, R., Zádorová, T., 2019. Soil Organic Carbon Mapping Using Multispectral Remote Sensing Data: Prediction Ability of Data with Different Spatial and Spectral Resolutions. Remote Sensing 11, 2947. doi:10.3390/rs11242947

Declaration of AI-Assisted Technologies in the Writing Process

During the preparation of this dissertation, the author used Grammarly and ChatGPT in order to improve the language and readability of the text. After using these tools, the author reviewed and edited the content as needed and takes full responsibility for the content of the work.