FROM SPACE TO SOIL:

MODELLING AND MAPPING SOIL ORGANIC CARBON DYNAMICS USING MACHINE LEARNING AND EARTH OBSERVATION

XUEMENG TIAN

Propositions

1. Soil data quality is largely ignored in digital soil mapping. (this thesis)

2. Open maps are open in form, but not in interpretation. (this thesis)

3. Results challenging hypotheses are as important to publish as results supporting them.

4. Scientific rigor erodes when research is treated as a product.

5. Environmental sustainability cannot be achieved in a socially unequal world.

6. Repeated feminist self-labelling reflects weak engagement with feminist theory.

Propositions belonging to the thesis, entitled

From Space to Soil: modelling and mapping soil organic carbon dynamics using machine learning and earth observation

Xuemeng Tian
Wageningen, 29 June 2026

From Space to Soil: modelling and mapping soil organic carbon dynamics using machine learning and earth observation

Xuemeng Tian

Thesis committee

Promotor:
Prof. Dr S. de Bruin
Professor of Geo-information Science and Remote Sensing
Wageningen University & Research

Prof. Dr K.M. de Beurs
Professor of Geo-information Science and Remote Sensing
Wageningen University & Research

Co-promotors:
Prof. Dr M. Herold
Professor of Remote Sensing
University of Potsdam, Germany

Advisors:
Dr B. Ahrens, Max Planck Institute for Biogeochemistry, Jena, Germany
Dr F. Schneider, Johann Heinrich von Thünen Institute, Braunschweig, Germany

Other members:
Prof. Dr G.B.M. Heuvelink, Wageningen University & Research
Prof. Dr H. Meyer, University of Münster, Germany
Prof. M. Wiesmeier, Technical University of Munich, Germany
Dr Y. Zeng, University of Twente

This research was conducted under the auspices of the C.T. de Wit Graduate School of Production Ecology & Resource Conservation (PE&RC)

From Space to Soil: modelling and mapping soil organic carbon dynamics using machine learning and earth observation

Xuemeng Tian

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on 29 June 2026
at 3:30 p.m. in the Omnia Auditorium.

Xuemeng Tian
From Space to Soil: modelling and mapping soil organic carbon dynamics using machine learning and earth observation
260 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2026)
With references, with summary in English

DOI: https://doi.org/10.18174/682189

Summary

Digital soil mapping (DSM) refers to the quantitative prediction of soil types or soil properties in digital form and at defined spatial and temporal resolutions. Over the past decade, DSM applications targeting Soil Organic Carbon (SOC) have expanded rapidly. This growth reflects two converging developments. First, SOC has gained prominence as a key indicator of soil functioning and a leverage point for climate change mitigation, creating a strong demand for spatially explicit and temporally dynamic information on its distribution. Second, DSM of SOC has become increasingly feasible due to the increasing collection of SOC measurements, a rising commitment to open data, expanding Earth Observation (EO) archives, and the widespread adoption of machine learning techniques for environmental modelling. Together, these factors have transformed SOC mapping from a research-driven activity into an emerging operational capability, able to provide spatially explicit and temporally dynamic information at continental scales.

Despite these developments, producing pan-European SOC maps that resolve three spatial dimensions and time (3D+T) remains a gap. Progress is constrained by the lack of a continent-wide harmonized SOC database, the limited availability of consistent, ready-to-use, and high-resolution EO data cubes tailored for environmental modelling, and the considerable computational resources required for large-scale spatiotemporal modelling. The overarching goal of this thesis is therefore to address these constraints by developing an operational, wall-to-wall SOC DSM pipeline for Europe at 30 m resolution, spanning multiple soil depths and time periods. In doing so, the thesis aims not only to deliver new SOC datasets, but also to identify the methodological and practical challenges encountered and explore viable solutions that support the future growth of 3D+T DSM.

Chapter 1 is an introduction to the thesis, presenting the motivation for monitoring SOC and the developments in DSM that enable large-scale spatiotemporal SOC modelling. It reviews existing large-scale SOC mapping products and highlights opportunities to extend them toward more spatially and temporally detailed representations. It concludes with an overview of how the subsequent chapters contribute to the main thread of the work: the key efforts undertaken to achieve pan-European SOC mapping at higher spatial and temporal resolution, the challenges identified along the way, and the exploration of a strategy to address them.

Chapter 2 lays the groundwork by constructing an analysis-ready and cloud-optimized EO data cube for Europe, derived from the Landsat ARD V2 archive (2000–2022). Although Europe is richly monitored from space, its EO resources are dispersed across sensors, formats, and preprocessing standards, and often contain substantial gaps due to poor-quality pixels. This fragmentation hampers their use in environmental modelling, which relies on coherent and temporally consistent data. This cube addresses these limitations by reorganizing more than two decades of Landsat observations into a harmonized, gap-reduced, temporally consistent, and thematically rich feature space. It provides 30 m predictors at bimonthly, annual, and long-term scales, capturing vegetation dynamics, soil exposure, hydrological conditions, and tillage signals. The chapter details the reconstruction of time series affected by low-quality pixels, along with a comprehensive quality assessment of the resulting long-term record, including evaluating gap-filling accuracy using artificially introduced gaps, conducting plausibility checks of spectral index proxies for land-surface processes at the European scale, visual inspection, and predictive tests using SOC regression and land cover classification. As the foundation for all subsequent chapters, the cube supplies the spatial and temporal context needed for meaningful and comparable SOC predictions across regions, depths, and years at high spatial resolution. Its open release further supports transparent, reproducible, and scalable DSM workflows.

Chapter 3 presents the measurements-to-maps spatiotemporal modelling framework developed to produce 3D+T high-resolution SOC density time series across Europe. Using harmonized SOC observations from across the continent together with the Landsat-based feature cube and additional EO-derived predictors, the chapter details how Random Forest (RF) and Quantile Regression Forests (QRF) were applied to generate biannual 30 m SOC density maps for multiple depth intervals down to 2 m for the period 2000–2022. In addition to predictions, the framework provides per-pixel uncertainty estimates and extrapolation risk layers to support transparent communication of map reliability and limitations to users. Independent validation shows good overall performance, with accuracy varying by land cover type, depth interval, and year. The resulting maps offer detailed spatial and temporal patterns of SOC across Europe, reflecting known soil–landscape relationships and long-term environmental gradients. Model driver analysis highlights soil depth as the dominant predictor, followed by vegetation indices, long-term bioclimate, and topography. Although pixel-level uncertainties remain substantial, particularly at greater depths, spatial aggregation reduces them. Together, these products constitute the first operational, temporally explicit SOC baseline for Europe at 30 m resolution.

Chapter 4 examines one of the major challenges in large-scale SOC mapping: the detectability of SOC change given the substantial uncertainties inherent in current modelling approaches. Detecting temporal SOC dynamics is inherently difficult because SOC changes slowly over time and exhibits high spatial variability. To quantify this challenge, the chapter introduces a model-based signal-to-noise ratio (SNR) framework, in which detectability is defined as the ratio between predicted SOC change and its associated uncertainty.

Applied across Europe using RF and QRF models and evaluated with repeated SOC observations, the framework compares two modelling strategies—*change first* and *state first*. Results show that SOC change is generally undetectable at the pixel scale, with SNR values typically below one. This outcome reflects the combined limitations of current DSM–ML–EO workflows: restriced SOC measurements quality and temporal length, EO-related uncertainties, and simplified model assumptions. Nonetheless, spatial aggregation improves SNR, enabling more reliable assessments of SOC change at regional and national scales. The chapter argues that routine reporting of SNR diagnostics is essential to increase transparency in future SOC dynamics modelling and mapping efforts.

Chapter 5 explores whether hybrid, soil-science–informed models can improve the reconstruction of SOC density ($g/cm3$) when bulk density (BD, $g/cm3$) measurements are sparse compared to the availability of SOC concentration (g/kg) data, even though both are required to compute SOC density. Conventional DSM–ML approaches typically model SOC density in a univariate manner, which not only under-utilises available information but may also overlook key physical relationships between SOC concentration and BD. Leveraging the abundant SOC concentration data from multiple Land Use/Cover Area Frame Statistical Survey (LUCAS) survey years, this chapter compares three neural network architectures: a univariate network, a multivariate network, and a hybrid soil-informed network that embeds established SOC–BD relationships into its latent structure. The models are evaluated on their ability to predict SOC density, reproduce the joint SOC–BD distribution, and reconstruct reliable SOC density time series. Results show that, compared to the univariate model, the multivariate model and especially the hybrid soil-informed model produces more physically consistent SOC density trajectories, mitigating the noisy or implausible patterns often observed in univariate predictions. Additional plausibility checks on the soil-informed model's latent BD components indicate that embedding domain knowledge can steer the model toward more soil-physically meaningful representations. This chapter demonstrates the potential of hybrid, soil-science–informed neural networks to enhance SOC density reconstruction in data-limited DSM contexts.

Chapter 6 reviews the gaps addressed, the challenges encountered, and the potential solutions identified. It also reflects on the broader lessons learned and outlines key questions for future research. Together, these elements position the thesis in continuity with prior developments while contributing to future advances.

These contributions were developed at a time when DSM of SOC is shifting from producing detailed, large-scale 3D+T SOC maps toward addressing broader questions of their interpretation, usability, and improvement. At the outset, generating consistent, high-resolution SOC maps with quantified uncertainty across Europe represented both a research and operational gap, as well as a practical need for soil-health monitoring, carbon accounting, and land management. The increasing availability of relevant studies and products, however, raises further questions: the practical limits of model accuracy

in data-poor settings; transparent and appropriate communication of uncertainty; the relevance of continental-scale products for local decision-making; and the methodological considerations that emerge once technical limitations are no longer the primary constraint. In several cases, this thesis reinforces known challenges and introduces new ones, rather than providing definitive solutions. By systematically documenting advances and limitations in data preparation, modelling, uncertainty quantification, SOC-change detectability, and integration with soil science, this work seeks to clarify the opportunities and constraints of current DSM–ML–EO approaches. In doing so, it aims to offer stepping stones for future researchers in developing more reliable, scalable, and transparent SOC monitoring frameworks, and to support the broader community in fostering open, honest, and constructive discussions.

Contents

Page

Summary v

Contents ix

Chapter 1 Introduction 1

Chapter 2 Preparation of high-resolution Earth Observation time-series for dynamic Soil Organic Carbon modelling 15

Chapter 3 High-resolution 3D+T mapping of Soil Organic Carbon density 67

Chapter 4 Modeling soil organic carbon changes using signal-to-noise analysis: a case study using European soil survey datasets 117

Chapter 5 Soil science-informed neural networks for soil organic carbon density modelling under scarce bulk density data 153

Chapter 6 Synthesis 175

Acronyms 195

References 197

Acknowledgements 237

About the author 241

Overview of completed training activities 243

Authorship statement 247

Chapter 1

Introduction

1.1 Background

1.1.1 From early soil maps to digital soil mapping

A soil map is, at its core, a representation of how soil attributes vary across space (Yaalon, 1989). Soil maps, in their many early forms, have a long and diverse history. In ancient societies, attempts to record differences in soil and land arose naturally from lived experience. In the Aztec world, a soil classification unique to their society was developed and used to characterize the soils of household lands (Hartemink et al., 2013). These early "soil map" supported agriculture, taxation, medicinal practices, and construction. In ancient China, an agricultural civilization in which soil held such significance that it became embedded in cultural ideology. Early accounts describe how Yu, a hydrological engineer who is generally dated to the *c.* 21st–19th centuries BCE, surveyed the world as it was known at the time during large-scale flood-control efforts. Based on these observations, he divided the territory into nine regions according to terrain and soil characteristics. If interpreted broadly, this can be seen as one of the earliest forms of soil classification mapping (Gong et al., 2003) (Fig 1.1). Yu later became the first ruler of China's earliest recorded dynasty, and this soil-based division profoundly shaped how later generations conceptualised the landscape. These early "map" were not cartographic products in the modern sense, but evolving repositories of empirical knowledge. Whether expressed as household land records or as large-scale territorial divisions with ideological meaning, both reflect a persistent need to organize spatial soil information.

With the emergence of soil science as a discipline, soil maps increasingly reflected scientific curiosity rather than practical needs. A major turning point came with Vasily Dokuchaev (1846–1903), the first to assert that soil is an independent natural body (Rusakova et al., 2022). He formulated the law of soil-forming factors and developed the theory of natural and soil zones (Dokuchaev, 1883). This work laid the foundation of soil science, and his ideas were subsequently adopted, expanded, and refined by later soil scientists (Jenny, 1941). Building on this theoretical framework and extensive surveys across Russia, Dokuchaev produced an "isohumus" map depicting humus content from quantitative soil measurements. His student, K. D. Glinka, further advanced soil–geographical concepts and, in 1908, prepared the first schematic world soil map, comprising 18 soil types (Hartemink et al., 2013).

However, as stated by Legros (2006), regarding this first world soil map, "Its accuracy is very limited and the map must not be considered in details." Despite Dokuchaev's explicit recognition of soil-forming factors and his reliance on quantitative surveys, the resulting system still oversimplified soil variation. Its zonal character, applied both to regions and soil properties, led to polygon-based maps lacking fine spatial detail and reducing soils to broad, weakly informative classes. This mapping approach was already criticized at the time by G. F. Nefedov, who argued that a single map summarizing soil taxa has little

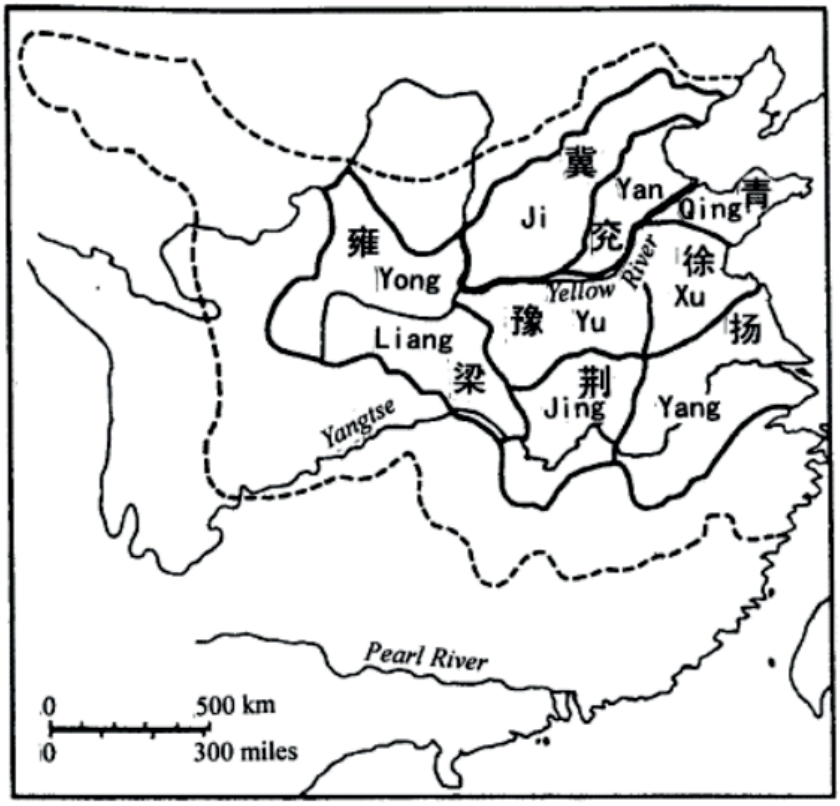


Figure 1.1: Soil map of the Nine Provinces as described in the book Yu Gong, traditionally attributed to Yu based on his flood control experiences. Adapted from Gong et al. (2003).

practical value and that soil maps should instead comprise a complex of individual maps showing the geographical distribution of each soil feature and property (Hartemink et al., 2013). His vision was far ahead of its time: the data resources required to implement such an approach simply did not exist.

Despite advances in soil science, soil mapping throughout much of the 20 th century continued to rely on the expertise and conceptual models of surveyors and mappers (Hudson, 1992; Minasny and McBratney, 2016). Considerable work was undertaken to investigate quantitative relationships among soil-forming factors (Yaalon, 1975), primarily to understand soil formation rather than to predict soil properties spatially. Yet, these findings proved difficult to translate into practical soil mapping. Soil is a complex system influenced by numerous interacting factors, and the inferred relationships often turned out to be highly landscape- and scale-dependent, with few generally valid equations. Moreover, many identified relationships were nonlinear, adding further complications to

their application in mapping. As a result, in the practice of soil mapping, the links between soil and environmental features, as well as the processes connecting them, remained largely qualitative and interpretive. It's argued that the knowledge contained within the conceptual model of the soil surveyor is more important than the maps, which is a poor realisation of that conceptual model (Hudson, 1992; Ryan et al., 2000). Consequently, soil maps produced by this conventional approach were still dominated by soil type maps (Hartemink et al., 2013).

It was not until the late 20 th century that Nefedov's vision began to come true. This was largely due to the rapid expansion of computing and information technology, which fundamentally transformed the way environmental and soil data could be created, organized, processed, and accessed (McBratney et al., 2003; McKenzie and Ryan, 1999). This technological shift, especially the application of geostatistics (Heuvelink and Webster, 2022), enabled the development and application of quantitative tools capable of quantifying soil–environment relationships and applying these relationships in soil mapping. With increasing availability of soil measurements, this led to the establishment of a new scientific discipline: Digital Soil Mapping (DSM). According to Lagacherie and McBratney (2006), DSM comprises three main components: inputs, which include field and laboratory observations such as legacy soil data, existing soil maps, and newly collected samples obtained through statistical sampling; processes, referring to spatial and non-spatial soil inference systems in which mathematical or statistical models relate soil observations to environmental covariates; and outputs, consisting of spatial soil information systems, usually raster predictions accompanied by uncertainty estimates, that can be updated as new information becomes available. Like everything else, soil mapping moved into the information age in the 21 st century.

The first Global Workshop on DSM was held in Montpellier, France, in 2004, marking a key milestone in promoting DSM as a means to meet the growing demand for up-to-date and updatable soil information. In 2006, discussions at the second Global Workshop on DSM led to the proposal of generating global gridded maps of key soil functional properties, which subsequently evolved into the GlobalSoilMap initiative—an international collaborative effort aimed at producing a fine-resolution global grid of soil properties (Arrouays et al., 2014a; Sanchez et al., 2009). As the initiative progressed, the consortium developed a consensus-based specification document that defined good practice for DSM and articulated standards for next-generation soil maps (Arrouays et al., 2014a). These specifications include a common prediction grid, standard depth intervals, and a core set of soil properties to facilitate seamless integration into a global product. Through these developments, DSM has evolved from a primarily theoretical and academic pursuit into an operational framework for delivering soil information to scientists, decision-makers, and policymakers (Arrouays et al., 2017; Arrouays et al., 2020a; Arrouays et al., 2020b; Kidd et al., 2020).

1.1.2 Towards operational SOC mapping

The transition of DSM from a primarily academic activity to an operational, cost-effective framework has been driven by the convergence of technological advances and evolving societal demands (Minasny and McBratney, 2016; Minasny et al., 2013). On the technological side, increasing availability of soil measurements, environmental covariates, and advanced spatial inference models has expanded the technical feasibility of large-scale soil mapping (Arrouays et al., 2017). In parallel, growing recognition of the central role of soils in addressing global challenges, such as food security, climate change mitigation, and land degradation, has led to sustained demand for timely, high-quality soil information (Lal et al., 2021). These technological and societal drivers have reinforced one another, accelerating the operationalisation of DSM.

Soil measurements remain the cornerstone of DSM (Schmidinger et al., 2025). Heightened awareness of the importance of soil has stimulated extensive soil sampling efforts worldwide (Batjes, 2009; Batjes et al., 2017; Leenaars et al., 2014). At the same time, open-data initiatives and international collaboration have substantially improved access to harmonised soil databases. Several major initiatives now provide large, standardised soil datasets, some of which include repeated measurements at fixed locations, enabling not only spatial prediction but also assessment of soil properties change over time (Orgiazzi et al., 2018).

Earth Observation (EO) data provide a complementary pillar for operational DSM. Derived primarily from satellite, aerial, and ground-based Remote Sensing (RS), EO offers spatially explicit information with broad geographic coverage, high spatial resolution, and repeated temporal observations. Many EO-derived variables are directly or indirectly linked to soil-forming factors, making EO an effective and scalable source of covariates for DSM (Angelopoulou et al., 2019; Loiseau et al., 2019; Pouladi et al., 2023; Richer-de-Forges et al., 2023).

Machine Learning (ML) methods have further enabled this transition by offering flexible and computationally efficient spatial inference tools. Compared with traditional geostatistical approaches, ML models are better suited to capturing complex and nonlinear relationships between soil properties and environmental predictors, while scaling efficiently to large datasets (Arrouays et al., 2014a; Lamichhane et al., 2019; Wadoux et al., 2020b). The increasing availability of user-friendly ML software and computational resources has lowered barriers to their widespread application in DSM.

Together, the integration of soil measurements, EO-derived covariates, and ML-based inference now makes it feasible to map Soil Organic Carbon (SOC) at finer spatial resolution, across larger geographic extents, and increasingly in a dynamic, time-explicit manner (Hengl and MacMillan, 2019; Hengl et al., 2018; Heuvelink et al., 2020; Tziolas et al., 2024; Veronesi and Schillaci, 2019; Wadoux et al., 2020b). Producing spatially detailed,

quantitatively explicit and temporally evolving maps of individual soil properties has become largely a matter of computational capacity and the availability of soil measurement data and soil covariates. Overcoming these challenges seems to be only a matter of time.

1.1.3 Soil organic carbon

In the era of DSM, SOC is one of the most frequently modelled and mapped soil properties (Chen et al., 2022a). When referring to SOC, three related but distinct quantities may be implied if not explicitly specified:

- **SOC content** (e.g., g kg^{-1}): the mass fraction of SOC in the fine-earth soil;
- **SOC density** (e.g., g cm^{-3}): the mass of SOC per unit volume, incorporating Bulk Density (BD);
- **SOC stock** (e.g., Mg ha^{-1}): the mass of SOC within a given soil depth.

SOC content is typically relevant for soil health assessments, whereas SOC stock is most relevant for climate mitigation and carbon accounting. In this thesis, the term "SOC" is used in its appropriate context for each chapter, with explicit specification provided where necessary. In the Introduction and Synthesis, SOC refers to this soil variable in a general sense.

Interest in SOC arises from its central role in soil fertility: SOC supports plant growth, enhances nutrient cycling, and contributes to crop yield (Ma et al., 2023; McBratney et al., 2014b). Dokuchaev's early "isohumus" map, published in his 1883 monograph "Russian Chernozem" and initially commissioned to support agricultural development in Russia, can be regarded as an early map of SOC (Rusakova et al., 2022).

Over time, the recognised functions of soil have expanded beyond biomass production to include water filtering and conservation, biodiversity support, carbon storage, etc (Andrews et al., 2004; Lehmann et al., 2020; McBratney et al., 2014a; Stockmann et al., 2013). For all of these functions, SOC plays a central role: it drives energy and nutrient cycles, improves soil structure, buffers pH, filters pollutants, and enhances water quality (Blanco-Canqui et al., 2013). High or increasing SOC content is associated with healthier, more resilient soils (McBratney et al., 2014b), which explains why SOC is among the most commonly measured soil indicators (Bünemann et al., 2018). Mapping its spatial distribution is therefore essential for monitoring, assessing, and managing soil health.

Among the various ecosystem services soils provide, their potential to act as a Green House Gases (GHG) sink has received particular attention (Lal, 2004; McBratney et al., 2014b). SOC is one of the largest global carbon sinks, storing approximately three times more carbon than vegetation and about twice as much as the atmosphere (Sanderman et al., 2017; Smith, 2012). Consequently, SOC sequestration is increasingly considered a

contribution to climate mitigation, but credible and reliable Monitoring, Reporting, and Verification (MRV) systems are required for carbon accounting and emissions trading (Smith et al., 2020). SOC maps can support these needs by providing baseline carbon information, monitoring SOC changes across time and identifying locations suitable for soil-based carbon sequestration projects (Minasny et al., 2017; Minasny et al., 2013). This also introduces new demands for SOC maps: they must explicitly represent temporal dynamics to quantify change, and provide sufficient spatial detail to distinguish differences across different SOC sequestration units (Brummitt et al., 2024).

1.2 Research gap

1.2.1 SOC maps in EU

Europe is among the most intensively mapped regions worldwide with respect to DSM of SOC (Chen et al., 2022a; Panagos et al., 2022). Many European countries maintain well-established soil monitoring networks, which have enabled the production of national- and regional-scale SOC maps, including for Germany (Behrens and Scholten, 2006; Poeplau et al., 2020c), France (Chen et al., 2023; Lemercier et al., 2022), Spain (Durante et al., 2024), Denmark (Adhikari et al., 2014), the Netherlands (Helfenstein et al., 2024b), and Hungary (Szatmári et al., 2019).

At the European scale, a well-established soil monitoring network and infrastructure is in place (Panagos et al., 2013; Panagos et al., 2022), supporting SOC mapping at the continental level. Most notably, the Land Use/Cover Area Frame Statistical Survey (LUCAS) topsoil survey first collected approximately 20 000 harmonised soil samples across EU Member States in 2009 and has since been repeated regularly (Orgiazzi et al., 2018). Early attempts to estimate SOC relied on expert-knowledge-based pedotransfer rules linking SOC to soil type, land use, and climate (Van Ranst et al., 1995). Subsequent work refined these rules and applied them to spatially continuous covariates, resulting in the first continental SOC maps at kilometre-scale resolution (Jones et al., 2005). Following the implementation of the LUCAS topsoil survey, EU-wide SOC maps at resolutions down to 250 m have been produced (de Brogniez et al., 2015). In addition, Europe is represented in global SOC products such as SoilGrids (Hengl et al., 2014; Hengl et al., 2017; Poggio et al., 2021).

Despite these advances, most existing SOC maps remain static, typically representing topsoil layers to a maximum depth of 30 cm for a single reference period (2D) (Chen et al., 2022a). Yet SOC varies not only across space, but also over time. By incorporating temporal covariates, DSM can be extended from static snapshots to time-explicit mapping (2D+T) (Heuvelink et al., 2020; Tayebi et al., 2021; Yang et al., 2022). In Europe, however, only a limited number of studies have mapped SOC dynamics. These include local-scale assessments in north-western France (Ellili et al., 2019), national-scale reconstructions

in Belgium and Hungary (Meersmans et al., 2011; Szatmári et al., 2021; Szatmári et al., 2024b), continental-scale products such as the WorldSoils project, which provides EU-wide SOC maps at 100 m resolution for periods after 2018 (Wesemael et al., 2024).

Soil properties also vary with depth, which has increasingly been incorporated into DSM. Depth can be addressed either through harmonisation and interpolation of point observations to standard depth intervals (Meersmans et al., 2011) or by explicitly modelling depth as part of the inference framework (Hengl et al., 2017; Poggio and Gimona, 2014). While no single approach has emerged as universally superior (Liu et al., 2016; Ma et al., 2021; Nauman and Duniway, 2019; Roudier et al., 2020), available mapping products have gradually evolved from static two-dimensional representations (2D), to include time (2D+T), and more recently, depth and time jointly (3D+T). Following early demonstrations of spatio-temporal interpolation for soil properties at local scales (Gasch et al., 2015), a notable example of 3D+T soil mapping at national scale is provided by Helfenstein et al. (2024a), who produced 25 m resolution spatio-temporal SOM maps for the Netherlands spanning 1953–2023.

Taken together, existing SOC mapping efforts demonstrate that high-resolution and time-explicit SOC mapping is increasingly feasible at local to national scales. However, extending such approaches to produce consistent 3D+T SOC maps at the pan-European scale remains an unresolved challenge. This limitation arises from a combination of data and operational constraints.

Although the LUCAS soil monitoring project (Orgiazzi et al., 2018) and numerous national or project-based soil monitoring programmes (Emmett et al., 2016; Poeplau et al., 2020c; Ramos et al., 2017; Serrano et al., 2022; Smart et al., 2009) have substantially expanded SOC data availability in both time and depth, no single harmonised dataset currently provides SOC measurements that are both repeatedly sampled and consistently observed at greater depths across Europe.

At the same time, advances in EO have greatly improved spatial resolution, temporal frequency, thematic relevance to environmental processes, and temporal coverage, offering unprecedented opportunities for large-scale, time-explicit soil modelling (Radeloff et al., 2024; Wulder et al., 2019). However, the increasing volume and complexity of EO data, together with fragmented data access, heterogeneous preprocessing standards, and limited processing capacity available to most users, remain major barriers to their effective use in continental-scale, spatio-temporal DSM applications.

These limitations highlight a critical research gap: despite major advances in soil monitoring, EO data availability, and modelling methodology, there is currently no operational framework that enables consistent, high-resolution 3D+T SOC mapping across Europe over multi-decadal time scales.

1.2.2 Assessing the plausibility of SOC change from model predictions

Moving from static SOC maps to time-explicit predictions raises a more fundamental question: whether the inferred SOC changes are detectable and meaningful given the uncertainty inherent in current DSM–ML approaches.

In the context of DSM, SOC change is most commonly derived from time series of predicted SOC states, either by calculating differences between time steps or by fitting temporal trends to predicted values (Chen et al., 2019; Helfenstein et al., 2024b; Li et al., 2022; Meersmans et al., 2011; Meng et al., 2024; Szatmári et al., 2024b; Yang et al., 2022). An alternative strategy is to model SOC change directly, using repeated observations to derive change metrics that are then predicted as the target variable. By focusing on change rather than absolute SOC states, this approach is expected to reduce correlated noise between time steps (De Rosa et al., 2024). However, its application remains limited due to the scarcity of long-term, repeated SOC measurements. As a result, practical implementations are rare. Notable examples include the prediction of long-term SOC trends for cropland in Bavaria based on a 35-year time series (Broeg et al., 2024b), and the modelling of SOC change from repeated LUCAS topsoil samples in cropland and grassland systems (De Rosa et al., 2024).

Regardless of the modelling strategy, validation based on repeated SOC measurements is essential for assessing the accuracy and consistency of SOC change estimates. Yet the thus required validation data are rarely available. Where it has been attempted, both Helfenstein et al. (2024b) and Broeg et al. (2024b) reported poor performance in modelling SOC change despite good predictive skill for SOC states using ML and EO. Although De Rosa et al. (2024) achieved more promising results, their approach was not based on ML but instead relied on a data-driven, equation-based modelling framework.

In the absence of direct validation using repeated measurements, uncertainty quantification becomes the primary means of evaluating SOC change estimates. This is particularly critical given that SOC changes are often small relative to prediction uncertainty (Paustian et al., 2019; Poeplau and Don, 2013). A range of methods has been developed to quantify such uncertainty, with Quantile Regression Forests (QRF) among the most widely applied approaches in DSM (Vaysse and Lagacherie, 2017; Wadoux et al., 2020b). In line with the suggested DSM practices by GlobalSoilMap specifications, uncertainty estimation has become an integral component of DSM practice to ensure transparency and reproducibility, and is typically reported in the form of prediction intervals (Arrouays et al., 2014b; Heuvelink, 2014; Minasny et al., 2013).

Despite these advances, few studies have explicitly tested whether predicted SOC changes exceed their associated uncertainty. A rare quantitative example is provided by Szatmári et al. (2024b), who considered uncertainty in SOC change estimation but focused primarily on trend direction rather than magnitude. In De Rosa et al. (2024), uncertainty was

presented alongside predicted SOC change, but only at an aggregated, longitudinal level. Chen et al. (2022a) emphasized that "the range of the estimated trend should be compared with the map uncertainty to test whether the trend is plausible or of the same order of magnitude as the accumulated errors linked to spatial predictions at different dates."

Without systematic evaluation of whether predicted SOC changes exceed their associated uncertainty, SOC dynamics inferred from DSM products remain difficult to interpret and potentially misleading. This highlights a critical methodological gap: advancing SOC mapping from states to dynamics requires not only higher-resolution predictions, but also explicit frameworks for assessing the plausibility and detectability of modelled SOC change.

1.2.3 Towards soil-science–informed DSM

ML has undoubtedly transformed DSM, enabling more efficient and cost-effective production of spatial soil information. However, while ML models are effective at reproducing patterns present in training data, their performance is fundamentally constrained by the quantity, quality, and structure of available observations (Minasny et al., 2024).

SOC density provides a clear example of this challenge. As a key variable for both soil health assessment and climate-change mitigation, it is typically modelled as a single target variable derived from SOC content, BD, and Coarse Fragment (CF). In practice, SOC content is measured far more frequently than BD and CF, and inconsistencies in measurement protocols often prevent their joint use. Consequently, large amounts of SOC content data cannot be fully exploited for SOC density modelling, especially in time-explicit applications where BD observations are sparse.

Conventional DSM approaches usually adopt a univariate modelling strategy, treating SOC density as an independent prediction target. While effective for capturing spatial patterns, this approach ignores the intrinsic physical relationships among SOC content, BD, and CF (Van Der Westhuizen et al., 2023). When extended across time, such univariate formulations may lead to temporally unstable or physically implausible SOC density trajectories, even when pointwise predictive accuracy appears high. Although multivariate modelling has a long tradition in DSM and recent ML studies have demonstrated its potential for jointly modelling correlated soil properties, multivariate ML frameworks have rarely been applied specifically to reconstruct SOC density from its constituent variables.

A further step beyond multivariate ML is soil-science–informed ML, in which domain knowledge is explicitly embedded into the model structure to represent known soil relationships, constrain learning, and guide predictions toward physically meaningful solutions (Ma et al., 2019; Weindorf and Chakraborty, 2024). The relationship between SOC content and BD is well established in soil science (Adams, 1973; Federer et al., 1993; Robinson

et al., 2022; Stewart et al., 1970), yet it is seldom incorporated into ML frameworks. From the perspective of SOC dynamics, embedding such physical constraints may help reduce implausible compensation effects between variables and improve the temporal coherence of SOC density reconstructions under data limitations.

The difficulty of reliably reconstructing SOC dynamics from model predictions points to a broader methodological issue highlighted by Minasny et al. (2024): ML approaches prioritise predictive accuracy for SOC states, while giving insufficient consideration to soil-science context, model interpretability, and physical consistency. Whether embedding soil domain knowledge into ML models can improve the physical consistency and temporal coherence of SOC density reconstructions, particularly under conditions of sparse BD measurements, remains largely unexplored.

1.3 This thesis

1.3.1 Research objectives

Summarizing the background and research gaps presented above, this thesis aims to address the following overarching research question: how can high-resolution SOC dynamics across Europe be reliably reconstructed from EO-based ML predictions, and what are the uncertainty, plausibility, and physical consistency of these reconstructed dynamics?

To answer this question, the thesis is structured into two subjects, comprising four specific research objectives:

- **Subject 1:** Operational feasibility of high-resolution dynamic SOC mapping at the European scale
  - **RO1:** Transforming EO time series into covariates for large-scale dynamic SOC modelling across Europe.
  - **RO2:** Modelling SOC using EO data and ML–based DSM across space, depth, and time at the European scale.
- **Subject 2:** Credibility of model-based SOC dynamics
  - **RO3:** Evaluating the reliability of SOC change detection derived from DSM-based predictions, explicitly accounting for model uncertainty.
  - **RO4:** Integrating soil-scientific knowledge into ML models to support SOC dynamics predictions.

1.3.2 Thesis structure

This thesis consists of six chapters. Chapter 1 provides a general introduction, while Chapter 6 synthesises the entire work. The four research objectives guide the work

presented in Chapters 2 to 5. The relationship between the research objectives and the individual chapters is illustrated in Fig. 1.2. Chapters 2–5 are based on peer-reviewed publications or manuscripts currently under peer review and can therefore be read as self-contained studies.

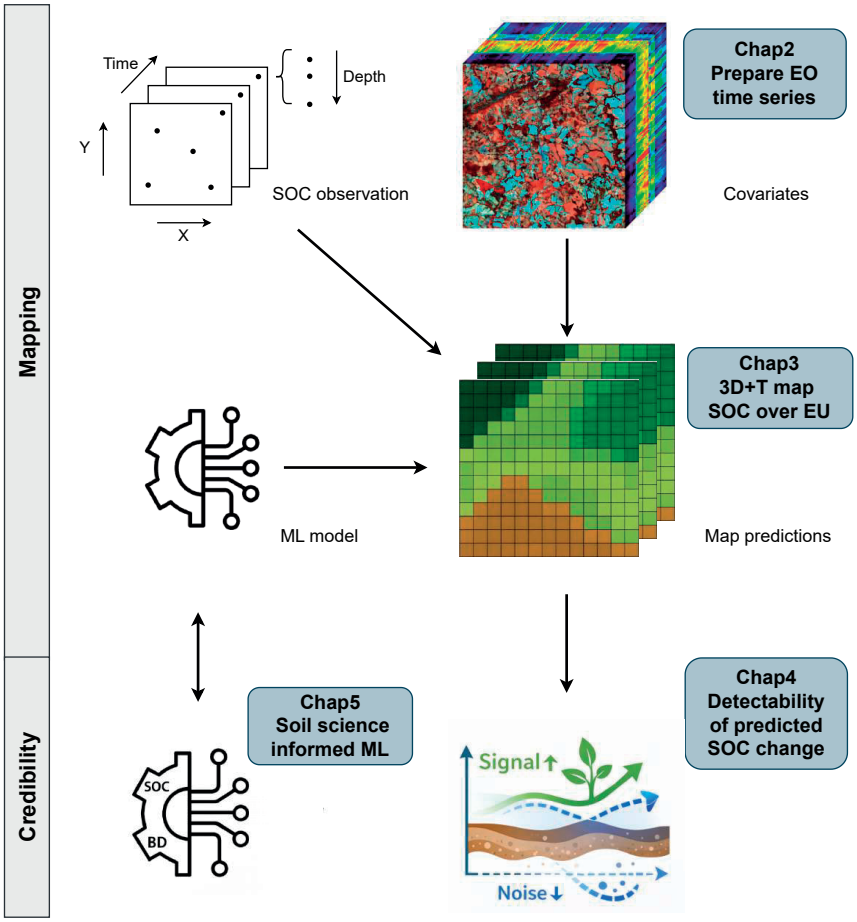


Figure 1.2: Conceptual framework of this thesis.

Chapter 1, the current chapter, introduces the thesis by outlining the scientific and societal motivations for monitoring SOC and reviewing recent developments in DSM that enable large-scale, spatio-temporal SOC modelling and mapping. It synthesises existing large-scale SOC mapping efforts and identifies key gaps related to data readiness, model uncertainty, temporal interpretation, and physical consistency. The chapter concludes by

formulating the research objectives that guide the thesis and outlining how the subsequent chapters address them.

Chapter 2 evaluates whether fragmented long-term EO archives can be transformed into a reliable, analysis-ready foundation for spatio-temporal SOC modelling across Europe. It presents a harmonised and cloud-optimised Landsat-based EO data cube (2000–2022) and systematically assesses its temporal consistency, gap-filling performance, and physical plausibility. This chapter establishes the suitability of EO data as scalable and fit-for-purpose predictors for subsequent SOC mapping and analysis.

Chapter 3 presents a measurements-to-maps ML modelling framework that integrates harmonised SOC observations with the EO feature cube and additional predictors. The framework produces biannual 30 m SOC density maps for multiple soil depth intervals down to 2 m for the period 2000–2022, together with per-pixel uncertainty estimates and extrapolation risk layers. By examining the accuracy, uncertainty, and dominant drivers of SOC mapping across Europe, the results confirm good overall predictive performance while also revealing that the magnitude and structure of prediction uncertainty impose clear limitations on the reliable detection of SOC change from modelled time series. This chapter therefore clarifies both the strengths and limitations of pan-European 3D+T SOC maps, demonstrating that they can serve as a valuable baseline for applications, while explicitly motivating the focus on SOC change detectability in the subsequent chapter.

Chapter 4 examines how the plausibility and detectability of SOC change can be systematically assessed given the uncertainty inherent in EO- and ML-based SOC predictions. It introduces a model-based Signal-to-Noise Ratio (SNR) framework, defining detectability as the ratio between predicted SOC change and its associated uncertainty. Applied to repeated SOC observations, results show that SOC change is generally not detectable at the pixel scale, reflecting limitations in data quality, temporal coverage, and model assumptions. However, spatial aggregation substantially improves detectability, enabling more reliable assessments at regional to national scales. This chapter demonstrates that SNR provides a practical diagnostic for evaluating when and under what conditions SOC change inferred from DSM products is meaningful.

Chapter 5 explores whether soil-science–informed hybrid ML models can improve the physical consistency and temporal coherence of SOC density reconstructions under data limitations. Addressing the imbalance between abundant SOC content measurements and sparse BD observations, the chapter compares three neural network architectures: a univariate model, a multivariate model, and a hybrid soil-informed model that embeds established SOC–BD relationships into its latent structure. Experiments using LUCAS data show that the soil-informed model produces more physically consistent SOC–BD joint distributions and more temporally coherent SOC density trajectories, reducing the implausible fluctuations observed in univariate predictions. This chapter demonstrates that

embedding soil-science knowledge into ML models can enhance SOC reconstruction beyond purely data-driven approaches, particularly in time-explicit and data-limited settings.

Chapter 6 synthesises the findings of the thesis by revisiting the research objectives and reflecting on the methodological and practical implications of large-scale, spatio-temporal SOC mapping. It discusses remaining challenges, identifies opportunities for improving SOC dynamics assessment, and outlines directions for future research.

Chapter 2

Preparation of high-resolution Earth Observation time-series for dynamic Soil Organic Carbon modelling

This chapter is based on:

X. Tian, D. Consoli, M. Witjes, F. Schneider, L. Parente, M. Şahin, Y.-F. Ho, R. Minařík, and T. Hengl (2025b). "Time series of Landsat-based bimonthly and annual spectral indices for continental Europe for 2000–2022". *Earth System Science Data* 17.2, 741–772. DOI: 10.5194/essd-17-741-2025

Abstract

The production and evaluation of the Analysis Ready and Cloud Optimized (ARCO) data cube for continental Europe (including Ukraine, the UK, and Turkey), derived from the Landsat Analysis Ready Data version 2 (ARD V2) produced by Global Land Analysis and Discovery team (GLAD) and covering the period from 2000 to 2022 is described. The data cube consists of 17TB of data at a 30–meter resolution and includes bimonthly, annual, and long-term spectral indices on various thematic topics, including: surface reflectance bands, Normalised Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), Normalised Difference Snow Index (NDSI), Normalised Difference Water Index (NDWI), Normalised Difference Tillage Index (NDTI), Minimum Normalized Difference Tillage Index (minNDTI), Bare Soil Fraction (BSF), Number of Seasons (NOS), and Crop Duration Ratio (CDR). The data cube was developed with the intention to provide a comprehensive feature space for environmental modeling and mapping. The quality of the produced time series was assessed by: (1) assessing the accuracy of gap-filled bimonthly Landsat data with artificially created gaps, (2) visual examination for artifacts and inconsistencies, (3) plausibility checks with ground survey data, and (4) predictive modeling tests, examples with Soil Organic Carbon (SOC) and land cover (LC) classification. The time series reconstruction demonstrates high accuracy, with RMSE smaller than 0.05, and R^2 higher than 0.6, across all bands. The visual examination indicates that the product is complete and consistent, except for winter periods in northern latitudes and high-altitude areas where high cloud and snow density introduce significant gaps, and hence many artifacts remain. The plausibility check further shows that the indices logically and statistically capture the processes. The BSF index showed a strong negative correlation (-0.73) with crop coverage data, while the minNDTI index had a moderate positive correlation (0.57) with the Eurostat tillage practices survey data. The detailed temporal resolution and long-term characteristics provided by different tiers of predictors in this data cube proved to be important for both soil organic carbon regression and LC classification experiments based on the 60,723 Land Use/Cover Area Frame Statistical Survey (LUCAS) observations: long-term characteristics (tier 4) were particularly valuable for predictive mapping of SOC and LC coming on the top of variable importance assessment. Crop-specific indices (NOS and CDR) provided limited value for the tested applications, possibly due to noise or insufficient quantification methods. The data cube is made available under a CC-BY license and will be continuously updated.

2.1 Introduction

Earth Observation (EO) data is increasingly recognized for its critical role in environmental sciences (Kansakar and Hossain, 2016; Salcedo-Sanz et al., 2020). Compared to traditional field surveys, it offers high-resolution, large-scale, and recurrent spatial environmental data at relatively low cost. These capabilities are important for comprehensive environmental studies, ongoing monitoring, and effective management (Chatenoux et al., 2021; Giuliani et al., 2021). EO data not only offer valuable insights into the Earth's surface through various spectral bands but can also highlight specific land surface features through spectral indices derived by processing them. In addition, satellite data serve as an essential input for various models that study physical processes, forecast future scenarios, and inform policy- and decision makers (Salcedo-Sanz et al., 2020).

As EO technology develops, more opportunities emerge. The development of a wider range of spectral indices allows a more detailed analysis using different combinations of satellite bands (Montero et al., 2023). Improvements in spatial and temporal resolutions enable more detailed observations. In addition, a longer record of satellite imagery facilitates long-term environmental studies. These advances help us better understand the environmental dynamics and enhance natural resource management strategies. However, they also introduce new challenges. One of the key challenges to fully unlock the potential of EO data for environmental applications, as highlighted by numerous researchers (Chatenoux et al., 2021; Giuliani et al., 2021; Killough, 2018; Montero et al., 2023; Wagemann et al., 2021), is the gap between the demand for detailed EO data and the limited processing capabilities of most users. As the volume of EO data and the complexity of spectral indices increase, processing typically requires specialized expertise and costly computational resources, both locally and in the cloud. For example, while platforms such as MODIS, Landsat, and Sentinel openly provide valuable satellite data, these data often require preprocessing to remove poor quality pixels, such as those affected by cloud cover and geometric distortions (Radeloff et al., 2024; Wulder et al., 2022). While many pre-processed layers are available, they tend to be scattered across various data portals and often focus on limited themes. Accessing these ready-to-use datasets usually demands domain-specific knowledge to even be aware of their existence. In addition, non-standardized data formats further complicate the ability of users to integrate these data sets for specific applications (Lokers et al., 2016; Wagemann et al., 2021). Thus, there is an essential need for solutions that ensure easy access to open environmental data. Furthermore, these solutions should also facilitate the easy integration of these data, enabling its combined use to effectively address critical environmental and economic challenges (Giuliani et al., 2017).

Numerous platforms have emerged in the last decade that aim to make the management, processing and analysis of big EO data operational, including Google Earth Engine (GEE, Gorelick et al., 2017), Sentinel Hub (SH, `https://www.sentinel-hub.com/`), Open Data Cube (ODC, `https://www.opendatacube.org/`), Copernicus Data Space Ecosystem

(https://dataspace.copernicus.eu/), and openEO Cloud (https://openeo.cloud/). Each platform has its strengths and weaknesses: GEE offers extensive datasets and robust processing, but is limited by its closed nature, restricting external contributions and governance. Sentinel Hub faces similar limitations. OpenEO provides flexibility with user-defined functions but lacks guaranteed compatibility across different back-end systems. Among these, the Open Data Cube (ODC) stands out for its open-source nature, flexibility, and strong community support, making it a leading option (Gomes et al., 2020). As a result, many analysis-ready data (ARD) cubes have been developed using ODC principles. ARD refers to a multidimensional time series stack of spatially aligned pixels that are ready for analysis, eliminating the need for additional data harmonization (Giuliani et al., 2021; Giuliani et al., 2017). These efforts reduce technical complexity and have been shown to be effective in delivering information efficiently, as demonstrated by successful implementations in various domains, such as vegetation dynamics (Obuchowicz et al., 2023), snow cover (Poussin et al., 2023), and drying trend (Poussin et al., 2021).

EcoDataCube.eu, developed by Witjes et al., 2023, is the first product to cover the entire EU with a sufficient temporal range to support a long-term analysis of land degradation and land use change dynamics. This work extends the concept of ARD to ARCO (analysis-ready and cloud-optimized) data cubes (Iosifescu Enescu et al., 2021; Stern et al., 2022), which helps to optimize cloud-based data management and processing. Cloud optimization allows for on-the-fly access, reduces latency through partial and parallel reads, and efficient metadata handling. Cloud-Optimized GeoTIFF (COG, https://www.cogeo.org/), adopted by EcoDataCube.eu, is a good example of such a format. COG files are structured to facilitate network access through HTTP range requests, ensuring compatibility with cloud object storage systems. This design supports integration with high-level analysis libraries and distributed frameworks. Additionally, EcoDataCube.eu also minimizes data gaps caused by clouds, which is a major obstacle to unlocking the potential of EO data (Baumann, 2024). This is achieved through time-reconstruction algorithms to ensure optimal cloud-free conditions (Consoli et al., 2024; Zhou et al., 2016). This approach not only prepares the data for spatio-temporal analysis but also proves beneficial for modeling, as models typically require complete input data to function effectively.

EcoDataCube.eu provides valuable insights through its quarterly temporal resolution and raw bands, making it a significant resource for Earth observation applications. To further reduce the burden on end users and increase the use and impact of EO data, we developed the ARCO data cube for the EU by refining the temporal resolution, processing reflectance bands to derive spectral indices, and updating the data layers until 2022. Our Landsat-based spectral indices data cube spans from 2000 to 2022, covering the pan-European area, including Ukraine, the UK, and Turkey, with data at a detailed 30m resolution. Using the Global Land Analysis and Discovery team (GLAD) Landsat Analysis Ready Dataset version 2 (ARD V2) presented by Potapov et al. (2020) as input, we adopt the

state-of-the-art approach of Consoli et al. (2024) to generate four tiers of environmental predictors:

1. Bimonthly aggregated cloud-optimized bands;
2. Bimonthly spectral indices derived from bands;
3. Annual indices derived by analyzing the bimonthly time series of indices;
4. Long-term indices reflecting features across two decades.

The foundational tier of the data cube consists of bimonthly cloud-optimized bands, derived using time reconstruction methods to address the gaps left by cloudy pixels (Consoli et al., 2024). Building upon this, we calculate bimonthly spectral indices as the second tier of predictors. This selection includes the most widely used indices, covering key aspects such as vegetation, crops, soil, and water (Montero et al., 2023): the Normalized Difference Vegetation Index (NDVI) is used to evaluate vegetation health and biomass. Complementing NDVI, the SAVI has been shown to be more effective in areas with sparse vegetation by minimizing the impact of soil brightness on vegetation sensing (Huete, 1988; Reddy et al., 2022; Rhyma et al., 2020). Additionally, the FAPAR provides a more direct measurement of plant productivity (Myneni and Williams, 1994). The NDWI, also known as the Normalized Difference Moisture Index (NDMI), provides insights into water dynamics and climatic characteristics (Gao, 1996). The NDSI helps identify snowy areas (Salomonson and Appel, 2006). The NDTI, also known as Normalized Burn Ratio 2 (NBR2), shows potential in tillage detection, post-fire recovery studies and soil sealing identification (Beeson et al., 2020; Daughtry et al., 2010; Eskandari et al., 2016; Ettehadi Osgouei et al., 2019; Sonmez and Slater, 2016; Storey et al., 2016; Xiang et al., 2022; Zheng et al., 2012).

Through time series analysis, we also derive several annual indices that capture temporal patterns within the year as our third tier of predictors. This includes the NOS and CDR from NDVI time series, which shed light on the intensity of cropland use (Estel et al., 2016; Li et al., 2014; Patel and Oza, 2014; Siebert et al., 2010). The BSF measures the duration of soil exposure, which is related to soil health (Demattê et al., 2020; Mzid et al., 2021; Sharma et al., 2018; Turmel et al., 2015). Although NDTI is positively correlated with the crop residue cover ratio (CRC), using it to distinguish between conventional, conservative, and zero tillage practices can be challenging without knowing the specific timing of tillage or planting (Zheng et al., 2013; Zheng et al., 2012). Therefore, its annual derivation, minimum NDTI (minNDTI), is adopted as a proxy for tillage due to its ease of application (Zheng et al., 2012). We also calculate annual percentile aggregations (25th, 50th and 75th) of NDVI and NDWI to quantify annual vegetation and water conditions. Finally, from these annual time series, we develop long-term indices that reveal decadal characteristics as the 4–tier predictors.

Given the comprehensive and extensive nature of this data cube, it is impractical to validate every derived spectral index against ground truth data. Plausibility checks were conducted selectively, focusing on 3rd tier predictors where feasible and necessary, and where relevant ground data were available. These checks involved matching the data with the ground survey statistics to highlight the advantages and limitations of the data. Our predictors data cube will be compared with EcoDataCube.eu, which has been utilized for land cover (LC) classification by Witjes et al. (2023). By benchmarking against EcoDataCube.eu, our aim is to highlight the improvements and advantages offered by our spectral indices data cube, specifically in terms of accuracy, robustness, and usability for land cover classification. To demonstrate the versatility and utility of the data cube for various environmental modeling purposes, we also present a case study focused on building regression modeling for SOC using thid data cube. Finally, we discuss the recommended uses, limitations, and future development of this data cube and compare it with other similar projects and initiatives. All the layers we produced are available under the open data license CC-BY and can be accessed and visualized via our `https://stac.ecodatacube.eu` as Cloud-Optimized GeoTIFFs (Witjes et al., 2023). To enable users to easily assess the suitability of datasets for specific applications, metadata is embedded within the layer name in a structured and standardized format, aligning with Data-Fitness-For-Use assessment methodologies (A. Wentz and Shimizu, 2018; Lacagnina et al., 2022; Lokers et al., 2016; Pôças et al., 2014; Yang et al., 2013). A copy of the data, including bimonthly reflectance bands, spectral indices, annual aggregations andlong-term analysis, is also available from Zenodo (Tian et al., 2024, https://doi.org/10.5281/zenodo.10776891); complete code used to derive all indices is available via `https://doi.org/10.5281/zenodo.12907281`.

2.2 Material and methods

2.2.1 Landsat ARD V2

The data cube presented in this study is based on the Landsat Analysis Ready Dataset version 2 (ARD V2), developed by the Global Land Analysis and Discovery (GLAD) team at the University of Maryland (Potapov et al., 2020). Landsat ARD V2, the second version of Landsat ARD, consists of 16-day tiled composites with 23 images per year from 1997 to 2022, totaling 598 images. It includes seven reflectance bands (blue, green, red, near-infrared, short-wave infrared 1, short-wave infrared 2 and thermal) and a detailed quality flag that classifies each pixel as land, water, cloud, cloud shadow, topographic shadow, hill shade, snow, haze, cloud proximity, shadow proximity, other shadows, or buffered proximity of the previously mentioned classes. These quality flags enable the identification of poor quality pixels, including those affected by clouds, cloud shadows, haze, or other shadowy conditions. The presence of these gaps requires significant preprocessing before the data can be used for direct modeling and analysis. As one of the few globally consistent archives for historical time series of normalized surface reflectances derived from

various Landsat satellite collections, Landsat ARD V2 offers long-term availability (since 1997) and detailed spatial resolution (30 meters). For this study, we cropped the global Landsat ARD V2 layers to a pan-European extent, including Ukraine, the UK and Turkey. The GLAD ARD tile system is adopted to support data organization, parallel processing, and spatial inference when necessary.

2.2.2 TSIRF framework and scikit-map

As a comprehensive framework for processing EO time series, Time Series Iteration-free Reconstruction Framework (TSIRF) enables diverse time-series processing techniques by simply adjusting the convolution kernel (Consoli et al., 2024). We adopted TSIRF for temporal aggregation and gap-filling to impute data gaps created after removing clouds using the ARD 2 quality assessment mask. This framework is implemented through the Python package scikit-map (https://github.com/openlandmap/scikit-map), which optimizes speed by leveraging data structures for parallel computing. In addition to TSIRF, scikit-map also offers straightforward capabilities for band operations, trend calculations, temporal statistics, parallel raster reading and writing, and processing. These functionalities enable efficient processing, analysis, and visualization of large multidimensional raster datasets. In this work, all data processing and map production were performed using scikit-map python library.

2.2.3 Spectral indices predictors data cube

The preparation of the spectral biophysical indices data cube involves developing four tiers of predictors from Landsat ARD V2, step by step. These tiers differ in temporal resolution and processing complexity, as outlined in Fig. 2.1 and detailed in the following subsections. Section 2.2.3.1 introduces the production of tier-1 bimonthly, gap-free Landsat reflectance bands. Tier-2 predictors are bimonthly time series for six biophysical indices (Section 2.2.3.2). From tier-2 indices, time series analyses are conducted to derive annual predictors (Section 2.2.3.3). Finally, further time series analyses are applied to tier-3 predictors to extract long-term temporal features representing persistent states (Section 2.2.3.4).

2.2.3.1 First tier of predictors: bimonthly Landsat bands

The first tier predictors are ARCO land surface reflectance bands, gap-filled using Landsat ARD V2 data. Pixels with insufficient quality were identified using the quality band and masked, creating gaps in the time series. These gaps were initially reduced by calculating a weighted average from several scenes (typically 6–7) within and adjacent to the two-month period of the original 16-day interval time series, resulting in a bi-monthly product (one image every two months). Despite this aggregation, significant gaps remained, as shown in Fig. 2.2, which were subsequently filled using the Seasonally Weighted Average

Generalization (SWAG) method developed by (Consoli et al., 2024). SWAG assigns weights for aggregation based on the clear sky fraction of the corresponding tiles. For gap-filling, SWAG respects seasonality and causality by reconstructing each pixel's time series using weighted averages, prioritizing images from the same period in previous years, and relying only on past data.

This Tier 1 product is part of the global bi-monthly aggregation framework established by (Consoli et al., 2024) and represents the European subset of that effort. Within this work, the data has been refined for regional applications by applying the pan-European landmask (https://doi.org/10.5281/zenodo.8171860) to exclude water bodies and assembling it into continental mosaics stored as COGs, as shown in step (1) of Fig. 2.1. These gap-filled bi-monthly ARCO Landsat surface reflectance layers provide the foundational material for producing spectral indices and their derivatives.

2.2.3.2 Second tier of predictors: bimonthly spectral index

The second tier of predictors is calculated per pixel using ARCO surface reflectance bands, as shown in step (2) of Fig. 2.1. Direct satellite-derived indices make use of the unique spectral signatures of different objects on the surface of the land to detect their presence within a pixel. Direct satellite-derived indices include NDVI, SAVI, FAPAR, NDTI, NDWI, and NDSI. The calculation formula we adopted is described in Table 2.1.

2.2.3.3 Third tier of predictors: annual aggregated index

In step (3), the aggregated annual indices are derived by extracting key temporal features or statistics from the bimonthly index series for each year. These indices are then provided at a 30m spatial resolution on an annual basis:

- **Annual P25, P50, and P75 aggregation of NDVI, NDWI and NDTI** are calculated by identifying the values at the 25th, 50th and 75th percentile of the sorted bimonthly NDVI and NDWI values for each pixel within a year.
- **Minimum Normalized Tillage Index (minNDTI)** is determined as the minimum value of the six NDTI values over a year (Zheng et al., 2012).
- **BSF** is calculated by dividing the number of pixels classified as bare surface within a year's time series (identified by the criterion of NDVI values below 0.35) by the total number of pixels analyzed in that year (Castaldi et al., 2019).
- **Number of season (NOS)** indicates the frequency of cropping cycles within a year, calculated by identifying peaks in the NDVI time series (Li et al., 2014; Liu et al., 2020). Time steps with NDVI values greater than 0.5 are first flagged as candidate peaks, possibly corresponding to an actual cropping cycle. A prominence filter of 0.25 ensures that each candidate peak is at least 0.25 higher than its surrounding troughs, addressing consistently high NDVI values in areas such as forests. Adjacent

Table 2.1: Overview of third tier of predictors

Index	Band Operation	Reference
NDVI	$\frac{\text{nir} - \text{red}}{\text{nir} + \text{red}}$	Tucker, 1979
NDTI	$\frac{\text{swir1} - \text{swir2}}{\text{swir1} + \text{swir2}}$	Van Deventer et al., 1997
NDWI	$\frac{\text{nir} - \text{swir1}}{\text{nir} + \text{swir1}}$	Gao, 1996
NDSI	$\frac{\text{green} - \text{swir1}}{\text{green} + \text{swir1}}$	Salomonson and Appel, 2006
SAVI	$\frac{(\text{nir} - \text{red})}{(\text{nir} + \text{red} + 0.5)} \times 1.5$	Huete, 1988
FAPAR	$\frac{(\text{ndvi} - 0.03) \times (0.95 - 0.001)}{0.96 - 0.03} + 0.001$	Robinson et al., 2018

peaks with intervals shorter than 2 months are merged, keeping the one with higher NDVI values, to ensure an accurate representation of distinct cropping cycles.

- **CDR** quantifies the length of active cropping periods, calculated as the ratio of time during which a pixel is in an active cropping state, defined as when the vegetation signals reach at least half the amplitude of the phenological curve's peak values (White et al., 1997). The peak values are determined by the average NDVI values of the peaks identified during the calculation of NOS.
- **Accumulated NDVI P50, NDWI P50, NDVI of the month of July and August, and minNDTI** The cumulative indices are calculated by summing their values over time from the year 2000. They are expected to reflect the cumulative effects of water content, vegetation health, and tillage practices on each pixel.

2.2.3.4 Fourth tier of predictors: long-term temporal feature index

In step (4), the indices that quantify long-term temporal features over 2000–2022 are calculated. The long-term changes trends are derived from annual aggregates to represent

the changes from 2000 to 2022. For this, we adopted the Theil-Sen estimator to fit the trend for 2000–2022. This is a statistical, non-parametric approach that is resistant to outliers, as it calculates the median slope among all point-pairs (Sen, 1968; Theil, 1950). This results in trend maps for BSF, NDVI P50, NDWI P50, and minNDTI. Additionally, the P25, P50, and P75 percentiles of NDVI, NDWI, and BSF are derived for the years 2000-2022 to provide a general overview of their distribution and general state over the this period.

2.2.4 Plausibility check

Due to the limited availability of land survey data that align temporally, spatially, and thematically with our data cube, we were unable to perform traditional accuracy validation on most index layers, which requires extensive ground-truth datasets for direct comparison. Instead, we conducted a plausibility check to assess whether the data align logically and statistically with established land surface processes. Apart from tier 1 product, we focus our quality assessment on the third tier of predictors—annual aggregated indices such as minNDTI, BSF, NOS, and CDR—since they are more complex and less established than simpler spectral indices like NDVI and NDWI.

2.2.4.1 Quality assessment of tier-1 product

The quality of the gap-filled time series for bimonthly Landsat surface reflectance bands was evaluated using 2,746 time series randomly sampled from Europe. Details are provided in the supplemental computational notebook available in the Section 2.5. To simulate real-world conditions, 10% artificial gaps were introduced into the data. Performance metrics were used to assess the accuracy of the TSIRF model, including R-squared (R2) which measures the proportion of variance explained by the model; Root Mean Squared Error (RMSE), quantifying the average magnitude of prediction errors; and Concordance Correlation Coefficient (CCC; Lawrence and Lin 1989), evaluating the agreement between observed and predicted values are reported to indicate the accuracy of TSIRF in reconstructing the tier-1 land surface reflectance bands.

2.2.4.2 Index statistics by crop type: BSF, NOS and CDR

To examine the effectiveness of crop-specific parameters — BSF, NOS and CDR-in reflecting the crop patterns — we used the harmonized LUCAS data, which details crop types information at specific locations and times (d'Andrimont et al., 2020). Three NUTS2 regions are selected to capture a diverse array of European agricultural practices and climates: Picardy region (FRE2), Piedmont province (ITC1), and Malopolska Province (PL21). Picardy is chosen for its temperate maritime climate and highly mechanized, large-scale farms. Piedmont is included to represent its unique damp rice paddies, which illustrate specialized crop cultivation. Malopolska offers insights into agriculture under a

more continental climate, contrasting with other regions. We identified points belonging to four predominantly grown crops in the selected NUTS2 regions—FRE2, ITC1, and PL21 from the most recent LUCAS 2018 dataset. These points were then overlaid onto the 2018 index maps for BSF, NOS and CDR, followed by the calculation of the statistics for crop-specific indices among the different crop types. By comparing the statistics of these indices across regions and crops, we conducted an evaluation of the indices' sensitivity and accuracy in capturing the distinctions in cropping patterns and reflecting the specific agricultural dynamics.

2.2.4.3 Correlation with land survey data: BSF

We matched BSF maps against two land survey datasets, with approximately 150 records of crop cover duration from 2007 to 2016 on cropland (Edlinger et al., 2022; Edlinger et al., 2023). The hypothesis is that the proportion of time that crops cover the soil is inversely related to the presence of bare soil during the same time period. By implementing a correlation analysis between BSF values derived from satellite data and ground-based measurements of crop cover duration, our objective was to assess the reliability of the BSF in capturing the extent of bare soil across agricultural landscapes.

2.2.4.4 Correlation with regional statistics: minNDTI

To validate the minNDTI's effectiveness in reflecting tillage practices across the EU, we compared it with Eurostat's tillage area statistics (https://doi.org/10.2908/EF_MP_PRAC). Eurostat's data originates from farm structure surveys conducted in 28 EU countries and is available in a vector shapefile format covering 319 NUTS2 regions, detailing arable areas under different tillage practices. The breakdown of arable land area based on tillage practices is as follows:

$$\sum(\text{TIL_CV} + \text{TIL_CSERV} + \text{TIL_ZERO} + \text{ARAXTIL}) = \text{ARA} \tag{2.1}$$

where TIL_CV refers to areas under conventional tillage, TIL_CSERV refers to areas under conservation tillage, TIL_ZERO refers to areas under zero tillage, ARA refers to total arable land areas, and ARAXTIL refers to areas land excluding from tillage survey; all areas recorded in hectares. Notably, ARAXTIL represents land that is excluded from tillage survey due to not being sown/cultivated during the reference year, which includes areas with multi-annual plants such as temporary grassland, leguminous plants, industrial crops (hops or aromatic plants) etc. These are recorded in hectares and together make up the total arable land area (ARA).

To spatially match the Eurostats data with our predictor layers, the data preparation process involved four steps:

1. **Calculate Tillage Shares**: For valid NUTS2 data records with all four types of tillage practices, calculate the share of each tillage practice within each NUTS2 region.
2. **Crop Masking**: Remove non-arable pixels from the minNDTI maps using the EU crop mask developed by d'Andrimont et al., 2021.
3. **Spatial Overlay**: Overlay the Eurostat survey data with crop-masked minNDTI map layers of matching years.
4. **Calculate Average minNDTI**: Calculate the average minNDTI value within each NUTS2 region.

The relationship between each tillage practice and minNDTI is assessed separately using Ordinary Least Squares (OLS) regression models, correlating the shares of each tillage practice with the average minNDTI values within each NUTS2 region. To capture the collective impact of all tillage practices, we introduced the concept of Weighted Crop Residue Cover (WCRC). This metric aggregates the influence of various tillage practices, considering minNDTI's correlation with crop residue cover (CRC):

$$\text{WCRC} = \sum_{\text{tillage types}} (\text{typical CRC value} \times \text{area share}) \tag{2.2}$$

where WCRC integrates typical CRC values for each tillage type, weighted by their respective area shares within each NUTS2 region.

We fitted typical CRC values for each type of tillage practice from their respective ranges, as outlined in Table 2.2 (Magdoff, Van Es, et al., 2000; Zheng et al., 2012). This optimization process identified the values within these ranges that maximize the correlation between WCRC and minNDTI. For excluded tillage practices, we assume a broad range of 0 to 100 percent to cover all possibilities, reflecting the uncertainty in its typical CRC values. The correlation analysis assesses the relationship between the estimated WCRC and the average minNDTI of the NUTS2 regions.

Table 2.2: Typical CRC value range for each tillage practice type

Tillage Practice Type	CRC Range (%)
Conventional tillage	0–30
Conservation tillage	30–70
Zero tillage	70–100
Excluded tillage practices	0–100

2.2.5 Comparison to EcoDataCube in LC classification

The Landsat-based predictor data cube are intended as an improvement of the version in the EcoDataCube (Witjes et al., 2023). The switch to a bimonthly temporal aggregation from a seasonal one with three percentiles leads to a smaller number of variables (6 instead of 12) and a higher level of temporal resolution. In order to quantify the difference of these aggregation techniques, we compare the performance of sets of random forest classifiers trained on both sets of Landsat data to predict seven types of land cover: Cropland, Grassland, Bare land, Forest, Artificial land, Shrubland, and Wetlands. We exclude the EcoDataCube non-Landsat data (DTM and Sentinel–2) and indices in this work to guarantee an objective comparison. Details about the training data will be discussed in the next section, while an overview of the experiment workflow is presented in Fig. 2.3.

2.2.6 Modeling experiments

To demonstrate the utility of the indices in the data cube, we performed two modeling experiments: SOC regression and LC classification. The general workflow is depicted in Fig. 2.3. The data points used for SOC regression and land cover classification come from the LUCAS land cover and top soil surveys. Only surveyed point data (*i.e.* identified by unique point id and sampling year) with available both land cover and topsoil SOC data were selected for modeling experiments. These point data were overlaid with EcoDataCube Landsat data, along with all Landsat reflectance and biophysical index layers generated in this study, to create the training dataset for modeling experiments. Using stratified sampling, we select about 10 % of test data points based on a 30 km grid tiling system to ensure that the test data accurately represent the spatial distribution throughout the study area. This results in 52.306 training data points and 8.417 test data points, in total 60.723 points.

The models' performance for both tasks is evaluated by training them with different combinations of predictors, organized into two main categories: themes and tiers. The first division categorizes the predictors by themes: reflectance bands, vegetation, water, soil, and crop (see Table 2.3). The second division organizes predictors by tiers, which represent different levels of temporal aggregation and processing, as detailed in Section. 2.2.3. These tiers, as described above, include bimonthly bands, bimonthly indices, annual indices, and long-term indices. By systematically testing models with these combinations, we can understand how indices developed for different purposes and how short-term variations, seasonal dynamics, and long-term temporal features contribute to the accuracy and robustness of predictive models. Both SOC regression and land cover classification models were trained using different groups from both divisions. We examined the feature space to identify the most influential predictors and compared their importance between the regression and classification tasks. This analysis aimed to highlight the flexibility and comprehensive nature of our data cube. Apart from comparing different feature

combinations, we also provide the top 20 most important features for the model trained on all predictors and all point data. This provides a focused analysis that identifies which specific predictors are the most influential across the entire dataset when no subsets are used.

Table 2.3: Theme-Based Division of Predictors

Theme	Predictors
Reflectance Bands	Bimonthly Landsat bands Annual P25, P50, and P75 of Landsat bands
Vegetation	Bimonthly NDVI, SAVI, FAPAR Annual NDVI P25, P50, and P75; acumulated NDVI P50 and NDVI of the bimonth July and August Long term P25, P50, P75 and trend of NDVI
Water	Bimonthly NDWI, NDSI Annual NDWI P25, P50, and P75 Long term P25, P50, P75 and trend OF NDWI
Soil	Bimonthly NDTI Annual minNDTI, NDTI P25, P50, and P75, and BSF Long term P25, P50, P75 and trend of BSF, and long term trend of minNDTI
Crop	Annual NOS and CDR

For the SOC regression experiment, we used the random forest (RF) as an estimator to predict the SOC content throughout the EU. The target variable, SOC, was obtained from the LUCAS soil survey. We evaluated the performance of the regression model using different combinations of predictors to identify which set provided the most accurate predictions. Performance metrics used to assess model accuracy include R2, RMSE, and CCC. We used a Random Forest regression model with the following parameters: a maximum depth of 20 to control complexity, `sqrt` for max features to improve performance, a minimum of 2 samples per leaf to prevent overfitting, a minimum of 5 samples per split for stable node splitting, and 800 estimators to ensure robust predictions. These parameters were selected through empirical testing and cross-validation, implemented using the scikit-learn library in Python. The metrics will also be calculated in the *log1p* space.

For the LC classification experiment, we employed a random forest classifier to predict LUCAS LC classes. Similarly to the SOC regression, we tested the model's performance with different combinations of predictors to determine the most effective feature sets.

Performance metrics used to evaluate the classification model include precision, assessing the accuracy of positive identifications; recall, evaluating the model's ability to capture all relevant cases; and F1-score (Taha and Hanbury, 2015), providing a balance between precision and recall for a comprehensive model assessment.

2.3 Results

This section begins with an accuracy assessment of time series reconstruction using the TSIRF method to generate tier-1 predictors. This is followed by a qualitative visual examination of spatial patterns in three zoomed-in areas, each representing distinct dominant land cover types and unique land dynamics over the past decades. Additionally, long-term trends of BSF, NDVI, and NDWI are analyzed using annual time-series data from 2000 to 2022 at three LUCAS points within these areas. These analyses provide both spatial and temporal insights into land cover changes and agricultural practices, demonstrating how these indices capture and reflect patterns and trends over time. Where relevant data are available, quantitative assessments are also incorporated to further validate our approach, as elaborated in the Section 2.2.4.

2.3.1 Accuracy of reconstructed tier-1 predictors

The Table 2.4 presents performance metrics for the reconstructed Landsat surface reflectance bands across tier-1 seven spectral bands. Overall, the reconstruction demonstrates high accuracy, with low RMSE, high R^2 and CCC across all bands. Thermal band consistently shows the best performance, while the NIR band exhibits the lowest performance, with slight variations observed among the other bands. In a similar global-scale experiment, Consoli et al. (2024) reported generally higher reconstruction performance. For a more comprehensive perspective on global time-series reconstruction, we refer readers to their work, which focuses exclusively on this topic.

Table 2.4: Performance metrics for the reconstruction of tier-1 product

Metric	Blue	Green	Red	NIR	SWIR1	SWIR2	Thermal
RMSE	0.04	0.04	0.04	0.05	0.03	0.02	0.01
RMSE/μ	0.86	0.51	0.56	0.20	0.15	0.22	0.01
RMSE/σ	0.56	0.56	0.58	0.60	0.41	0.40	0.32
$\mathbf{R^2}$	0.67	0.67	0.67	0.62	0.83	0.84	0.90
CCC	0.83	0.82	0.81	0.79	0.91	0.91	0.95

2.3.2 Visual examination of tier-3 predictors

Figure 2.4 presents zoomed-in examples of tier-3 predictors: NDVI P50, NDWI P50, BSF, NOS, and CDR. They provide an overview of vegetation health, moisture levels, bare soil exposure, number of growing seasons, and crop duration. These visual representations complement the statistical analysis by highlighting spatial patterns that may not be evident through quantitative methods alone, providing an intuitive understanding of the spatial patterns of the indices. The three presented regions have distinct features (Fig. 2.4): (A) an agricultural area in northern France, characterized by well-organized grid-like field patches of various crops; (B) a rice cultivation region that featured less regular field texture in northern Italy; and (C) the former Szczakowa sand mine in southern Poland (Pietrzykowski, 2008; Pietrzykowski and Krzaklewski, 2007).

NDVI and NDWI delineate agricultural patches with 30 m spatial resolution (Fig. 2.4-A). Forested areas show high values of NDVI and NDWI, while urban areas (scattered towns in Fig. 2.4-A and Fig. 2.4-B) and exposed land from mining activities (Fig. 2.4-C) exhibit low values. High NDWI values in region B, attributed to flooded paddy fields, do not always coincide with high NDVI (Ranghetti et al., 2016). For example, there are several patches in Fig. 2.4-B exhibiting high NDWI but moderate NDVI values. This observation could be attributed to the paddy fields in region B, which are flooded in a controlled way to meet the water demand of the rice plants (Ranghetti et al., 2016).

Although derived from the NDVI annual time series, the BSF shows different patterns. Urban and exposed soil areas have high BSF values, while woodlands have low values. Despite lower NDVI P50 values, region B has lower BSF values due to stable NDVI levels above a threshold, indicating less bare soil exposure throughout the year. This highlights that BSF reveals different aspects of the landscape compared to NDVI P50, making it a valuable complement to assess exposure to bare soil and vegetation dynamics. The crop-specific NOS and CDR provide insights into growing seasons and their duration. CDR highlights the duration for which crops occupy a pixel within a year and generally inversely relates to BSF. Most croplands in regions A and B have a NOS value of 1, indicating a single annual growing season, while urban and forested areas have a NOS value of 0 (Fig. 2.4-C). Some areas show NOS values of 2, indicating possible practices such as winter cropping or double-cropping systems.

2.3.3 Trend analysis with land cover dynamics

Four trend maps were produced to analyze land surface trends between 2000 and 2022: NDVI annual median, NDWI annual median, BSF, and minNDTI (Fig. 2.5). The NDVI trends reveal a positive trend in most parts of Europe. However, specific regions, including the Alps, Northern Europe, northern Scotland, Iceland, and Scandinavia, exhibit clear negative NDVI trends. These areas, characterized by high latitudes or altitudes, often experience persistent snow cover and frequent cloudiness. The NDWI trend map for these

regions shows a positive trend, indicating an increase in vegetation water content. In other parts of Europe, the NDWI trends display a mix of positive and negative changes, lacking a definitive overall trend. The decreasing trend in NDVI in regions of high altitude and high latitude is also reflected in the BSF trend map (Fig. 2.6). In Spain and Turkey, strong positive and negative trends are scattered, with the positive trend being more dominant. Generally, our results show an increase in minNDTI from 2000 to 2022 across Europe. However, there are exceptions, such as the negative trends observed in regions such as Turkey.

In Region A, Northern France, all BSF, NDVI, and NDWI trends show scattered increases and decreases, likely influenced by well-planned crop patches and their varying management practices (Fig. 2.7). Most areas exhibit positive NDVI and NDWI trends, except for urban areas such as Saint-Quentin city, where both vegetation and water content are declining. The time series also shows significant variation, influenced by crop types and management practices.

In Region B, northern Italy, the NDVI trends exhibit variability, but the overall trend is not as positive as in Region A (Fig. 2.8). BSF generally remains stable, while some areas exhibit an increasing trend, indicating greater soil exposure. The dominant decline in NDWI highlights a drying trend, likely due to the increasing adoption of dry seeding practices in rice cultivation (Ranghetti et al., 2016). The time series for the three indices has remained relatively stable over 20 years, reflecting consistent rice cultivation practices.

Region C in Poland, which has transitioned from a sand mine to a reforested area (Pietrzykowski, 2008; Pietrzykowski and Krzaklewski, 2007), shows the most significant positive trends in BSF, NDVI, and NDWI (Fig. 2.9). This indicates a steady increase in vegetation cover and water content, which is also reflected in the time series data. Although it was a coniferous forest from 2009 to 2018, the increases in NDVI and NDWI, coupled with the decrease in BSF, are likely due to the continuous growth of trees. Fig. 2.10 shows satellite images from 2000, 2010, and 2020, each accompanied by the corresponding BSF images for comparison, clearly illustrating the progression of reforestation during the last two decades.

2.3.4 Crop specific indices statistics across European regions

Fig. 2.11 shows the average values of BSF, NOS, and CDR for maize, sugar beet, rice, which are common crops in the NUTS2 regions PL21, FRE2, and ITC1, respectively. The NDVI time series for 2018, from which these indices are derived, are shown in the bottom panel. Sugar beet in the FRE2 region has the highest average NOS value, due to its distinct NDVI time series fluctuations, indicating pronounced seasonal variations. The NDVI time series for sugar beet shows a distinct peak during the summer months (July and August), a deep valley in March and April, and a less pronounced peak in winter. In comparison, the NDVI time series for maize in the PL21 region and rice in the ITC1

region display more uniform patterns, with smaller differences between peak and valley NDVI values, leading to lower NOS values. Sugar beet in the FRE2 region also exhibits the highest BSF values due to a long and deep valley period in the NDVI time series. In contrast, rice in the ITC1 region shows the lowest BSF values, reflecting its stable vegetation coverage throughout the year. Consequently, rice in the ITC1 region also has the highest CDR value. Although BSF and CDR often show a negative correlation, this is not always the case. For example, sugar beet in the FRE2 region has a higher BSF compared to maize in the PL21 region but also has a higher CDR. This highlights that, although related, BSF and CDR can differ due to their different focus: BSF measures the duration of bare soil, whereas CDR emphasizes the active growing period.

Fig. 2.12 shifts the focus from different crop types across regions to the same crop type—common wheat—grown in the FRE2, PL21, and ITC1 regions. The average values of BSF, NOS, and CDR for common wheat reveal notable regional variations, despite the NDVI time series overlap. Common wheat in the FRE2 region displays significantly higher NOS values compared to the other two regions, due to a secondary growth peak in winter. It also has the highest CDR value, indicating the most active growing cycles. In contrast, common wheat in the ITC1 region has the lowest BSF.

2.3.5 Assessing BSF in cropland

An inverse correlation, as illustrated in Fig. 2.13, is evident between the recorded crop coverage data and the BSF values from 2007 to 2016 in the cropland. Given the complexities involved, including the varying temporal resolutions of observation, the effects of crop residues, and the emergence of natural vegetation and weeds, a perfect one-to-one correlation between these variables is not expected. This observed correlation, in line with our hypothesis, confirms the validity of the BSF data.

2.3.6 Compare minNDTI with Eurostat at NUTS2 level

Fig. 2.14 shows the Pearson correlation between the shares of each individual type of tillage practice and their impact on the minNDTI values. Although the correlation coefficients for all the tillage practices suggest only a moderate correlation between the mean minNDTI values and the respective shares of each tillage practice, comparing the regression coefficients provides valuable information about their relative impacts. The sensitivity of the average minNDTI to each tillage practice is represented by the regression slope coefficient obtained from the OLS model fitting. Conventional tillage practices, characterized by soil inversion and minimal soil cover post-harvest, exhibit the most negative influence on average minNDTI in NUTS2 regions. This is followed by conservation tillage practices. In contrast, zero-tillage practices, which minimize soil disturbance, show the most positive effect on the average minNDTI. The area excluded from the tillage practices analysis, denoted

as *"araxtil"*, consists mainly of multiannual plants, resulting in a positive impact on the average minNDTI as indicated by the sensitivity analysis.

The correlation analysis between the WCRC and the average minNDTI of NUTS2 regions indicates that the cumulative impact of all tillage practices on the average minNDTI is more significant than any individual practice (Fig. 2.15). The optimal CRC for the tillage practices excluded from the analysis is relatively high at 75.08 %, consistent with Eurostat's statement that the majority of land excluded from tillage practices comprises multi-annual plants.

To further illustrate the spatial distribution and temporal trends of minNDTI, we used the minNDTI map, masked by the crop mask of d'Andrimont et al. (2021), to generate maps showing the average minNDTI for 2016 across the croplands of each NUTS2 region. Furthermore, we created a map that illustrates the average minNDTI trend between 2000 and 2022 per NUTS2 region (Fig. 2.16). In 2016, NUTS2 regions in Scandinavia and Ireland exhibited high minNDTI values, in contrast to regions in Spain and eastern Europe, which generally displayed lower minNDTI values on average. The trend map for minNDTI reveals that NUTS2 regions around the Baltic Sea are experiencing a relatively strong negative trend, indicating a possible increase in the intensity of tillage. In contrast, areas in eastern Britain, the Netherlands and Austria exhibit a moderate positive trend. In general, the European Union presents a negative trend in minNDTI. However, caution must be exercised when interpreting minNDTI data, as the correlation between minNDTI values and tillage practices is not robust enough to draw definitive conclusions without additional information.

2.3.7 LC classification results compared to EcoDataCube.eu

Both LC models achieved a weighted F1–score of 0.77, indicating similar performance. However, the model trained on this Landsat-based predictor data cube (*i.e.* 6 bimonthly and 3 annual Landsat features per band) achieved a macro F1–score of 0.47, versus a macro F1-score of 0.45 for the model trained on Landsat bands from EcoDataCube.eu (*i.e.* 4 quarterly Landsat data with 3 percentiles per quarter). This is mainly due to the higher classification accuracy for the relatively rare *"wetlands"* class (202 points, see Fig. 2.17).

2.3.8 Feature importance analysis of SOC regression and LC classification

2.3.8.1 Examination across tiers

From Table 2.5, we see the performance of the SOC regression model evaluated using different combinations of predictors categorized by tiers. When models are trained solely on one tier of predictors, tier 4 predictors (long-term characteristics) outperform the other

tiers across all three metrics. In contrast, tier 2 predictors (bimonthly spectral indices) show the least optimal CCC and R-square, while tier 1 predictors (bimonthly reflectance bands) have the least optimal RMSE. For models trained on combinations of two tiers of predictors, the combination of tier 1 and tier 4 predictors performs best across all three metrics, while the combination of tier 2 and tier 4 predictors performs the least well. When examining combinations of three tiers of predictors, most metrics are quite close, but the combination of tier 1, tier 2, and tier 4 still stands out slightly. When comparing across the number of tier sets combined, the general observation is that the more predictors combined, the better the model performed, with one exception: the model trained only on tier 4 predictors, despite being the smallest feature set, surprisingly achieved the best performance across all combinations.

Table 2.5: Metrics for SOC regression using different combinations of predictor based on tiers.

Number of tiers combined	Tiers	Feature number	CCC	RMSE	R2
1	tier1	42	0.635	0.651	0.489
	tier2	36	0.623	0.660	0.476
	tier3	38	0.637	0.652	0.487
	tier4	13	**0.682**	**0.613**	**0.547**
2	tier1+tier2	78	0.647	0.642	0.503
	tier1+tier3	80	0.648	0.642	0.503
	tier1+tier4	55	0.661	0.630	0.521
	tier2+tier3	74	0.648	0.643	0.501
	tier2+tier4	49	0.644	0.643	0.502
	tier3+tier4	51	0.659	0.634	0.516
3	tier1+tier2+tier3	116	0.652	0.639	0.507
	tier1+tier2+tier4	91	0.664	0.629	0.523
	tier1+tier3+tier4	93	0.663	0.630	0.522
	tier2+tier3+tier4	87	0.662	0.631	0.520
4 (all)	all tiers	129	0.665	0.629	0.523

An examination of the model performance across different tiers, as shown in Fig. 2.18, reveals that including tier 1 and tier 2 features generally has a negative effect. In contrast, incorporating tier 3 and tier 4 features positively impacts performance. Notably, tier 4 features show to have the most positive impact on model performance, consistent with the findings from Table 2.5.

Unlike the SOC regression experiment, the LC classification experiment indicates that combining features from tier 1, tier 2, and tier 4 results in the highest accuracy (see

Table 2.6). Although tier 4 is the best single-tier feature set for models in terms of F1-score, precision, and recall, it does not outperform all possible feature combinations. Models utilizing tier 3 predictors (annual spectral bands and indices) show the least optimal performance. For models using two-tier combinations, the pairing of tier 1 and tier 4 yields the best results across all three metrics, followed by the combination of tier 2 and tier 4. When examining models that incorporate three tiers, the combination of tier 1, tier 2, and tier 4 stands out, achieving the highest F1-score, precision, and recall among all possible feature set combinations, even slightly outperforming the feature set with all features.

Table 2.6: Metrics for LC classification using different combinations of predictors based on tiers.

Number of tiers combined	Tiers	Feature number	F1-score	Precision (U.A.)	Recall (P.A.)
1	tier1	42	0.769	0.776	0.784
	tier2	36	0.776	0.781	0.789
	tier3	38	0.761	0.762	0.773
	tier4	13	0.785	0.782	0.791
2	tier1+tier2	78	0.782	0.787	0.794
	tier1+tier3	80	0.774	0.777	0.787
	tier1+tier4	55	0.785	0.792	0.799
	tier2+tier3	74	0.777	0.781	0.790
	tier2+tier4	49	0.785	0.791	0.797
	tier3+tier4	51	0.775	0.777	0.787
3	tier1+tier2+tier3	116	0.779	0.783	0.792
	tier1+tier2+tier4	91	**0.790**	**0.796**	**0.802**
	tier1+tier3+tier4	93	0.782	0.785	0.795
	tier2+tier3+tier4	87	0.784	0.788	0.796
4 (all)	all tiers	129	0.786	0.789	0.798

Fig 2.19 shows that including tier 1 (bimonthly reflectance bands), tier 2 (bimonthly indices), and tier 4 (long-term characteristics) all improve model performance. Similar to the SOC regression results, models incorporating long-term trends (tier 4) exhibit the best performance metrics on average, while models including annual features tend to decrease model performance.

2.3.8.2 Examination across themes

Table 2.7 presents the performance of the model for different combinations of predictors based on themes. When examining models trained solely on one predictor theme, reflectance band predictors perform the best across all metrics, while crop predictors perform the worst. It is important to note that crop predictors only include two variables: NOS and CDR. For models trained with combinations from two themes, those involving band predictors generally perform much better than others, especially when combined with water and vegetation predictors. In contrast, the inclusion of crop predictors appears to degrade the performance of the model. When three predictor themes are combined, the best performance is found with the combination of band, soil, and water predictors, closely followed by combinations of band, soil, and vegetation, as well as band, water, and vegetation. Models lacking band predictors but including crop predictors perform worse, which is also true for models trained on four themes. Adding crop predictors to models that already include all other predictors does not significantly change the performance of the model. When comparing model performance across all possible theme-based combinations, it is evident that crop predictors do not add value to the model and can even harm performance when insufficient other predictors are available. In contrast, soil, vegetation, and water predictors all enhance the model, and the combination of these three themes provides the most significant improvement in performance.

Fig. 2.20 shows the R-square comparison of using or not using features from a certain theme as input for the SOC regression model. Including more thematic groups of features almost always bring in added value, especially reflectance bands, except for crop indices. The general performance patterns of the LC classification models, trained on different feature sets, resemble those of the SOC regression models (see Table 2.8). For LC classification models trained on a single theme feature set, crop theme features result in the worst metrics, while reflectance bands yield the best performance. Similarly to SOC regression, two-theme combinations that include band theme features generally perform better. For three-theme combinations, the combination of band, soil, and water themes is the most effective. For models trained on combinations of four themes, the best performance is observed when crop features are excluded. In particular, the model trained on the entire feature space achieved the highest accuracy metrics, with a minimal difference to the model trained on all features except crop indices, specifically using the combination of band, soil, vegetation and water themes (0.786331 versus 0.786330).

On average, models that incorporate Landsat bands achieved the highest F1-scores, while those including crop indices had the lowest, as shown in Fig. 2.21. Including additional thematic predictors, in addition to crop indices, generally improved model performance.

2.3.8.3 Comparison of feature space between SOC regression and LC classification

Comparing the results of the tier-based and theme-based for both SOC regression and LC classification, the improvement in model performance from combining more tiers is less significant than the enhancement seen when increasing the number of combined themes. The contribution of features from different tiers varies between SOC regression and LC classification: including tier 3 and tier 4 features improves the SOC regression model, while tier 3 is the only feature tier that decreases model performance for LC classification. In both experiments, tier 4 demonstrates its importance. Regarding the feature examination across themes, both experiments show similar results: reflectance bands provide a solid foundation for the models, and different thematic features add valuable information. However, crop indices contribute very little and can even degrade model performance when other variables are insufficient.

2.4 Discussion

2.4.1 Comparison to other open EO datasets

Our ARCO data cube, with almost 17TB of data for Europe at a 30 m spatial resolution from 2000 to 2022, represents a significant improvement over existing pan-EU EO data sets in terms of completeness, consistency, and comprehensiveness. Through an automated workflow, we ensured that all data layers adhere to uniform standards in naming conventions, projection systems, spatial extents, and resolutions. This consistency allows users to easily combine layers over space and time, enabling detailed studies using various indices. Our multidimensional index data cube, optimized as ARCO mosaics to be cloud-free and gap-free, significantly reduces preprocessing efforts compared to using raw data from open data archives such as Sentinel (https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data), MODIS (https://modis.gsfc.nasa.gov/data/dataprod/), or Landsat (https://www.usgs.gov/products/data), which are freely provided by agencies like USGS, NASA, and the European Commission's Copernicus and Sentinel programs.

The concept of ARD has been applied in projects to manage multidimensional data cubes from remote sensing images for specific countries and regions, such as the Australian Data Cube (Lewis et al., 2017), Swiss Data Cube (Giuliani et al., 2017), and Africa Regional Data Cube (Killough, 2019). Although Witjes et al., 2023 offers the first pan-European data cube, it is limited to quarterly temporal resolution and only cloud-optimized reflectance bands. Our data cube extends this work by refining the temporal resolution and developing four tiers of predictors. In addition to bimonthly indices that capture detailed environmental dynamics, annual aggregations enable yearly comparisons, and long-term characteristics represent general states and trends of vegetation, water, and crop management across Europe. This advancement benefits from two advantages of the

Landsat program, the world's longest continuously running EO mission and detailed 30 m spatial resolution (Wulder et al., 2022). These indices complement ground survey data, which often lack temporal continuity and are limited in duration due to the significant time and effort required.

Although some data sets provide continuous time series of spectral indices, they are often scattered and inconsistent in format, resolution, and coverage (Wagemann et al., 2021). For instance, most reconstruction efforts focus on NDVI (Zhou et al., 2016), while indices such as NOS, CDR, and BSF are less available (Demattê et al., 2020; Estel et al., 2015; Estel et al., 2016; Ettehadi Osgouei et al., 2019). Our data cube's multidimensional nature provides comprehensive insights into various aspects. The BSF index, for example, helps in crop management by mitigating soil exposure risks and revealing larger land use changes such as deforestation (Demattê et al., 2020). Combined indices provide a holistic view, *e.g.* with NOS and CDR, along with BSF, detailing crop cycles and identifying intensively cultivated areas (Estel et al., 2016). Coupled with climate and human footprint data, these indices facilitate studies of the interactions between climate, human activities, and environments (Brown et al., 2012; Seifert and Lobell, 2015).

Although our Landsat indices data cube encompasses a wide range of indices, it is less specialized than studies designed to quantify specific processes. For example, BSF, proposed in previous work (Demattê et al., 2020; Mzid et al., 2021), is mapped here for the first time annually across continental Europe as a complete, gap-free time series and has proven useful in modeling experiments. However, it is not a direct representation of soil surface. In contrast, many studies mapping bare soil focus on identifying the "barest" pixel to create a Bare Soil Composite (BSC) (Demattê et al., 2020; Diek et al., 2017; Mzid et al., 2021; Rizzo et al., 2023; Rogge et al., 2018; Safanelli et al., 2020). These studies explored more sophisticated methods to ensure the algorithm accurately identifies truly bare pixels while effectively filtering out non-soil signals, such as vegetation. Typically, this involves a combination of indices and, in some cases, land use masks (Diek et al., 2017; Heiden et al., 2022; Rogge et al., 2018; Safanelli et al., 2020). As a result, these BSC products provide a more direct and reliable representation of soil properties. While its applicability may be limited in regions dominated by permanent vegetation (Hill and Guerschman, 2022) and it is typically produced as a static snapshot, BSC remains a valuable predictor for Digital Soil Mapping (DSM) (Broeg et al., 2024a). Recent studies suggest that combining the strengths of BSF and BSC is feasible. For instance, Zepp et al. (2023) investigated the temporal windows required to generate recurrent BSC products, highlighting the potential to create BSC time series that capture dynamic soil surface properties. Building on this, a natural next step for this study is to apply this approach and generate BSC time series as a new feature, leveraging the extensive Landsat archive at our disposal. Additionally, BSF could be integrated as an indicator to dynamically refine the temporal frame for BSC generation.

Many studies on Land Surface Phenology (LSP), *i.e.*, the descriptors of spatio-temporal patterns of the vegetated land surface derived from EO data (De Beurs and Henebry, 2005), present different methods to quantify crop cycles (Estel et al., 2016; Henebry and De Beurs, 2013; Liu et al., 2020; Meroni et al., 2021). All these methods highlight potential avenues for refining the current CDR and NOS data products in the future. Most LSP studies and products have moderate to coarse spatial resolutions (Caparros-Santiago et al., 2021; Estel et al., 2016; Ganguly et al., 2010; Zhang et al., 2018), due to the trade-off between spatial resolution and revisit frequency (Bolton et al., 2020; Caparros-Santiago et al., 2021; Meroni et al., 2021). Although the Harmonized Landsat Sentinel (HLS) product—created by harmonizing data from Landsat-8/9 and Sentinel-2A/B—offers new opportunities to map LSP indicators at fine spatial resolutions (Bolton et al., 2020), it is only available starting from 2013. Lastly, despite the potential to quantify the intensity of tillage with EO data such as Landsat-7/8/9 and Sentinel-2, the availability of information on tillage practice on a continental scale remains limited.

In summary, the Landsat indices we described in this paper can be seen as part of a comprehensive effort to produce useful biophysical indices to support soil, land, water and land use monitoring. While our methods may be less specialized than those of dedicated thematic studies, this reflects a compromise to balance computational resource demands and the need for complete, gap-free coverage. Despite this, they are designed as a foundational resource, offering a broad and adaptable set of predictors in a comprehensive, gap-free form that requires minimal effort from users for effective application in various environmental modeling and monitoring tasks.

2.4.2 Implications from plausibility checks

Our quality assessment primarily focused on third and fourth-tier predictors, such as minNDTI, BSF, NOS, CDR and their long-term trends, due to their complexity and the availability of relevant ground data. We also visualized the annual P50 of NDVI and NDWI to provide basic information on general vegetation and water conditions. Visual examination indicated that annual NDVI, NDWI, and BSF effectively differentiate various landscapes, while NOS and CDR distinguish different crop patches. Although interpreting these indices without detailed crop management information is challenging, regional statistics of NOS, CDR, and BSF reveal significant differences in crop cycles among different crop types and regions across Europe. We demonstrated that the BSF index effectively measures cropland bareness, with a high negative correlation coefficient of -0.73 with the duration of crop cover in agricultural areas between 2007 and 2016. However, the moderate correlation of minNDTI with tillage practices statistics from EUROSTAT suggests the need for further investigation.

In Fig. 2.5, a negative NDVI trend is observed in regions characterized by high altitude and latitude. This finding appears to be in conflict with the greening phenomenon in

cold / snowy ecosystems, as reported by Obuchowicz et al. (2023). Furthermore, the positive trend observed in the NDWI trend map is inconsistent with Poussin et al. (2021)'s observation of a slight decline in NDWI values in the Alps region of Switzerland during the period from 1984 to 2019. Several factors could be influencing these discrepancies, such as the different temporal scales used for trend calculation, as noted by Jong and Bruin (2012), or the increased dependence on gap-filling due to limited available high-quality EO data in these regions. These contrasting observations call for further investigation to understand the underlying causes and implications of them.

Although these efforts are limited by the availability of relevant ground data and sometimes yield inconclusive results, it is important to note that these indices are proxies for ground objects and processes, rather than exact measurements of ground conditions. They are designed to complement, not replace, ground survey data. Therefore, plausibility checks are not expected to provide one-to-one validation but to highlight the advantages and limitations of these indices. In summary, well-established indices such as NDVI and NDWI, and the relatively straightforward BSF index, are reliable and suitable for direct analysis, as demonstrated in Sect. 2.3.3. However, for more complex indices like NOS and CDR, robust assessments require careful integration with ground-truth data to ensure accuracy and validity.

2.4.3 Insights from modelling experiments

By looking at predictors of different themes' impact on SOC regression and LC classification model, band theme predictors are vital for both models. This is unsurprising, as they form the foundation from which other spectral indices are derived and provide a general representation of surface objects and processes (Zhou et al., 2021). Apart from band theme predictors, models trained on single-theme predictors show that vegetation predictors have a stronger positive impact on both LC and SOC modeling than other themes. Vegetation is fundamental in defining land covers, naturally ranking high in feature importance. Its close relationship with living soil, through through physical, chemical, and biological interactions (Kooch et al., 2022), positions it as a vital*"organism"* factor in the soil formation process and a key driver of SOC change (McBratney et al., 2003; Wiesmeier et al., 2019). This importance is reflected in numerous SOC modeling studies, where vegetation indices like NDVI and EVI consistently rank among the top predictors, sometimes even used as the only spectral index (Gomes et al., 2019; Sothe et al., 2022; Wadoux et al., 2020a; Zeraatpisheh et al., 2022).

Soil and water theme parameters also contribute valuable insights for SOC modeling. The indices of these two themes already differ statistically across most LCs, even before considering LC classification (Szabo et al., 2016). For SOC modeling, tillage practices and soil exposure, as indicated by soil theme predictors, negatively alter soil structure and impact carbon dynamics (Chahal et al., 2021; Demattê et al., 2020; Mzid et al.,

2021; Wiesmeier et al., 2019; Yang et al., 2020b). Similarly, water dynamics and climate, represented by water theme predictors, are key elements of the *climate"* factor in the soil-forming *"scorpan"* model (Maynard and Levi, 2017; Wiesmeier et al., 2019; Zeraatpisheh et al., 2022). When direct climate data, such as precipitation and temperature, are unavailable, water theme indices can serve as useful substitutes.

Combining predictors from multiple themes generally enhances model performance. This is consistent with the findings of Witjes et al. (2023) for LC classification. For SOC modeling, most studies in DSM ensure that all soil-forming factors are incorporated within the *"scorpan"* model (McBratney et al., 2003; Wadoux et al., 2020a; Wiesmeier et al., 2019). Zeraatpisheh et al., 2022 demonstrated that while combining feature groups may not always result in higher prediction accuracy, it does produce SOC predictions with lower uncertainty. The exception is crop theme predictors, likely due to limitations in our methods for quantifying NOS and CDR. Numerous studies have shown that phenological parameters can improve soil property and LC modeling, at least for croplands (He et al., 2021; Maynard and Levi, 2017; Nguyen et al., 2020; Xue et al., 2014; Yang et al., 2020a). However, these studies often focus on small areas with much less data, allowing for localized and detailed extraction of phenological parameters. In contrast, our approach to extracting phenological parameters like NOS and CDR, and assessing their relevance to SOC and LC modeling, remains limited. Further investigation is needed to determine whether phenological parameters can be effectively scaled to large, continental-level studies and, if so, to determine the methods for their application.

Unlike theme-based predictors, predictors of different tiers exhibit varying levels of importance for SOC regression and LC classification modeling. For SOC regression, predictors that represent longer periods are more important, aligning with the findings of Zepp et al., 2023. This is clear in Fig. 2.18, where predictors with extended temporal processing significantly enhance model performance. Table 2.5 shows that model trained solely on tier 4 predictors (long-term characteristics) consistently outperform others across all three metrics. This suggests that long-term characteristics are highly valuable for SOC regression and can sometimes outperform more complex combinations of features. A study by Yang et al. (2022) demonstrates that temporal variables are effective in modeling temporal SOC variation, though their scope spans 30 years, while the LUCAS topsoil survey covers only nine years. Dynamic information appears less beneficial for SOC modeling, which is not surprising since SOC is a slow-changing soil property that typically varies over decades.

In contrast, incorporating detailed dynamic information (tier 1 and tier 2: bimonthly bands and indices) improves the model performance for LC classification. This improvement may be attributed to the dynamic nature of LC, which changes more frequently or suddenly compared to SOC, making it better captured with more frequent revisit information. The temporal dimension has always been critical in LC classification. Researchers have

emphasized incorporating temporal features into models to enable the mapping of LC at finer temporal scales, such as annual intervals (Brown et al., 2022; Gómez et al., 2016; Zhang et al., 2020). Surprisingly, the tier 3 (annual) indices are shown to be less important for LC classification. This could be because the temporal characteristics within a year neither provide the detailed phenological information that bimonthly predictors do (as phenological process aren't defined strictly by a calender year), nor offer the long-term characteristics that tier 4 predictors provide.

Combining predictors from multiple themes adds more value than combining predictors from multiple tiers. This is likely because different tiers mainly vary in their level of temporal feature engineering, conveying similar messages. However, as demonstrated by our SOC regression and LC classification experiments, using more features does not always lead to better performance. This can increase the risk of multicollinearity and overfittin, as well as add complexity and require greater computational resources (Wadoux et al., 2020a). Feature importance analysis highlights that feature selection should be guided by a clear understanding of the predictors' characteristics, modeling objectives, and task constraints. Factors such as the temporal and spatial resolution of covariates, the spatial extent of modeling, and the balance of sample data also affect the determination of the appropriate features (Chen et al., 2022b; Nazari et al., 2020; Shi et al., 2018).

2.4.4 Limitations

As with any product derived from the Landsat ARD V2, users should be aware of inherent limitations. A key step in harmonizing Landsat products into ARD is surface reflectance normalization (Potapov et al., 2020), which differs from physically-based atmospheric and surface anisotropy corrections. Our plausibility checks demonstrate that the bimonthly Landsat reflectance bands (tier-1) are well-suited for index derivation and modeling purposes, indicating sufficient product quality. However, similar to the Landsat ARD V2 data, this data cube may not be suitable for very precise studies of land surface reflectance.

As noted by Potapov et al. (2020), the GLAD Landsat ARD algorithm struggles with wintertime image processing in temperate and boreal climates at high latitudes. This is due to low sun azimuth angles and the similarity between snow cover and clouds during the winter season. As a result, some images and their resulting 16-day composites may still contain snow-covered observations. To address this issue, additional filtering—such as applying the "snow/ice" observation quality flag—was implemented in this study to remove affected observations. However, this process resulted in larger data gaps in certain regions. Consequently, the TSIRF algorithm, which depends on other images in the sequence for reconstruction, can produce inaccurate results when the time series contains too few valid pixels (Consoli et al., 2024). This is also reflected in the reduced time series reconstruction accuracy of the European dataset compared to the global dataset, which

can be attributed to Europe's higher average latitude compared to the global dataset. Consequently, artifacts are observed in high-altitude and high-latitude regions, such as central Sweden and southern Norway, as illustrated in Figures 2.5 and 2.6. While these artifacts do not significantly impact the overall analysis, users should be aware of them and exercise caution when interpreting results, especially in regions with challenging data conditions.

As with many Land Surface Phenology (LSP) studies, our temporally-aggregated indices such as BSF, NOS, and CDR rely on threshold-based methods (De Beurs and Henebry, 2010). These thresholds can vary depending on crop types and regional characteristics, often requiring expert knowledge for accurate application, especially crop relevant indices (Chen et al., 2016). Given the broad spatial and temporal scope of our data cube, we adopted a generalized threshold approach, which might not capture local variations accurately (Dematthê et al., 2020; Estel et al., 2016; Henebry and De Beurs, 2013). The limitation of this method lies in its inability to distinguish between meaningful changes and natural fluctuations exceeding the threshold (De Beurs and Henebry, 2010). Therefore, users are advised to proceed cautiously and conduct preliminary tests when a meaningful conclusion needs to be drawn from direct analysis with this limited local or domain knowledge.

Another challenge posed by advances in EO technology is the expanding list of satellite spectral indices, each uniquely calibrated for specific environmental attributes. For example, the Awesome Spectral Index (ASI), developed by Montero et al., 2023, lists 127 vegetation indices by its publication date. Incorporating all of these indices into a single data cube is impractical. To balance data volume and complexity, we adopted a hybrid approach. The predictors data cube includes the indices most commonly applied, covering several important aspects of the environment, such as vegetation health, soil conditions, and water dynamics. Additionally, developed in alignment with the principles of the Open Data Cube (ODC), our ARCO data cube includes guided documentation and scripts to assist users in querying, exploring, performing analyses, and tracking the provenance of all contained data. This approach facilitates quality control and updates, allowing users to customize the data cube according to their specific needs or to explore new indices that could offer additional insight into the environment.

The data cube includes a wide array of indices with various applications. However, some indices may yield irrelevant values outside their intended regions. For example, many indices lose significance in snowy environments at high latitudes and altitudes, and minNDTI does not convey useful information about tillage practices outside of arable land. It is crucial not to apply these indices indiscriminately across all land covers. Other indices within the data cube can help delineate appropriate application areas. For example, the NDSI can exclude snowy areas, ensuring a more accurate application of other indices.

2.4.5 Future work

While NOS and CDR provide useful insights into crop cycles and their duration, their limited contribution to SOC regression models suggests a need for further refinement. Improving the methods for quantifying these indices or exploring alternative indicators might enhance their utility in SOC modeling. As discussed in Section 2.4.1, there is significant value in learning from specialized studies that focus on accurately quantifying specific processes. Future updates to this work will incorporate more localized and domain-specific knowledge could help to tailor these indices to better capture relevant environmental processes.

Although the original Landsat archive has a 16-day temporal revisit, we aggregated the data set into a bimonthly resolution to address quality issues such as gaps and cloudy pixels. Although bimonthly resolution effectively captures annual dynamics, a finer 16-day resolution could provide richer details, especially for NDTI values observed from harvest windows crucial for mapping tillage practices (Zheng et al., 2012). However, finer temporal resolution requires extensive gap-filling, presenting a significant trade-off between temporal detail and data consistency.

Our data set currently relies exclusively on Landsat ARD V2 to derive NOS, CDR, and minNDTI. However, integrating multiple sensors could enhance accuracy (Caparros-Santiago et al., 2021; Henebry and De Beurs, 2013). For example, using minimum surface temperature thresholds can prevent false crop cycle identifications (Li et al., 2014), and integrating precipitation data can improve crop residue estimates (Beeson et al., 2020; Dvorakova et al., 2023). Furthermore, combining Harmonized Landsat-Sentinel (HLS) data could further improve index usability, especially for years covered by Sentinel-2 (Wu et al., 2022).

As indicated by the survey results of Wagemann et al., 2021, improving the usability of spectral indices requires the participation of various stakeholders, including farmers, policymakers, and sector-specific experts. Their local knowledge and domain-specific insight can refine algorithms and improve the utility of data. This approach aligns with the objectives of the Horizon Europe AI4SoilHealth project (2023–2027), aiming to bridge the gap between EO data and practical applications.

2.5 Conclusions

In this paper, we present the production process and quality assessment of a 17TB ARCO data cube, primarily based on the Landsat indices. It is available at a 30 m resolution from 2000 to 2022, comprising multiple spectral indices. Multi-tiers of preditors are derived to represent different levels of temporal information extraction: bimonthly, annual, and long-term analyses covering continental Europe, including Ukraine, the UK, and Turkey. The main advantages of our dataset include: (1) its fine spatial resolution and extensive

temporal coverage, which are crucial for analyzing and tracking the spatial variations and temporal dynamics of different environmental aspects, such as soil; (2) its broad coverage of indices, each providing unique insights into different aspects, together offering a holistic view of environment; (3) it is analysis-ready for modeling and mapping applications.

To maximize the effectiveness and reliability of the data, the following usage recommendations are proposed:

- **Established Indices**: For well-established indices like NDVI, NDWI, SAVI, FAPAR, NDSI, and those demonstrating high reliability such as BSF, the data cube provides valuable information and insights, especially when ground truth data are lacking.
- **Complex Indices**: For more complex indices that were not validated in this study due to the lack of corresponding ground truth data, caution is advised. These should be used alongside ground truth data to ensure accuracy and validity, and their application should consider domain and local knowledge.
- **Modeling**: All indices in the data cube have proven effective as predictors for modeling purposes, making them useful for various environmental and ecological models.
- **Feature Selection**: It is beneficial to perform feature selection to identify the best combinations of indices from the comprehensive predictor sets for specific modeling tasks, optimizing the model's performance and relevance to the study's objectives.

Data availability

This dataset is registered with a valid DOI (Tian et al., 2024, https://doi.org/10.5281/zenodo.10776891). While each data bucket has a unique DOI, we kindly ask users to cite the DOI on the Zenodo landing page to ensure consistency. This DOI represents all dataset versions and always resolves to the latest release.

Due to Zenodo's storage limitations and the large size of the dataset, only portions of the data are stored in Zenodo, distributed across multiple buckets. To streamline access, we have created a Zenodo landing page that includes a navigation catalog linking to the individual buckets. Additionally, a central access catalog, provided as a Google Sheet, offers complete links for public access to the data stored in Wasabi Cloud Storage. Both the Zenodo landing page and the central catalog include a detailed explanation of the naming conventions used.

To lower the barrier to data use, metadata is embedded directly into file names. Each dataset layer follows a standardized naming format, structured into 10 key fields: generic variable name, variable procedure combination, position in the probability distribution or variable type, spatial support, depth reference, time reference (including start and

end times), bounding box, EPSG code, and version code. Each metadata field serves a specific purpose in assessing the datasets' fitness for use. For the datasets presented in this study, the metadata specifies a uniform spatial support of 30 m resolution, a depth reference denoted as *s* (land surface), a bounding box identified as *eu*, and an EPSG code of *EPSG:3035*. For the other fields: the generic variable name helps users identify the required predictor layer; the variable procedure combination provides indications of how the data was derived and its source; the time reference, comprising the start and end dates, ensures users can select layers matching their relevant temporal scope; and the version code, which reflects the creation date of the corresponding layer, facilitates tracking and version control. This metadata framework enhances geodata trustworthiness and usability by allowing users to assess dataset relevance and suitability directly from the name. This metadata framework enhances geodata trustworthiness and usability by allowing users to assess dataset relevance and suitability directly from the name.

Code availability

Complete code used to derive all indices and generate the visualization figures is available at https://doi.org/10.5281/zenodo.12907281. Even more spectral indices can be derived using the https://github.com/awesome-spectral-indices/ library provided by Montero et al. (2023).

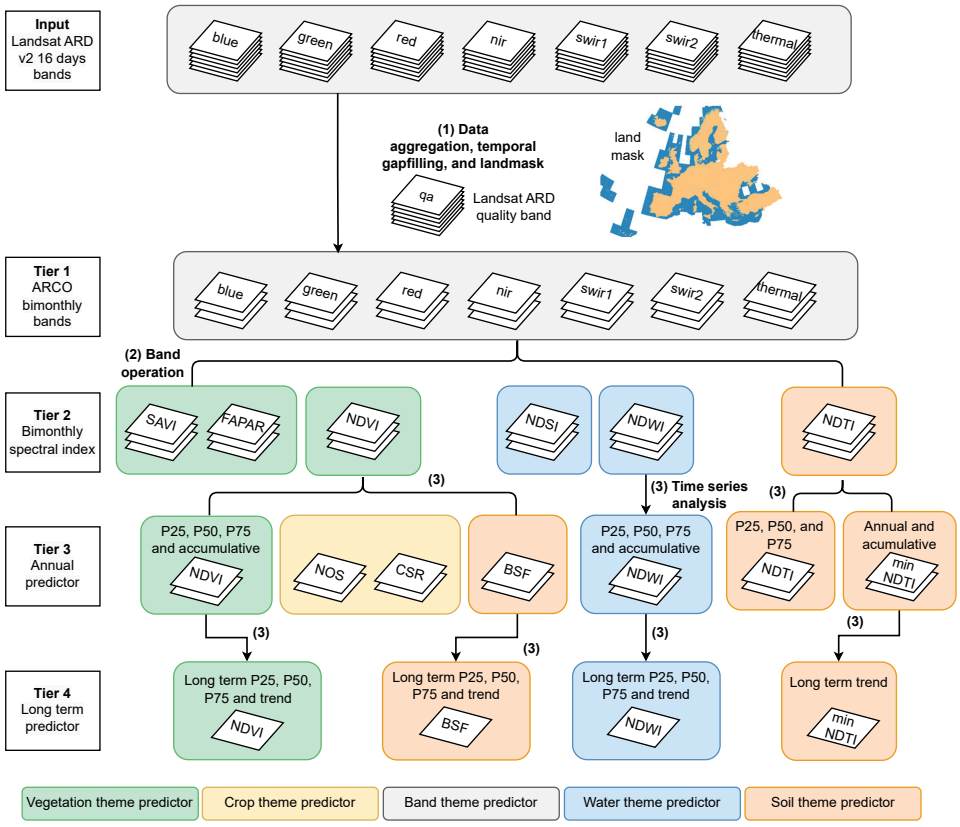


Figure 2.1: General workflow for processing Landsat-based spectral index predictors. They formed four different tiers based on their level of processing and different temporal resolutions through the workflow. The predictors could also be categorized into five thematic groups, each framed by different colors in the figure, including vegetation, water, band-specific properties, crops, and soil characteristics. The "(3) Time series analysis" in the workflow incorporates three different temporal operations: temporal aggregation, which extracts statistical measures over specific periods to represent those intervals; trend analysis, which evaluates the directional changes of a predictor over time; and cumulative analysis, which sum up the corresponding annual predictor values from a starting baseline year (*i.e.* 2000) to the specified index year. Land-mask source: https://doi.org/10.5281/zenodo.8171860.

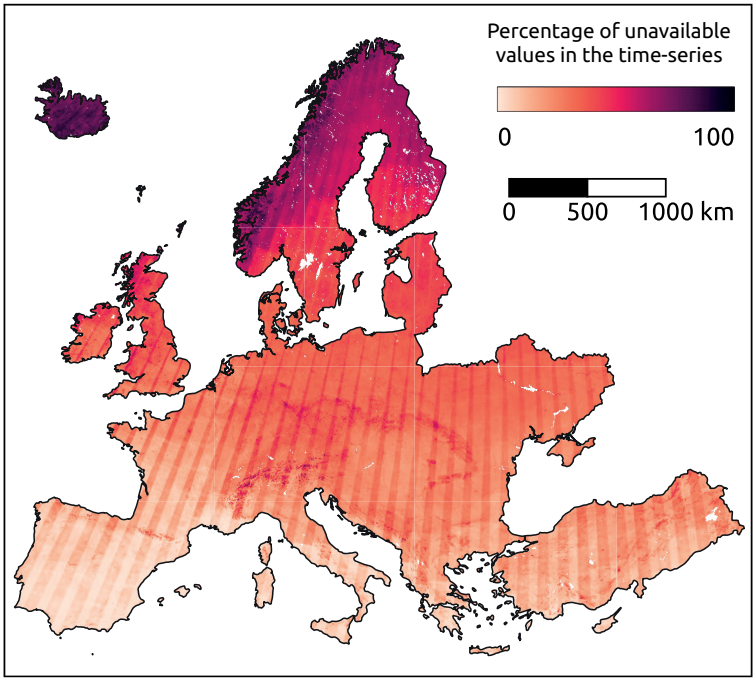


Figure 2.2: Per-pixel count of available values in the bimonthly aggregated Landsat time-series from 1997 to 2022. Darker areas are more affected by presence of data-gaps in the time-series. In addition to cloud-presence and snow-cover, it is possible to notice patterns determined by overlapping scenes in the original Landsat raw images. Adapted from Consoli et al. (2024).

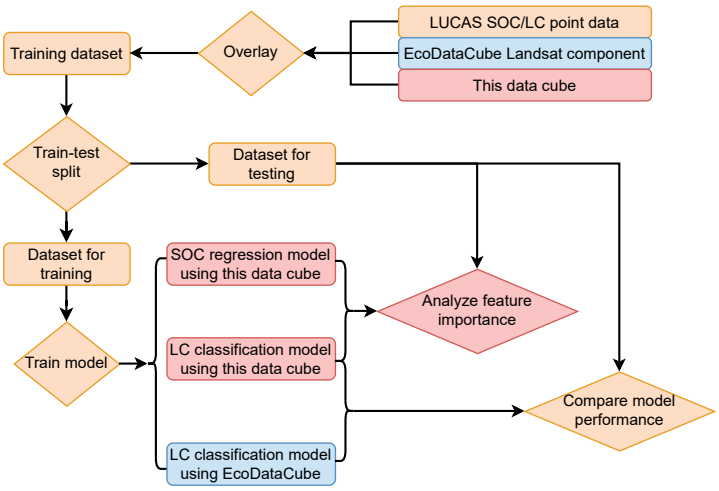


Figure 2.3: Workflow for modeling experiments, illustrating the comparison between EcoDataCube and the proposed data cube in the LC classification experiment, as well as the feature importance analysis conducted for the LC classification model and the SOC regression model using the proposed data cube. Components specific to EcoDataCube are marked in blue, those unique to the data cube produced in this work are marked in red, and shared general components are marked in yellow.

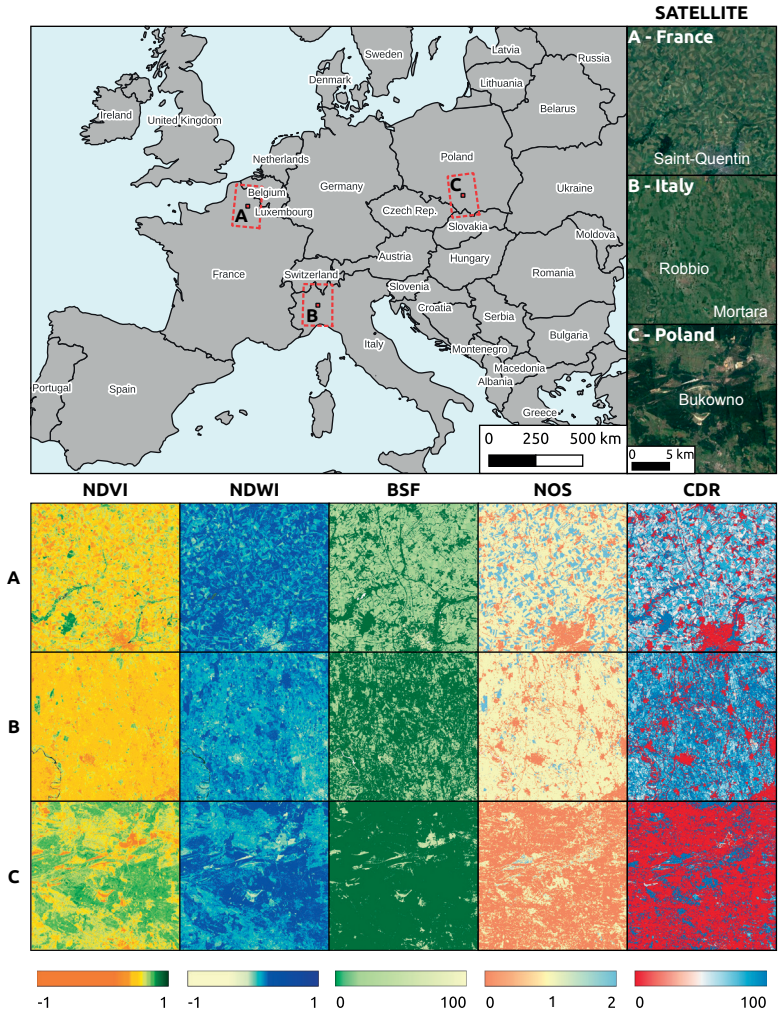


Figure 2.4: Zoomed-in examples of NDVI and NDWI median in the year 2018 at (A) Northern France, with the city of Saint-Quentin at south-east corner; (B) Northern Italy, in the western Po Valley, east of Vercelli; (C), Southern Poland, within the Przemsza River basin, featuring the Szczakowa sand mine. Each image tile has a side length of 20km. Satellite imagery source: © Google Earth (https://earth.google.com/web/).

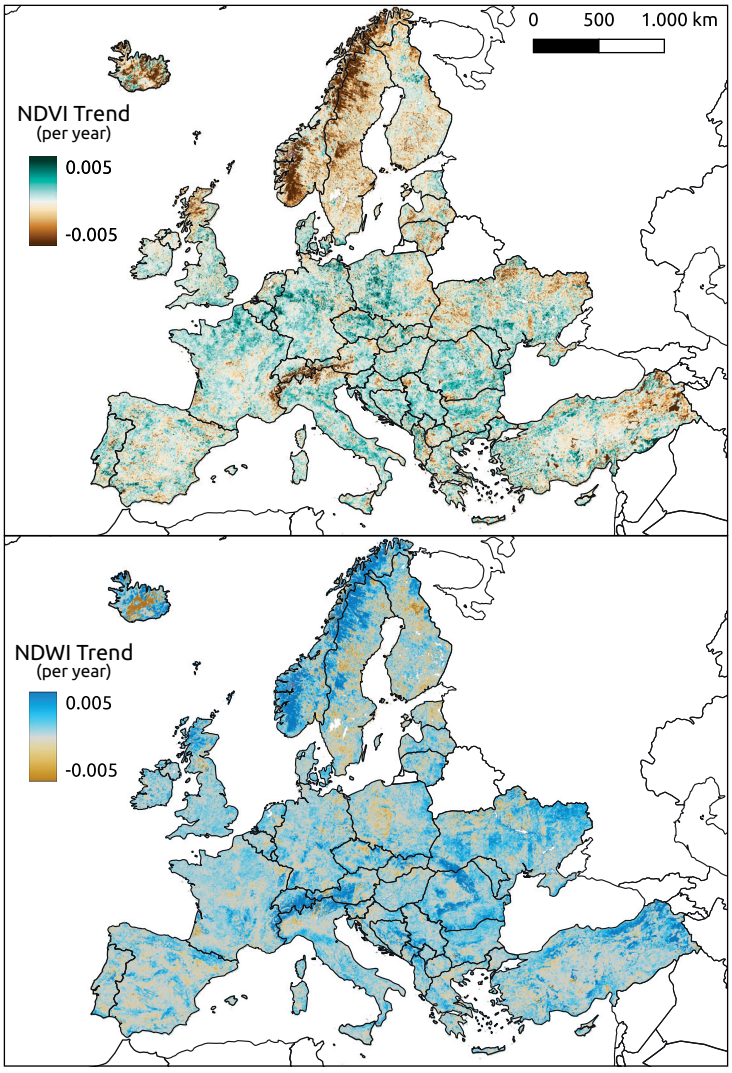


Figure 2.5: Long term changing trend between 2000–2022 of NDVI (top panel) and NDWI (bottom panel), as the annual rate of change in index values.

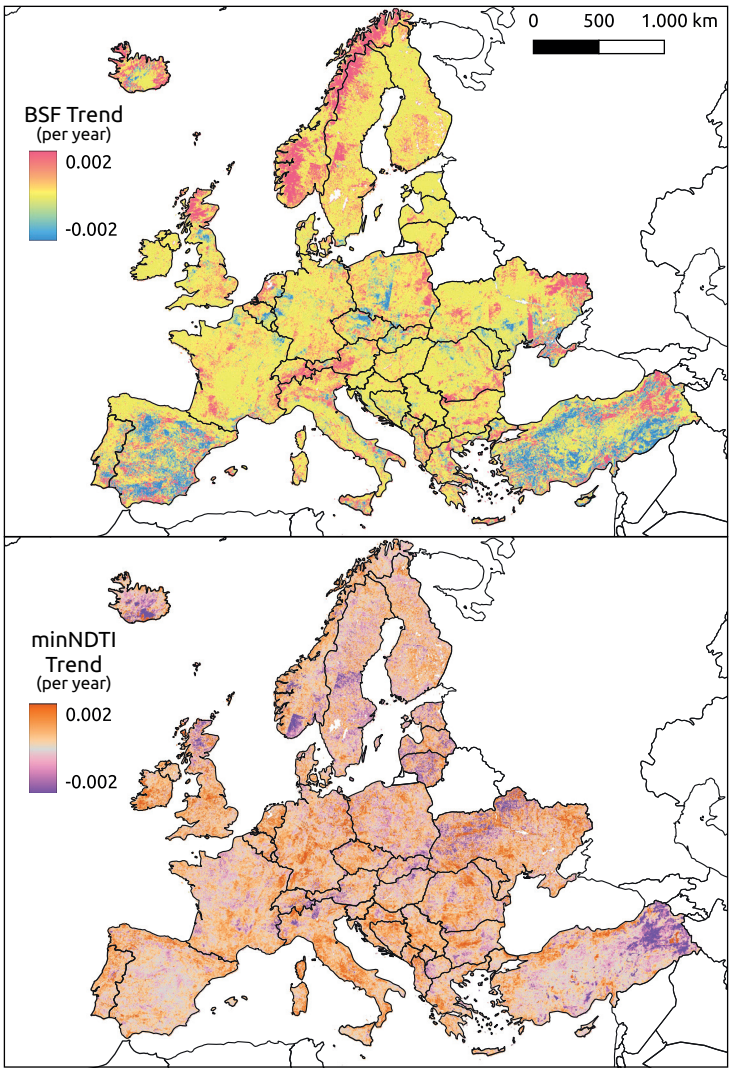


Figure 2.6: Long term changing trend between 2000–2022 of BSF (top panel) and minNDTI (bottom panel), as the annual rate of change in index values.

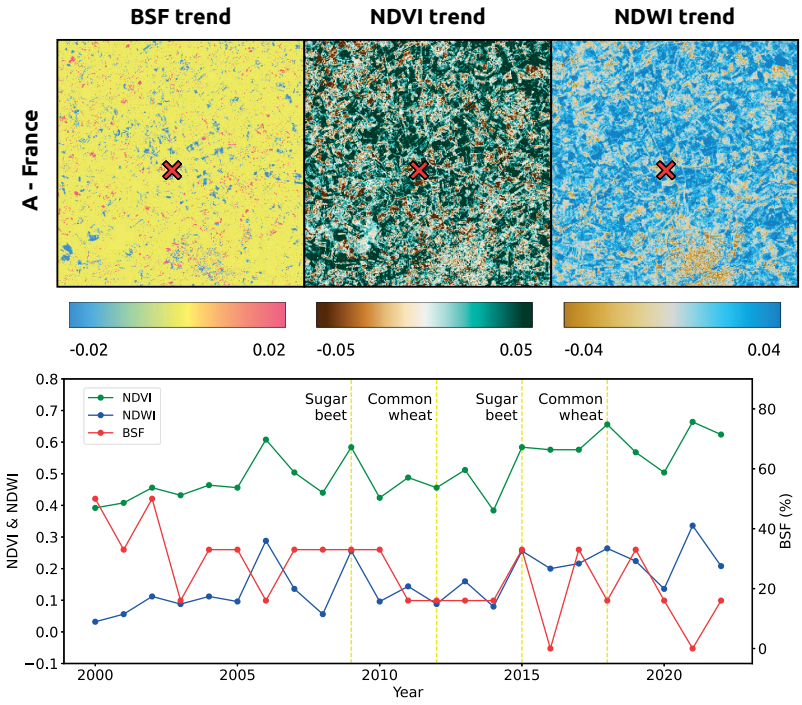


Figure 2.7: The top panel provides a zoomed-in view of NDVI, NDWI and BSF trends for Saint-Quentin region, north France (region A, identical in spatial location and extent to region-A of Fig. 2.4), which has a consistent agricultural use with varies crop types across years. Images generated by the authors using data produced in this study. Red crosses in the top panel mark the LUCAS point (ID: 38363000; lat: 49.91122N, lon: 3.237571E), detailed in the bottom panel. The bottom panel shows the time series of NDVI, NDWI, and BSF from 2000 to 2022 at this LUCAS point, which is revisited four times in the years 2009, 2012, 2015, and 2018, with the surveyed land cover indicated beside yellow vertical lines.

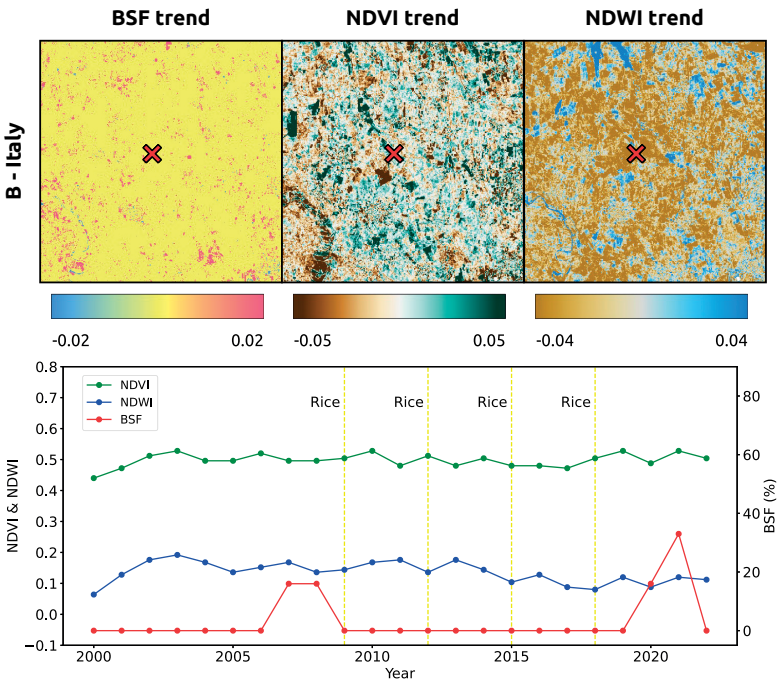


Figure 2.8: The top panel provides zoomed-in view of NDVI, NDWI and BSF trends for rural area besides city Vercelli in north Italy, with dominant rice cultivation (region B, identical in spatial location and extent to region-B of Fig. 2.4). Images generated by the authors using data produced in this study. Red crosses in the top panel mark the LUCAS point (ID = 42122479; lat=45.33338N, lon=8.611634E), detailed in the bottom panel. The bottom panel shows the time series of NDVI, NDWI, and BSF from 2000 to 2022 at this LUCAS point, which is revisited four times in the years 2009, 2012, 2015, and 2018, with the surveyed land cover indicated beside yellow vertical lines.

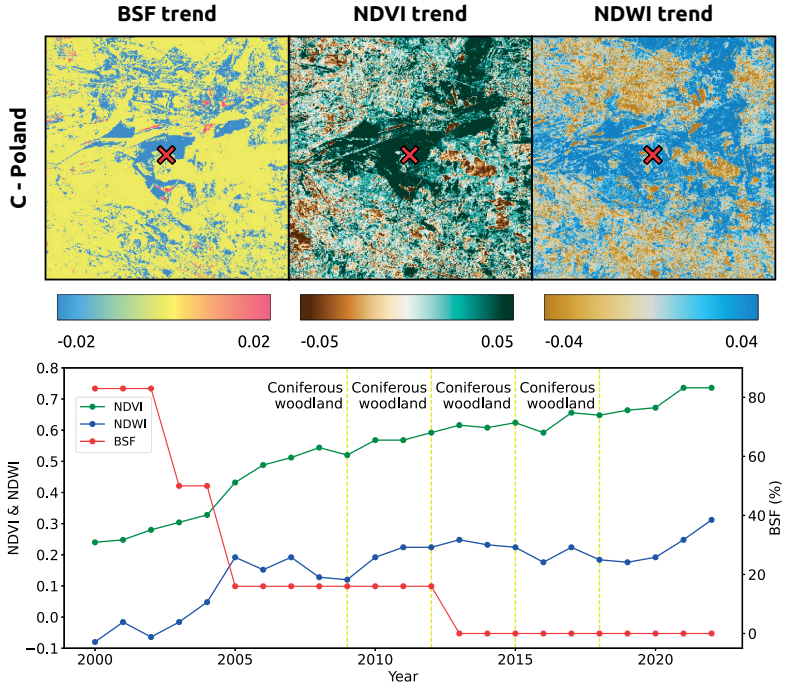


Figure 2.9: The top panel provides zoomed-in view of NDVI, NDWI and BSF trends for the former Szczakowa sand mine, south Poland (region C, identical in spatial location and extent to region-C of Fig. 2.4). Images generated by the authors using data produced in this study. Red crosses in the top panel mark the LUCAS point, detailed in the bottom panel (ID=49923058; lat=50.24478N, lon=19.435963E). The bottom panel shows the time series of NDVI, NDWI, and BSF from 2000 to 2022 at this LUCAS point, which is revisited four times in the years 2009, 2012, 2015, and 2018, with the surveyed land cover indicated beside yellow vertical lines.

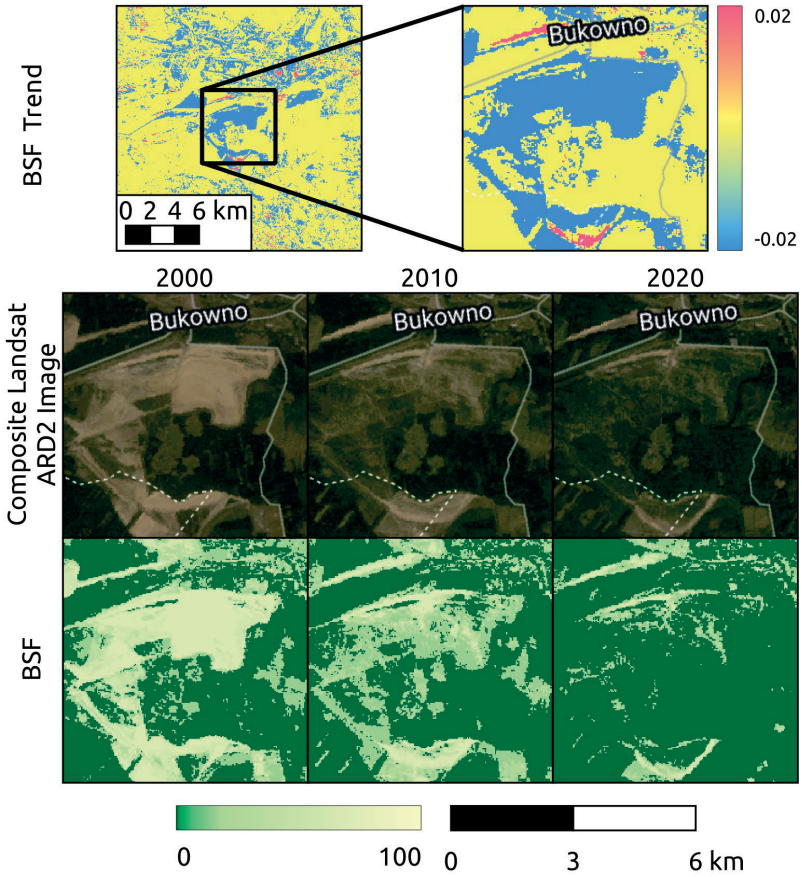


Figure 2.10: Time series comparison showing zoomed-in satellite and BSF images of reforested post-sandmine areas near Bukowno town for the years 2000, 2010, and 2020. On the top left are the BSF trend image, the black frame shows the zoomed-in area of which the satellite and BSF images are shown. The top left image is the same as in Fig. 2.9. BSF images generated by the authors using data produced in this study.The Composite Landsat ARD2 Image was generated by the authors using data from (Potapov et al., 2020), as described in Sect 2.2.1.

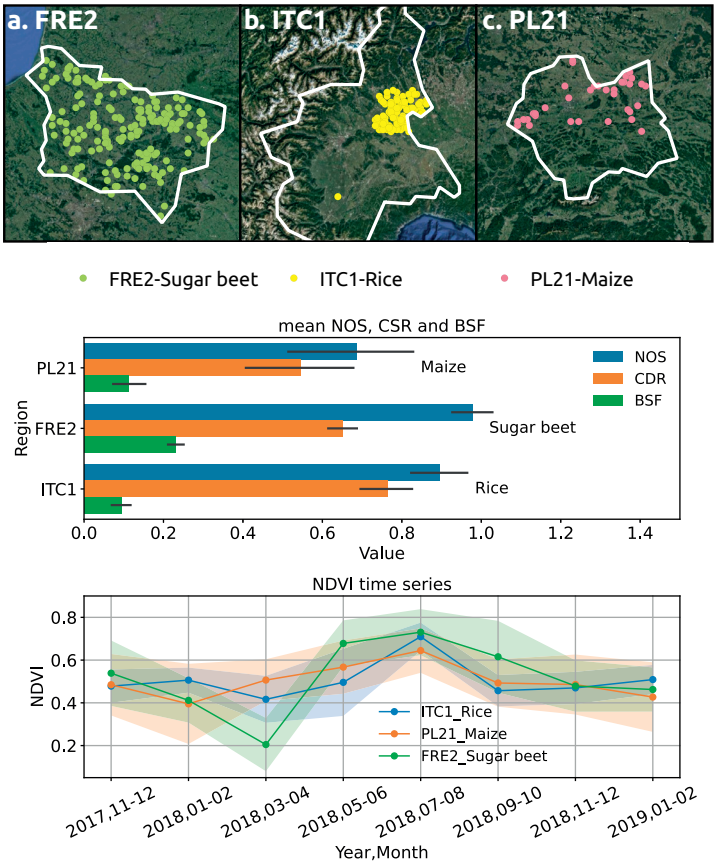


Figure 2.11: Illustration of the locations of LUCAS 2018 survey points of selected crop types (top panel); statistics of various indices calculated from these LUCAS 2018 points across crops in the selected regions (middle panel); and average NDVI dynamics for the year 2018, including adjacent values from neighbouring years (bottom panel). The LUCAS 2018 points data are from 3 NUTS2 regions: (a) the central part of the Picardy region (FRE2) in north France; (b) the western part of the Piedmont province (ITC1) in Italy; and (c) the Malopolska Province (PL21) in southern Poland. These 3 regions are close to or partially overlap with the areas shown in Fig. 2.4, thus they are denoted with lowercase letters instead of capital ones to show the difference and similarity. Satellite imagery source: © Google Earth (https://earth.google.com/web/).

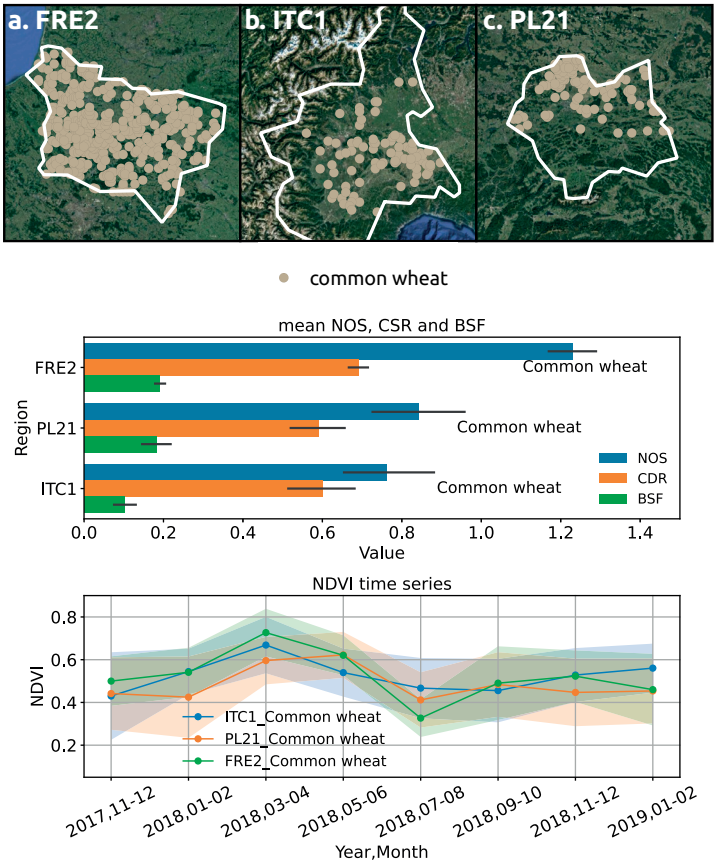


Figure 2.12: Illustration of the locations of LUCAS 2018 survey points of common wheat (top panel); statistics of various indices calculated from these LUCAS 2018 points of common wheat across regions (middle panel); and average NDVI dynamics for the year 2018, including adjacent values from neighbouring years (bottom panel). The LUCAS 2018 points data are from the same 3 NUTS2 regions as in Fig. 2.11. Satellite imagery source: © Google Earth (https://earth.google.com/web/).

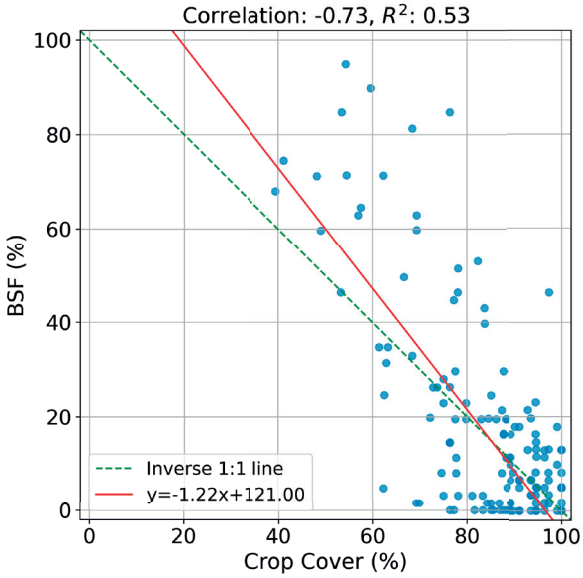


Figure 2.13: Relationship between crop cover proportion and BSF from 2007 to 2016, depicted with an inverse correlation measured by Pearson's product moment correlation coefficient and coefficient of determination R^2.

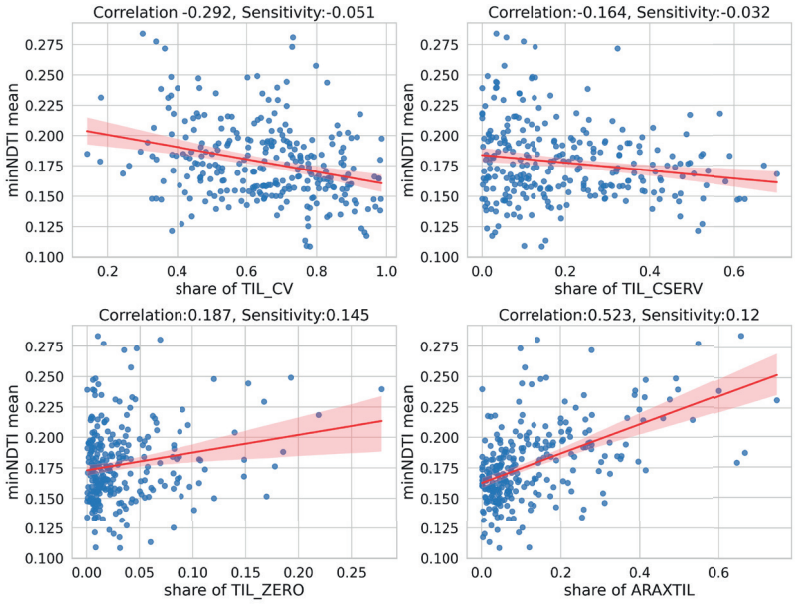


Figure 2.14: Relationship between individual share of tillage practices and average minNDTI of NUTS2 regions recorded in Eurostat. Each point represents a NUTS2 region. Four types of tillage practices shown are TIL_CV: conventional tillage; TIL_CSERV: conservation tillage; TIL_ZERO: zero tillage; ARAXTIL: land excluding from tillage survey. Coefficient refers to the Pearson correlation coefficient, quantifying the linear correlation between two variables. Sensitivity measures how changes in tillage practices influence minNDTI, obtained as a regression slope coefficient from the OLS model fitting.

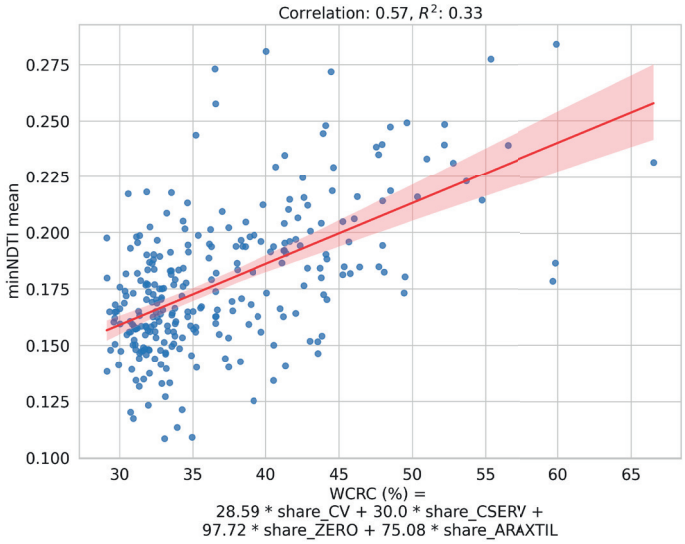


Figure 2.15: Relationship between weighted crop residue cover (WCRC) and mean minNDTI values of NUTS2 regions recorded in Eurostat. Each point stand for a NUTS2 region. Below the figure, the equation for calculating WCRC is presented, where the coefficients correspond to the fitted CRC for each type of tillage practice. Four types of tillage practices considered are TIL_CV: conventional tillage; TIL_CSERV: conservation tillage; TIL_ZERO: zero tillage; ARAXTIL: land excluding from tillage survey.

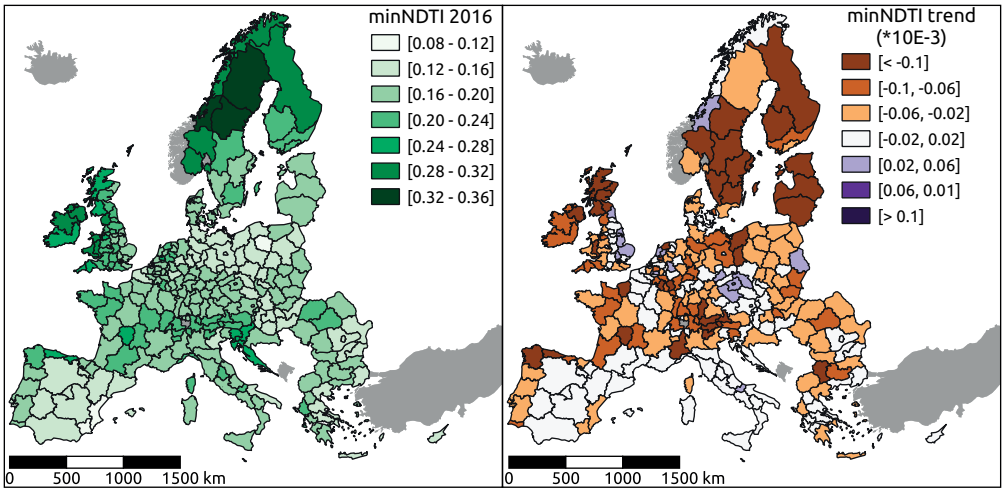


Figure 2.16: Average minNDTI values in the year 2016 (left panel) and average minNDTI trend values (right panel) by NUTS2 regions.

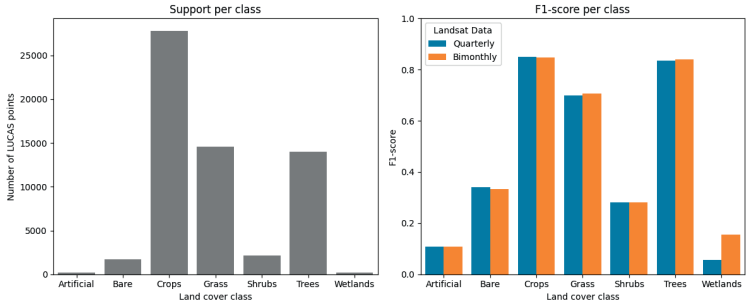


Figure 2.17: Comparison of classification results between Landsat data aggregated to 4 quarters and 3 percentiles per quarter (EcoDataCube Witjes et al., 2023) and the data from this work, aggregated to 6 bimonthly variables per band, as well as 3 annual percentiles. Note that overall accuracy was similar, except for the 'Wetlands' class, which was classified more accurately by the model trained on the bimonthly and annual data.

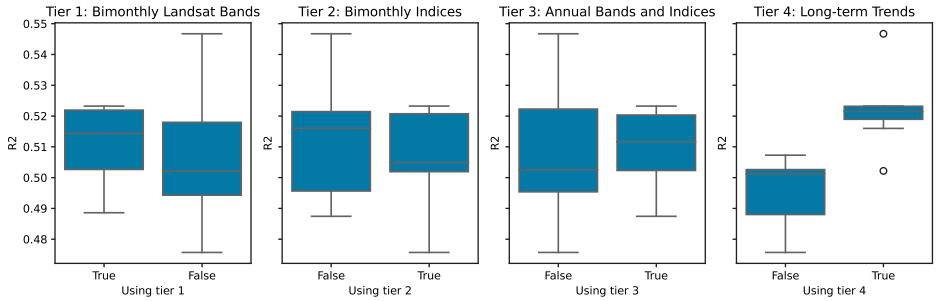


Figure 2.18: Accuracy comparison of SOC regression models trained on 15 combinations of 4 tiers of predictors.

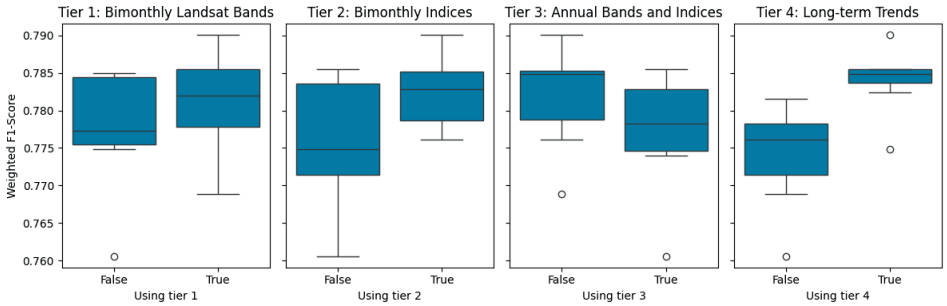


Figure 2.19: Accuracy comparison of LC classification models trained on 15 combinations of 4 tiers of predictors.

Table 2.7: Metrics for SOC regression using different combinations of predictors based on themes.

Number of themes combined	Feature combination	Feature number	CCC	RMSE	R2
1	band	63	0.635	0.652	0.488
	crop	2	0.061	0.898	0.028
	soil	17	0.582	0.690	0.426
	vegetation	27	0.594	0.680	0.443
	water	20	0.591	0.681	0.441
2	band+crop	65	0.636	0.651	0.489
	band+soil	80	0.649	0.642	0.504
	band+vegetation	90	0.651	0.639	0.508
	band+water	83	0.651	0.639	0.508
	crop+soil	19	0.581	0.691	0.425
	crop+vegetation	29	0.593	0.680	0.442
	crop+water	22	0.590	0.682	0.440
	soil+vegetation	44	0.631	0.655	0.483
	soil+water	37	0.626	0.658	0.479
	vegetation+water	47	0.618	0.661	0.474
3	band+crop+soil	82	0.650	0.641	0.505
	band+crop+vegetation	92	0.650	0.640	0.507
	band+crop+water	85	0.651	0.639	0.507
	band+soil+vegetation	107	0.659	0.634	0.516
	band+soil+water	100	0.660	0.634	0.516
	band+vegetation+water	110	0.658	0.634	0.516
	crop+soil+vegetation	46	0.630	0.656	0.482
	crop+soil+water	39	0.625	0.659	0.477
	crop+vegetation+water	49	0.621	0.659	0.476
	soil+vegetation+water	64	0.646	0.643	0.502
4	band+crop+soil+vegetation	109	0.660	0.634	0.516
	band+crop+soil+water	102	0.660	0.633	0.517
	band+crop+vegetation+water	112	0.657	0.634	0.516
	band+soil+vegetation+water	127	**0.666**	**0.628**	**0.524**
	crop+soil+vegetation+water	66	0.645	0.643	0.502
5 (all)	all themes	129	**0.666**	**0.628**	**0.524**

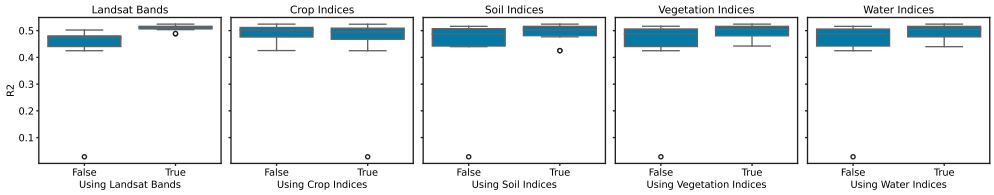


Figure 2.20: Accuracy comparison of SOC regression models trained on 31 combinations of 5 themes of predictors.

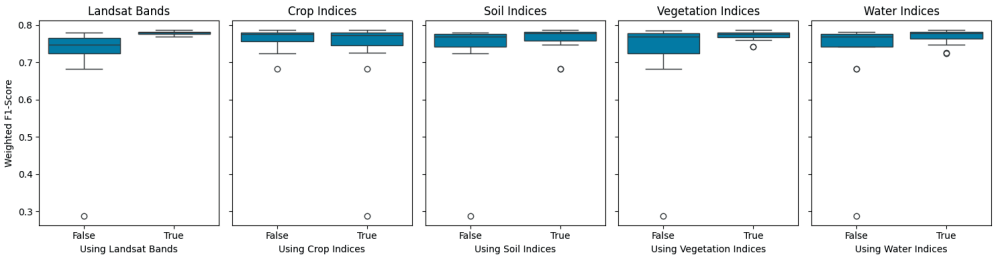


Figure 2.21: Accuracy comparison of LC classification models trained on 31 combinations of 5 themes of predictors.

Table 2.8: Metrics for LC classification using different combinations of predictors based on themes.

Number of themes combined	Themes	Feature number	F1-score	Precision	Recall
1	band	63	0.769	0.774	0.784
	crop	2	0.288	0.308	0.458
	soil	17	0.683	0.685	0.700
	vegetation	27	0.741	0.744	0.756
	water	20	0.724	0.727	0.739
2	band+crop	65	0.771	0.776	0.785
	band+soil	80	0.778	0.783	0.792
	band+vegetation	90	0.775	0.780	0.789
	band+water	83	0.777	0.781	0.790
	crop+soil	19	0.683	0.685	0.700
	crop+vegetation	29	0.741	0.745	0.756
	crop+water	22	0.725	0.730	0.741
	soil+vegetation	44	0.762	0.764	0.775
	soil+water	37	0.749	0.753	0.762
	vegetation+water	47	0.769	0.773	0.781
3	band+crop+soil	82	0.778	0.784	0.792
	band+crop+vegetation	92	0.774	0.780	0.788
	band+crop+water	85	0.776	0.781	0.789
	band+soil+vegetation	107	0.781	0.785	0.794
	band+soil+water	100	0.784	0.788	0.797
	band+vegetation+water	110	0.780	0.785	0.793
	crop+soil+vegetation	46	0.760	0.763	0.773
	crop+soil+water	39	0.747	0.751	0.761
	crop+vegetation+water	49	0.769	0.774	0.782
	soil+vegetation+water	64	0.779	0.782	0.791
4	band+crop+soil+vegetation	109	0.781	0.785	0.794
	band+crop+soil+water	102	0.784	0.789	0.797
	band+crop+vegetation+water	112	0.780	0.785	0.793
	band+soil+vegetation+water	127	**0.786**	**0.790**	**0.799**
	crop+soil+vegetation+water	66	0.778	0.782	0.791
5 (all)	all themes	129	**0.786**	**0.790**	**0.799**

Chapter 3

High-resolution 3D+T mapping of Soil Organic Carbon density

This chapter is based on:

X. Tian, S. de Bruin, R. Simoes, M. S. Isik, R. Minarik, Y.-F. Ho, M. Şahin, M. Herold, D. Consoli, and T. Hengl (2025a). "Spatiotemporal prediction of soil organic carbon density in Europe (2000–2022) using earth observation and machine learning". *PeerJ* 13, e19605. DOI: 10.7717/peerj.19605

Abstract

The paper describes a comprehensive framework for Soil Organic Carbon Density (SOCD) (kg/m^3) modeling and mapping, based on spatiotemporal Random Forest (RF) and Quantile Regression Forests (QRF). A total of 45,616 SOCD observations and various Earth Observation (EO) feature layers were used to produce 30 m SOCD maps for the EU at four-year intervals (2000–2022) and four soil depth intervals (0–20 cm, 20–50 cm, 50–100 cm, and 100–200 cm). Per-pixel 95% probability prediction intervals (PIs) and extrapolation risk probabilities are also provided. Model evaluation indicates good overall accuracy ($R^2 = 0.63$ and $CCC = 0.76$ for hold-out independent tests). Prediction accuracy varies by land cover, depth interval and year of prediction with accuracy the worst for shrubland and deeper soils 100–200 cm. PI validation confirmed effective uncertainty estimation, though with reduced accuracy for higher SOCD values. Shapley analysis identified soil depth as the most influential feature, followed by vegetation, long-term bioclimate, and topographic features. While pixel-level uncertainty is substantial, spatial aggregation reduces uncertainty by approximately 66%. Detecting SOCD changes remains challenging but offers a baseline for future improvements. Maps, based primarily on topsoil data from cropland, grassland, and woodland, are best suited for applications related to these land covers and depths. We recommend that users interpret the maps in conjunction with local knowledge and consider the accompanying uncertainty and extrapolation risk layers. All data and code are available under an open license at https://doi.org/10.5281/zenodo.13754343 and https://github.com/AI4SoilHealth/SoilHealthDataCube/.

3.1 Introduction

Organic carbon stored in soils is of increasing interest to policy makers as a key indicator of land productivity and a potential global solution to offset greenhouse gas emissions from agriculture and combat climate change (Friedlingstein et al., 2014; Jackson et al., 2017; Lamb et al., 2016; Panagos et al., 2024a; Smith et al., 2020). This storage is quantified as Soil Organic Carbon Density (SOCD, kg/m^3), which represents the total amount of organic carbon in a given volume of soil. SOCD is sometimes referred to as *"SOC stocks"* (in kg/m^2 or t/ha) when specified to a particular depth interval. It also serves as a key indicator for the Land Degradation Neutrality initiative under the United Nations Convention to Combat Desertification (Mattina et al., 2018). Given its critical role, there is a growing demand for up-to-date, dynamic, and spatially continuous information of SOCD (De Rosa et al., 2024; Keel et al., 2021; National Academies of Sciences, Engineering, and Medicine, 2021).

Recent advances in data-driven Digital Soil Mapping (DSM), especially leveraging Machine Learning (ML) to handle complex non-linear relationships and the increasing availability of EO data, have significantly improved SOCD modeling (Hengl and MacMillan, 2019; Tziolas et al., 2024; Veronesi and Schillaci, 2019; Wadoux et al., 2020b). de Brogniez et al. (2015) generated a 250 m EU-wide Soil Organic Carbon (SOC) map as a single snapshot, while Poeplau et al. (2020c) and Chen et al. (2023) assessed topsoil carbon stocks for single time periods in Germany and France, respectively. To capture SOC dynamics, temporal variables have been added to DSM, including trends in environmental factor time series (Yang et al., 2022), decayed NDVI (Heuvelink et al., 2020), and time-series EO data (Tayebi et al., 2021). These advances have extended SOC mapping over temporal scale. For example, Szatmári et al. (2021) estimated SOC stock changes in Hungary between 1992 and 2010 using 100 m SOC stock maps, while the World Soils project (Wesemael et al., 2024) produced 100 m resolution EU SOC maps for periods after 2018. Beyond Europe, Ugbaje et al. (2024) mapped SOC stocks in Australia at 90 m resolution for 1990 and 2018, MapBiomas (2023) and Sousa et al. (2024) provided annual topsoil SOC stock predictions for Brazil at 30 m resolution from 1985 to 2021, and Venter et al. (2021b) modeled SOC stock long-term averages and trends for Southern Africa spanning 1984–2019. As a critical dimension of soil, depth has recently been incorporated into DSM. Following a spatio-temporal interpolation approach for soil properties tested at a small scale (Gasch et al., 2015), Helfenstein et al. (2024a) produced detailed 25 m three dimensional space and time (3D+T) SOM maps for the Netherlands, spanning 1953–2023. With these advancements, the availability of mapping products has expanded, evolving from static spatial (2D) maps to time-series maps (2D+T) and incorporating depth (3D+T).

However, generating high-resolution maps often comes at the expense of broader spatial coverage or temporal resolution due to computational constraints, making large-scale high-resolution SOCD maps relatively uncommon. Although the Land Use/Cover Area Frame

Statistical Survey (LUCAS) soil monitoring project (Orgiazzi et al., 2018) has significantly expanded data availability, to our knowledge, no high-resolution SOCD dataset currently covers the European continent over a long period (> 10 years), i.e. which allows for a long-term trend assessment.

In this paper, we provide complete and consistent predictions of SOCD across pan-EU using a reproducible and updatable automated soil mapping framework. This is achieved by integrating more than 45,000 reference SOCD measurements — harmonized from LUCAS and national legacy soil datasets — with a diverse range of environmental features, particularly time-series EO data at 30 m resolution (Tian et al., 2025b), and applying dynamic SOCD mapping in 3D+T. In addition to detailing the production of these maps, we include a thorough validation of the model's performance across different land covers, soil depth intervals, and years. We also use Shapley values to gain deeper insight into how environmental features influence model predictions, providing a better understanding of the model's behavior and underlying drivers (Wadoux et al., 2020b).

Our methodology follows four main steps:

1. Develop a framework for modeling SOCD that optimizes performance while efficiently using computational resources, utilizing data collected and harmonized from various sources across Europe;
2. Produce a time series of 30 m SOCD maps for continental Europe at different depth intervals from 0 to 200 cm between the years 2000 and 2022. Additionally, generate quality indicator maps, including quantified uncertainty and extrapolation risk probability.
3. Explore the influence of environmental factors on SOCD predictions through explainable ML techniques.
4. Assess the accuracy and reliability of the models and maps across different spatial supports and assess their suitability to detect temporal changes in SOCD.

The data sets and maps produced are publicly available at https://doi.org/10.5281/zenodo.13754343. Detailed implementation of the modeling framework and analysis experiments is provided in Jupyter notebooks at https://github.com/AI4SoilHealth/SoilHealthDataCube/tree/main/SOCD_map.

3.2 Material and methodology

3.2.1 Spatiotemporal ML for 3D+T data

To produce consistent and seamless SOCD predictions across the EU, we developed a standardized and modular automated soil mapping framework using spatiotemporal ML and 3D+T data (Fig. 3.1). We adopted RF as the core algorithm to predict SOCD

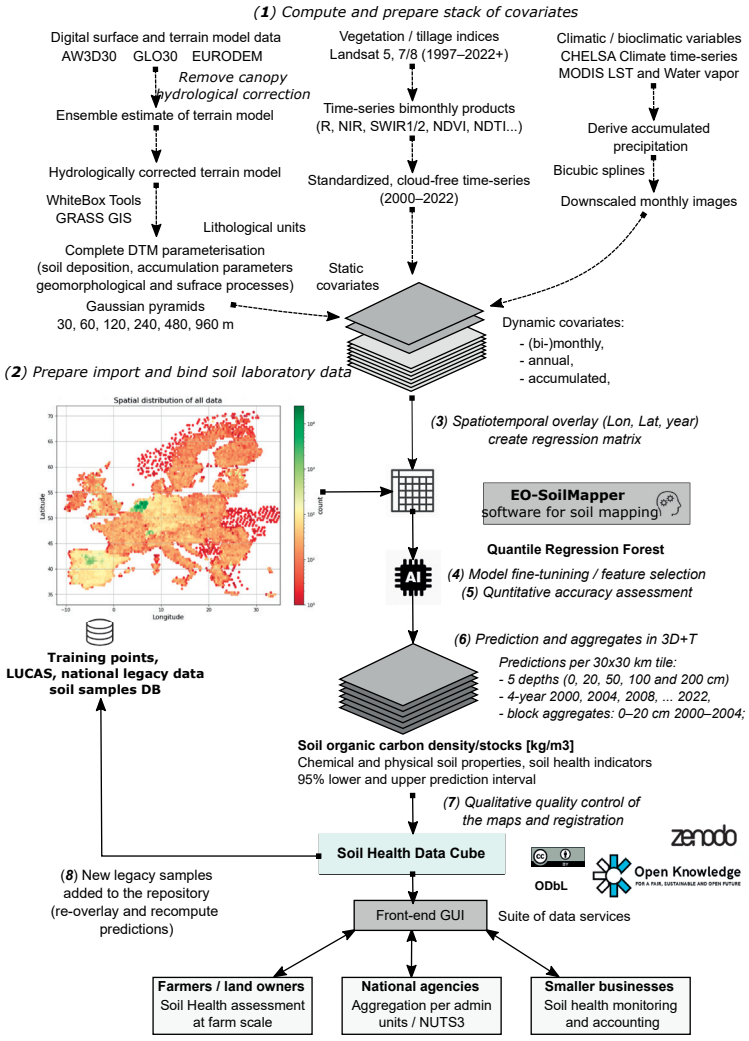


Figure 3.1: General eight-step framework for generating predictions as a part of the Soil Health Data Cube (AI4SoilHealth project). This is implemented as an automated workflow, allowing predictions to be updated and improved as new legacy soil data are harmonized and added to the training pool. Abbreviations: AW3D30 — ALOS World 3D 30 m Digital Surface Model (Japan Aerospace Exploration Agency, 2021), GLO30 — Copernicus GLO-30 Digital Surface Model (European Space Agency, 2024), NIR — Near Infrared, SWIR — Short-wave infrared, NDVI — Landsat Normalized Difference Vegetation Index, NDTI — Normalized Difference Tillage Index, MODIS — The NASA's Moderate Resolution Imaging Spectroradiometer, NUTS3 — EU's small regions based on the NUTS (Nomenclature of territorial units for statistics) classification, LST — land surface temperature, MODIS.

and estimate associated uncertainties due to its proven effectiveness in SOCD mapping (Hengl et al., 2018; Wadoux et al., 2020b) and its adaptability to model uncertainty, namely through the QRF (Meinshausen, 2006; Vaysse and Lagacherie, 2017). The spatiotemporal ML framework is based on eight steps that include (Fig. 3.1): (1) preparing feature layers, (2) preparing, importing and binding soil laboratory (training) data, (3) spatiotemporal overlay and generation of regression matrix, (4) model training, including feature selection, parameter fine-tuning, (5) quantitative accuracy assessment through cross-validation (CV) and individual test, (6) prediction, (7) qualitative quality control through visual checks to identify artifacts, gaps, overfitting, and potential problems, and (8) map updates. This framework is most similar to the frameworks suggested by Venter et al. (2021b) and Yuzugullu et al. (2024), with the difference that our framework is optimized for producing predictions in 3D+T with uncertainty mapped per pixel. This general modeling design is the basis of the Soil Health Data Cube, which is fully documented at https://shdc.ai4soilhealth.eu/.

Our models and output maps (predictions) are designed to be updatable; when reviewed by experts, spatial, temporal, or feature-space gaps can be identified, indicating where additional data and features are needed. As new data or improved features become available, the framework can be re-run to refine and update the maps. The eighth step, re-analysis, is especially dependent on receiving more European legacy soil laboratory data (training points) through emerging collaborations and initiatives, allowing us to continuously expand and update our pool of training points and gradually make more accurate predictions.

Note that detecting SOCD changes from LUCAS soil points is challenging, even though LUCAS provides repeated soil samples. Soils in general have complex and slow-changing nature (Poeplau and Don, 2013; Smith et al., 2020), and this challenge also applies to modeled SOC, even on an extended time scale such as 1953–2022, as highlighted by Helfenstein et al. (2024b). Thus, prediction uncertainties need to be reported at the pixel level, for example, by using the 95% prediction intervals (PI). This helps facilitate effective communication with end users and ensure proper application (Arrouays et al., 2020a).

We conducted an exploratory analysis to assess the potential for detecting temporal changes by evaluating whether the modeled time series exceed the prediction intervals (PIs), providing preliminary insight into the suitability of predictions for change detection. This temporal analysis is conducted at multiple levels, including the pixel level and, in this study, small regions (e.g., county) of Nomenclature of Territorial Units for Statistics (NUTS3) administrative units level through a case study, similar to the approach taken by Szatmári et al. (2021) and Szatmári et al. (2024b) in Hungary. We believe that such multi-scale analysis is relevant for applications across varying spatial contexts (Piikki and Söderström, 2019).

In addition to PI, we calculated anomaly probability maps to indicate the probability of extrapolation (risk) in the model. These maps provide users with guidance on areas where additional caution is required. Beyond generating and assessing maps and models, we analyzed feature importance with explainable ML techniques to understand how environmental features influence predictions, offering transparency into the model's behavior, and guiding future refinement.

3.2.2 Point data collection and harmonization

To calculate SOCD (kg/m^3), three key properties are required: SOC (content, g/kg), Bulk Density (BD, g/cm^3), and Coarse Fragments (CF, %). As described in Poeplau et al. (2017), there are two paths to compute SOCD, depending on the availability of the data. The first approach uses total Bulk Density (BD) (BD_{tot}) and coarse fragments by mass (CF_{mass}), as described in Eq. 3.1:

$$SOCD = \frac{SOC \cdot M_{fe}}{V_{tot}} = \frac{SOC \cdot M_{tot} \cdot (1 - 0.01 \cdot CF_{mass})}{V_{tot}} = SOC \cdot BD_{tot} \cdot (1 - 0.01 \cdot CF_{mass}) \tag{3.1}$$

The second approach utilizes fine earth BD (BD_{fe}) and coarse fragments by volume (CF_{vol}), as shown in Eq 3.2:

$$SOCD = \frac{SOC \cdot M_{fe}}{V_{tot}} = \frac{SOC \cdot M_{fe} \cdot (1 - 0.01 \cdot CF_{vol})}{V_{fe}} = SOC \cdot BD_{fe} \cdot (1 - 0.01 \cdot CF_{vol}) \tag{3.2}$$

Data on SOC, BD, and Coarse Fragment (CF) were collected from a range of regional, national, and pan-European surveys at the sample level, with each depth treated separately. The quality of these datasets varied, some datasets provided detailed information on soil depth, sampling time, sampling methods, measurement methods, and units, while others lacked such essential information. To ensure data consistency and reliability, standards were applied during data filtering and cleaning. Precise information on sampling year and geographic coordinates was required for spatial-temporal overlay with feature data to enable spatial-temporal modeling. Soil depth data was harmonized as follows: where a single value was recorded, it was directly used as the soil depth; where upper and lower depth limits were recorded, their mean value was calculated and used. The soil depth was then used as a feature in the modeling process, enabling the prediction of SOCD at various depths. Records missing any of these critical details were excluded.

For BD, we used laboratory measurements of the oven-dry mass of soil per unit volume, explicitly distinguishing between the fine earth density and the total BD. CF data were included only if they specified fragments larger than 2 mm and clearly indicated whether the measurements referred to mass or volume. SOC harmonization was guided by the LUCAS dataset, which served as the benchmark due to its high-quality data, extensive coverage, and standardized methodology across Europe. Measurements compatible with

the dry combustion method, as used in the LUCAS soil survey to determine SOC, were converted using established conversion factors. For example, although the conversion factor between the Walkley-Black method and dry combustion varies slightly between studies, the differences are generally minimal, and therefore we adopted the widely recognized factor of 1.3 (Kumar et al., 2019; Shamrikova et al., 2022; Walkley and Black, 1934). Data measured using incompatible methods or those lacking methodological details were excluded. An example of an incompatible method is loss on ignition (LOI), where conversion factors vary significantly across studies, ranging from 0.3 to 1.1 (Hernández et al., 2023; Konen et al., 2002; Kumar et al., 2019; Soon and Abboud, 1991). Furthermore, point data with zero SOCD values were excluded from the study if the associated land cover data did not indicate bare earth, rock, or sand.

After cleaning and harmonizing the data, we retained records that contained either all three properties—SOC, BD_{tot}, and CF_{mass}—or SOC, BD_{fe}, and CF_{vol}. These records were used to calculate the SOCD values using Eq 3.1 and 3.2, as described in Table 3.1. In addition, we included a small *"pseudo-zero"* SOCD dataset to address gaps in the feature space caused by the absence of samples from areas with minimal or no soil. These locations are often unintentionally omitted from field sampling campaigns due to practical reasonss (Barbet-Massin et al., 2012; Hengl et al., 2017). The *"pseudo-zero"* dataset was derived from land cover points identified as bare rock or shifting sand in the GLanCE project, where 20 cm resolution Very High Resolution (VHR) imagery was used to characterize such areas (Stanimirova et al., 2023). It was used during model calibration and training to help the model recognize covariate patterns associated with non-soil environments. Without such examples, the model could overestimate SOCD values in barren areas.

In total, we compiled 45,616 SOCD point data from 30,762 unique locations across Europe with valid environmental feature values (Table 3.1). The spatial distribution of this dataset is illustrated in Fig. 3.2-A. The LUCAS dataset serves as the cornerstone of the SOCD dataset. Although not the largest data source, it ensures a broad spatial distribution of samples across Europe. We specifically used BD_{fe} data derived by Pacini et al. (2023) from raw LUCAS BD (BD) data, which includes corresponding CF_{vol} values, facilitating the derivation of SOCD. National datasets further enriched the SOCD data but also introduced some clustering in specific regions. Another major data source is the systematic Spanish soil survey datasets, Parcelas COS (to 30 cm depth) and Parcelas INES (top 10 cm) (Serrano et al., 2022), which provided the majority of SOCD data, primarily from topsoil. BZE-LW, the core dataset of the first German agricultural soil inventory (Poeplau et al., 2020a; Poeplau et al., 2020b), is the second largest data source, offering more diverse data across soil depths. The INFOSOLO database from Portugal (Ramos et al., 2017) also contributed valuable data. The SOCD data estimated from GLanCE datasets account for the majority of data from bare rock and shift sands in the topsoil.

Table 3.1: Count of SOCD point measurements by sources and associated information.

Source	Derivation method	SOC method	Point count	Site count	Max depth (cm)
LUCAS	Eq (3.2)	Dry combustion	5806	5806	10
BZE-LW	Eq (3.2)	Dry combustion	16234	2938	184.5
Parcelas COS	Eq (3.1)	Walkley-Black	1576	788	20
Parcelas INES	Eq (3.1)	Walkley-Black	20656	20652	5
Infosolo	Eq (3.1)	Dry combustion	252	100	90
Infosolo	Eq (3.1)	Springer & Klee	106	25	179
Infosolo	Eq (3.1)	Walkley-Black	266	126	140
GLanCE	Estimation	Estimation	720	720	5

Our compiled trainind dataset spans 2000 to 2019, with notable data concentrations in 2012 and 2018 (Fig. 3.2-B) which are the LUCAS points. The overlap of BZE-LW and Parcelas INES datasets characterizes 2012, while 2018 corresponds to the LUCAS BD survey. The distribution of SOCD data is highly skewed, with most values concentrated around a central point and a long tail extending toward larger values (Fig. 3.3-C). Before 2010, the majority of data points of each year were from non-cropland areas; however, after 2010, the distribution shifted, with more data points come from cropland.

Depth-wise, most samples represent the topsoil (0–20 cm), as shown in Fig. 3.3, while data for deeper soil layers are more limited. Data are available for all types of soil cover in the topsoil, while data for deeper depths is limited to samples from *cropland*, *grassland*, and *woodland*. Generally, the mean measured SOCD values decrease with increasing soil depth up to 100 cm. Beyond this depth, from 100 cm to 200 cm, mean SOCD values appear to increase again; however, this inconsistency is likely due to limited data availability. We can also observe that the addition of the pseudo-zero SOCD dataset significantly improves the representation of *bare land and lichens or moss* land covers,

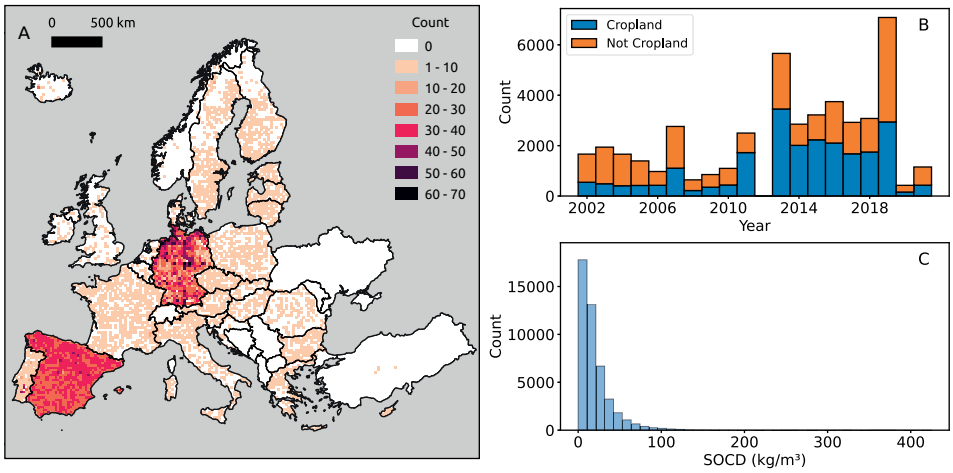


Figure 3.2: Distribution of point SOCD (kg/m^3) data: (A) Spatial distribution of SOCD across continental Europe, with each grid cell representing an area of 25 km^2; (B) Temporal distribution of SOCD from 2000 to 2020; (C) Histogram of SOCD values.

which account for approximately 80% of the data counts (720 out of 933). Note that the land cover information associated with the soil samples is based on the LUCAS land cover classification system, with the main classes shown in Fig. 3.3. For more details, see the LUCAS Technical Reference Document: C3 Classification (Land Cover & Land Use).

The SOCD values and the availability of the data also vary by land cover type (Fig. 3.3). The SOCD data from *water areas and wetlands* seem to exhibit the highest mean values, while the samples from *bare land and lichens or moss* show the lowest values. *Cropland* samples form the majority of the dataset, followed by grassland and woodland, while samples from other land cover types are quite limited. The land cover classifications were derived from the original data and standardized to align with the LUCAS land cover system. Note that 146 points lack valid land cover records and were excluded from this plot and any subsequent analyzes involving land cover strata.

All detailed data harmonization procedures are outlined in the Soil Health Data Cube manual, which comes with the production and analysis code.

3.2.3 Predictive features

A data-driven approach was used for feature selection. A wide range of features (also referred to as *"predictors"* or *"covariates"*) was included to represent various environmental factors such as vegetation, climatology, topography, human activity, and water content, etc. This redundancy resulted in an initial feature pool of 582 variables. Some features provided overlapping information; for example, Landsat-derived Bare Soil Fractions (BSF)

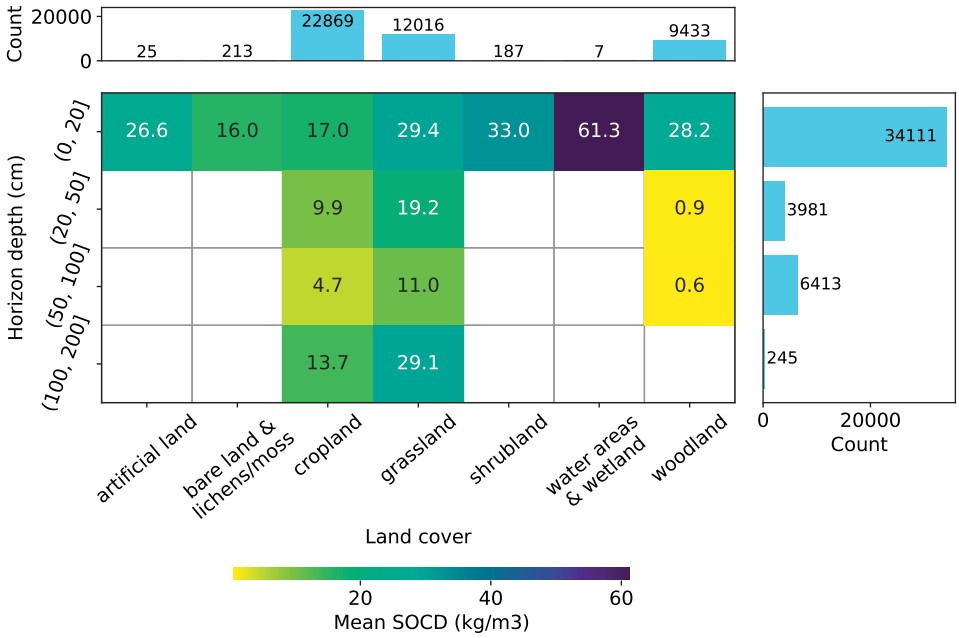


Figure 3.3: Distribution of SOCD observations (excluding *pseudo-zero* data) across land cover types and soil depth intervals. Grid cells and colors represent the mean SOCD for each land cover–depth combination.

from time-series analysis may overlap with bare soil (BS) fractions derived by Sun et al. (2023) using spectral unmixing.

For time-series data with missing values for recent years (e.g., 2020 and later) or earlier years (e.g., before 2001), we substituted the most recent available data from surrounding years. For example, CHELSA Climate time-series data ends in 2019, so data for 2020, 2021, and 2022 were copied from 2019. All data layers were resampled to a 30 m resolution using cubic splines (as implemented in GDAL) where necessary, ensuring consistency in spatial extent, resolution, and coordinate reference system. The following paragraphs detail the preparation of feature groups and the feature selection method.

The topographic variables were derived from the Ensemble Global Digital Terrain Model (EDTM) at 30 m spatial resolution and based on AW3D30, GLO30 and continental DTMs (Ho et al., 2023), as the topography-factor attributes for SCORPAN model (McBratney et al., 2003). These include terrain parameters such as slope (in degrees), minimum and maximum curvature, hillshade, negative, and positive openness. Hydrology parameters, including specific catchment area, slope length and steepness (LS) factor, flow accumulation, and topographic wetness index (TWI), were obtained using SAGA GIS (Conrad et al., 2015). Furthermore, northernness and easternness (representing the north-facing and

east-facing component of the slope, used as a proxy for sunlight exposure), and geomorphon classes were generated using GRASS GIS (GRASS Development Team, 2023) to capture the landscape orientation and geomorphological features, which plays an important role in the pedogenetic process (Sena et al., 2020). The EDTM was resampled from 30 m to 60 m, 120 m, 240 m, 480 m and 960 m and the derivation of parameters is repeated to produce a Gaussian pyramid (Behrens et al., 2018). Coarse-resolution layers (240 m, 480 m, and 960 m) derive terrain and hydrology parameters on a continental scale. Fine-resolution layers (30 m, 60 m and 120 m) are first divided into tiles and padded with 3400 pixels for geomorphon classes and hydrology parameters, whereas 100 pixels for the rest. The tiles are mosaicked and then reprojected to EPSG:3035 and finally cropped to the same spatial extent as the other layers.

The climate data used in this study include precipitation, daytime and nighttime Land Surface Temperatures (LST), water vapor, and long-term BIOCLIM variables (Karger et al., 2017). The precipitation data consist of cumulative monthly and annual cumulative sums, as well as long term average of monthly precipitation, all derived from the CHELSA daily precipitation product (Karger et al., 2021). Temperature data covers time series for monthly and annual median values, 5th and 95th percentiles, and standard deviations (SDs) of daytime and nighttime LST, sourced from the MODIS LST product (Wan, 2006). Water vapor data includes monthly mean and standard deviation, as well as annual median, 25th, and 75th percentiles and standard deviation, derived from the MODIS water vapor product (Lyapustin and Wang, 2018), as processed by Parente et al. (2023). Long-term climate characteristics, such as annual trends (e.g., mean annual temperature, annual precipitation), seasonality (e.g., annual temperature and precipitation ranges) and extreme or limiting environmental factors (e.g., temperatures of the coldest and warmest months and precipitation during the wettest and driest quarters), are represented by CHELSA BIOCLIM variables (Brun et al., 2022; Karger et al., 2017).

We produced a gap-free lithology map for Europe by imputing missing information from the EGDI/OneGeology surface geology map using a RF classifier trained on topographic variables and soil region maps (BGR) (Bundesanstalt für Geowissenschaften und Rohstoffe, 2005). For Türkiye, lithology was digitized and harmonized using the geological map from MTA `https://atag.itu.edu.tr/kataloglar/gis/Geology.zip`.

To consider the influence of human activities on SOCD, several variables were included to capture the extent and intensity of human pressure, including Human Footprint Index, nightlight data and cropland extent. Human Footprint Index, developed by Mu et al. (2022) combines data on land cover, accessibility, nightlights, and population density into a scoring scheme to estimate human pressure. Annual visible nightlight data (V2) from NASA/NOAA's Visible Infrared Imaging Radiometer Suite for the period 2012–2019 (Elvidge et al., 2021), extrapolated for 2000–2011 using logistic regression by Hengl (2023), were included as a proxy for human activity and land-use intensity, which could potentially

influence land productivity there influence SOCD (Hackländer et al., 2024). The global maps for cropland extent developed by Potapov et al. (2022) are adopted to indicate the possible human impact on the soil related to agricultural practices, such as crop residues and soil disturbance patterns. This cropland data set includes annual and perennial herbaceous crops, excluding woody crops and permanent pastures.

Land cover dynamics, particularly those involving vegetation and soil components, is essential to understand SOCD dynamics. To quantify land cover changes over time, we used two datasets derived from distinct methods, both providing continuous numerical representations of land cover rather than hard classes. The first data set represents the annual mean, maximum and standard deviation of photosynthetic vegetation (PV), non-photosynthetic vegetation (NPV), and bare soil (BS) fractions. These metrics were derived from the monthly Vegetation and Soil Fractions (2001–2022) product, generated by spectral unmixing of reflectance images (Sun et al., 2023). The second data set is the Plant Functional Types (PFT) Maps (1992–2020) (Harper et al., 2023), which were produced using the European Space Agency's Climate Change Initiative (CCI) land cover data product. This dataset, created using a cross-walk table and auxiliary data, provides annual compositions of 14 PFTs at 300 m resolution, capturing intra-class spatial variability for detailed surface process representation.

To better represent soil surface structure, we used the normalized mean radar backscatter coefficients from the Sentinel-1 Global Backscatter Model (S1GBM), which was produced using SAR images covering 2016–2017 period and was corrected for incidence angle variations to achieve consistency across the globe (Bauer-Marschallinger et al., 2021). The mosaics of C-band dual polarization backscatters (VV/VH) were downsampled from 10 m to 30 m spatial resolution to comply with high resolution features. The use of synthetic aperture radar (SAR) images has been shown to be effective in mapping various soil compositions, such as soil moisture (Celik et al., 2022), soil type (Deodoro et al., 2023), and soil organic carbon (Nguyen et al., 2022).

Although most other features are initially available at coarser resolutions and later resampled to finer scales, we used a 30 m Landsat-based spectral indices data cube as the foundation for our 30 m spatial resolution SOCD mapping. This data cube, developed by Tian et al. (2025b) from Landsat ARD V2 data (Potapov et al., 2020) using the method of Consoli et al. (2024), includes both surface reflectance bands (e.g., red, blue) and derived spectral indices such as the Normalized Difference Vegetation Index (NDVI), Normalized Difference Tillage Index (NDTI), Bare Soil Fraction (BSF), etc. These indices capture various environmental factors, including vegetation, soil, crops, and water. In addition to its extensive range of indices, the data cube is available at multiple temporal resolutions—bimonthly, annual, and long-term (2000–2022)—providing temporal characteristics at various levels.

3.2.4 Model calibration

3.2.4.1 Spatial-temporal overlay

The spatial-temporal overlay links point observations with corresponding environmental feature values based on location and year, generating a regression matrix for model calibration and evaluation. Environmental features are broadly divided into two groups: dynamic features and static features. Dynamic features represent environmental processes that vary spatially and temporally, with values changing over time. In contrast, static features are location-specific and remain constant over time for a given location. These typically include relatively stable environmental factors (e.g., lithology or topographic features) or aggregated representations of dynamic features over a long period (e.g., the 50th percentile of NDVI from 2000 to 2022).

For dynamic features, the sampling year of each point observation was used to align it with the corresponding feature values. For features with finer temporal resolutions, such as bimonthly (one value per two months) or monthly (one value per month), which vary within a year, all temporal values available for the feature within the observation year are overlaid with the point observations. The overlay operation was performed using the `SpaceTimeOverlay` function from the `scikit-map` package (Consoli et al., 2024).

3.2.4.2 Data split, usage and transformation

The overlaid dataset was then divided into three subsets: a calibration set, a training set, and a test set. This split is done using a stratified approach to ensure representativeness across different conditions, with strata defined by combinations of depth intervals and land cover types, as shown in Fig. 3.3. Pseudo-zero points were excluded from the hold-out test set, as they are not based on actual measurements and the cluster of data at origin could cause potential distortion of performance metrics. Approximately 10% of the data was selected as the calibration set, which was used to establish the model structure, including feature selection and hyperparameter tuning. This separation helps improve model generalizability and reduce overfitting. The training set was used for model training and CV, while the independent test set was excluded from both calibration and training to provide an unbiased validation. The final subsets included 4,985 point measurements for model calibration, 38,348 for training, and 2,283 for testing.

The distribution of SOCD data is highly skewed, making data transformation beneficial for effectively applying geostatistical methods (Martin et al., 2014; Osborne, 2010). An experiment detailed in the supplementary notebook 005_transformation_comparison shows that transforming SOCD data also enhances the performance of RF models, despite their non-parametric nature. Therefore, the RF model in this study was applied to log-transformed SOCD data, with predictions back-transformed before being presented to users as maps.

3.2.4.3 Extended Random Forest

RF was selected for this study to model and map SOCD due to its demonstrated effectiveness in handling large datasets and capturing complex non-linear relationships (Hengl et al., 2018; Wadoux et al., 2020b). Its extension, QRF (Meinshausen, 2006), quantifies uncertainty by providing estimates of conditional quantiles from the full distribution of predictions across all trees in the forest, rather than focusing solely on the conditional mean (Schmidinger and Heuvelink, 2023; Szatmári and Pásztor, 2019; Vaysse and Lagacherie, 2017).

In this study, we extended the RF implementation to run RF and QRF models simultaneously in a single execution. Specifically, we modified the `RandomForestRegressor` function to store the full output distribution from each tree in the forest. This modification allows us to compute both the conditional mean, consistent with standard RF predictions, and specific quantiles (e.g., P0.025 and P0.975) for QRF functionality. This integrated approach enhances the efficiency of the modeling and mapping process, allowing both predictions and PIs to be obtained simultaneously, reducing computational overhead. The resulting implementation is available as the `trees_rf` function in the Python library scikit-map (Consoli et al., 2024).

3.2.4.4 Feature selection

The Repeated Subsampling-Based Cumulative Feature Importance (RSCFI) method was developed to select the most relevant features from the candidate feature pool. Similar to Recursive Feature Elimination with Cross Validation (RFECV) — a standard approach in digital soil mapping (Wadoux et al., 2020b) — RSCFI recursively eliminates features based on CV results. However, unlike RFECV, which removes a fixed number of features at each step, RSCFI removes features with Cumulative Feature Importance (CFI) values below a predefined threshold at each iteration. RSCFI is designed to balance model performance and computational efficiency. When many features have low CFI values, RSCFI efficiently removes most irrelevant features. Conversely, if features have comparable CFI values, RSCFI removes a stable number of features recursively, requiring a similar amount of time as RFECV. Both methods were tested on the calibration dataset, and the method yielding optimal performance across metrics — including the coefficient of determination (R^2), concordance correlation coefficient (CCC; Lawrence and Lin, 1989), mean absolute error (MAE), and median absolute error (MedAE) — was selected for feature selection.

3.2.4.5 Hyper-parameter fine tuning

With preprocessed data and selected features, hyperparameter fine-tuning was performed using 5-fold CV with the `HalvingRandomSearchCV` function (Pedregosa et al., 2011), optimizing for CCC as the criterion. The initial hyperparameter space is detailed in Table 3.2. The number of trees (`n_estimators`) was set to 120, following the recommendation of

Oshiro et al. (2012), to balance model performance, computational efficiency, and memory usage. This choice is also supported by our experimental results, which show no significant performance improvement when the number of trees exceeds 120 (see supplementary notebook 007_explore_hyper.parameter).

The hyperparameter space defines key settings that influence the model's performance and complexity. The number of trees (`n_estimators`) determines the size of the ensemble. The **criterion** parameter specifies the function used to evaluate the quality of the splits, such as *squared error* or *absolute error*. The maximum depth of a tree (`max_depth`) controls its complexity to prevent overfitting, while the `max_features` parameter determines the number of features considered for the best split, expressed as a fraction or function. The `min_samples_split` parameter sets the minimum number of samples required to split an internal node, influencing the model's simplicity. Similarly, `min_samples_leaf` specifies the minimum number of samples needed at a leaf node, restricting the size of terminal nodes and helping to avoid overfitting.

Table 3.2: Hyperparameter space for the RF regression model.

Hyperparameter	Feature Space
n_estimators	{120}
criterion	{squared_error, absolute_error, poisson, friedman_mse}
max_depth	{10, 20, 30}
max_features	{0.3, 0.5, 0.7, log2, sqrt}
min_samples_split	{2, 5, 10}
min_samples_leaf	{1, 2, 4}

3.2.5 Model evaluation

Once the model structure was established with the calibration set, the RF model was trained and evaluated using the training and test datasets. Predictions were generated by averaging the output distribution of all trees. Model performance for SOCD predictions was assessed through three components: (1) 5-fold inverse sampling-intensity weighted cross-validation (ISIW-CV) on the training dataset, where data were randomly partitioned into folds and lower weights were assigned to more spatially clustered observations to address spatial sampling bias (De Bruin et al., 2022); (2) leave-one-year-out cross-validation (LOYO-CV) on the training dataset to assess the model's temporal transferability, in which data from an entire year were held out as the test set in each iteration (Huang et al., 2024); and (3) independent testing on a withheld test set, stratified by depth and land cover type. In both (1) and (2), CV was performed on depth-specific observations. Performance metrics—including MAE, bias, R^2, and CCC—were calculated for all assessment components using only real measurements. Pseudo-zero points were included

solely during model training and excluded from metrics calculation, as their inclusion could disproportionately affect the metrics due to their clustering near the origin. For the independent test, performance metrics were also reported for different years, and combinations of land cover and soil depth strata.

PIs are derived by calculating the conditional quantiles of SOCDs, specifically the 95% PI, with lower and upper bounds set at the 2.5th and 97.5th percentiles (P0.025 and P0.975) (Meinshausen, 2006). To evaluate the reliability of the PIs, several validation metrics are calculated on the test set, as recommended by Schmidinger and Heuvelink (2023), Goovaerts (2001), and Vaysse and Lagacherie (2017):

- **Prediction Interval Coverage Probability (PICP):** Assesses accuracy by calculating the fraction of true values falling within the prediction intervals.
- **Prediction Interval Width (PIW):** Evaluates sharpness by measuring the width required to capture the predictions accurately.
- **Quantile Coverage Probability (QCP):** Measures the asymmetry in the uncertainty coverage. It has the same underlying logic as PICP, but it evaluates single-quantile predictions, thereby highlighting the symmetry of the PIs.
- **Accuracy plot:** Evaluate the model's ability to predict local uncertainty by means of a scatter plot of the estimated versus observed fractions for different PI values.

Among these, the PIW and PICP are also reported for different combinations of land cover and soil depth strata.

3.2.6 Extrapolation risk probability from anomaly scores

Extrapolation decreases ML performance, but is unavoidable in large-scale spatial mapping, making communication of such uncertainties essential. Various methods exist to identify predictions made in dissimilar feature spaces, such as the area of applicability (AOA) (Meyer and Pebesma, 2021), Isolation Forest (Liu et al., 2008), and homosoils (Nenkam et al., 2022). Considering the computational demands and the extensive mapping scope of this study, in the end, we decided to choose Isolation Forest for its efficiency and applicability for non-normal multivariate data sets (Liu et al., 2008). In most of applications of Isolation Forest, it is used in fact to determine probability of inclusion using occurrence only records (Song and Estes, 2023).

Isolation Forest identifies regions outside the training data range by randomly partitioning the data and isolating samples. The `ensemble.IsolationForest` implementation from `scikit-learn` (Pedregosa et al., 2011) computes an anomaly score by averaging the path lengths over a forest of (i.e., ensemble of random trees), with shorter path lengths to indicate higher anomaly. Samples located in low-density or unfamiliar regions of the feature space typically require fewer splits and thus have shorter average path lengths.

The average number of splits—i.e., the average path length across the dataset—is used as a threshold to distinguish between regions that lie within or outside the training feature space (Liu et al., 2008). To communicate extrapolation risk, we transformed the anomaly scores produced by `IsolationForest` into a normalized scale from 0 to 1, where higher values represent greater extrapolation risk for a given sample or pixel. The threshold for distinguishing in-sample vs. out-of-sample regions was also rescaled to match this normalized scale, ensuring consistency with the extrapolation risk probability maps provided to end-users.

3.2.7 Map prediction

The final SOCD model was trained using all available SOCD data for mapping purposes. The map production was performed using the Python library `scikit-map` (Consoli et al., 2024). To balance temporal resolution and computational resources, a four-year interval was chosen for this analysis, with potential extensions to finer temporal resolutions, such as annual predictions. This final single model is applied to generate SOCD predictions were generated for multiple depths (0 cm, 20 cm, 50 cm, 100 cm, and 200 cm) across different years (2000, 2004, 2008, 2012, 2016, 2020, and 2022). A *spacetime block* method was developed to adjust the output value of each RF tree considering a particular depth interval and a specific time interval by averaging predictions from consecutive depths and years to create ensembles of averaged tree outputs (see Fig. 3.4). This approach provided mean predictions and uncertainty estimates via prediction intervals for four depth intervals (0–20 cm, 20–50 cm, 50–100 cm, and 100–200 cm) and six time periods (2000–2004, 2004–2008, 2008–2012, 2012–2016, 2016–2020, and 2020–2022). The standard deviation of the mean decreases by about $1/\sqrt{N}$ where N is the number of input prediction spacetime locations (e.g. 4).

The purpose of the aggregation over *spacetime block* is to create block predictions over specific time periods and standard depth intervals, ensuring they are both representative and practical for end users. Aggregating predictions over fixed space-time intervals helps mitigate significant inter-annual variability in high-resolution covariates, such as Landsat indices, which are influenced by factors like rainfall variability, forest fires, and floods. Such variability can result in oscillating predictions that are difficult to interpret and beyond the scope and budget of this project. Aggregation reduces these fluctuations, yielding smoother distributions and more stable predictions. Furthermore, most users of SOCD data focus on SOC stocks for standard depth intervals, such as 0–20 cm or 0–100 cm. The block method simplifies the calculation of total SOCD stocks for these intervals. Consequently, our predictions can only be evaluated using block or composite measurements, such as mixing soil samples from 2–3 depths (e.g., 0–20 cm) and repeating the process over 2–3 years to produce a spacetime block estimate of SOCD.

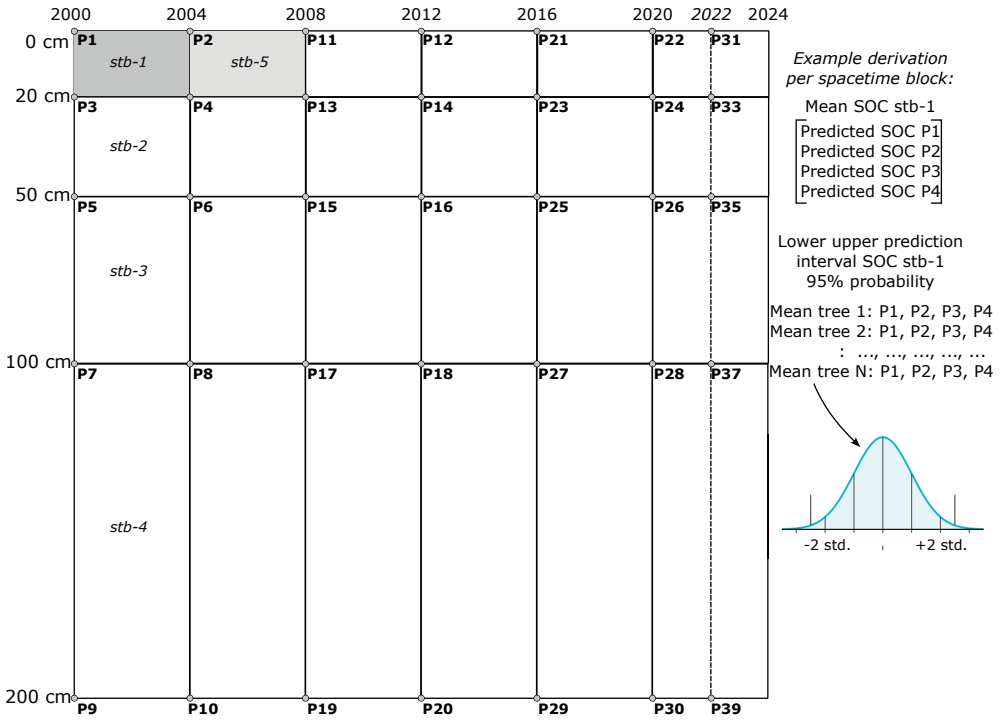


Figure 3.4: Illustration of SOCD prediction and PI generated in spacetime blocks across different temporal and depth intervals (3D+T) during map production. One spacetime block (e.g. stb1) is produced by averaging four predictions P1–4, while the uncertainty is derived by estimating distribution for N means per tree. This aggregation was chosen to reduce computational complexity and to reduce inter-annual variability effects of weather on mapping SOCD.

For each tile, all features selected by the model were loaded into memory for the corresponding area, year, and soil depth. The resulting tiled SOCD predictions, including uncertainties obtained by the model considering the PI, were saved as Cloud Optimized GeoTIFF files (COG). After all pan-EU tiles were produced, they were mosaicked, saved as COG, and stored in a local S3-based cloud storage, to allow visual inspection. Land masking was performed during the mosaicking process. Pan-EU land mask data used after the tiles of each product were mosaicked to create the EU-wide map.

In this study, we chose to apply the model across the entire continent, and inform users about the reliability of the predictions through guidance layers—namely, uncertainty maps, extrapolation risk maps, and performance metrics. We limit masking to buildings and

permanent water bodies—areas where SOCD has no practical relevance—which do not directly correspond to the LUCAS land cover classes used in model performance evaluation. Land masks—developed based on areas of agreement across datasets to exclude permanent water bodies and built-up areas—were applied (DLR Geoservice of the Earth Observation Center, 2019; Tian et al., 2023). Note that we mask only buildings, so most urban areas, parks, and smaller farms are still included in the maps. Compared to masking land cover types with weaker performance, this approach minimizes the risk of excluding potentially valuable information. It allows results to be interpreted in the context of local knowledge, with the final decision left to the users. This rationale also applies to soil depth, for which we provide complete predictions down to 200 cm, accompanied by corresponding uncertainty information.

Extrapolation risk probability maps are provided alongside predictions, with a recommended threshold to mask areas with high extrapolation risks in the SOCD maps. Users can also adjust the threshold based on their specific needs or specific risk tolerance.

3.2.8 Feature impact analysis with Shapley values

Shapley values, derived from cooperative game theory, are used in this study to interpret RF model predictions. This model-agnostic approach quantifies the contribution of each feature to the predictions, captures local variations and the importance of global features, and provides intuitive visualizations (Padarian et al., 2020; Shapley, 1997; Wadoux and Molnar, 2022; Wadoux et al., 2023). Using the `shap` package in Python, we calculated Shapley values on the test set, generating: (1) a global summary of average absolute Shapley values for general feature importance; (2) Partial Dependence Plots (PDPs) for the top 10 features; and (3) contribution plots for two locations, illustrating how features influence specific predictions. The Shapley values of the dynamic and static features are compared and analyzed, providing insight into their relative contributions and temporal dynamics in the model. It is important to note that the computation of Shapley values is intensive; therefore, it was performed only on the test dataset. Although this method provides valuable insights into the model decision-making process, its high computational cost makes it challenging to scale to large datasets.

3.2.9 Quantify uncertainty of spatial aggregates

SOCD prediction uncertainty was examined over time to assess change detectability by comparing PIs with predicted changes. This analysis was conducted at two spatial supports: the pixel level and the NUTS3 regional level. At the pixel level, two individual pixels from different land covers were analyzed, while at the NUTS3 level, two regions — Unterallgäu and the area of Mindelheim Stadt within it, located in Bavaria, Germany — were merged and analyzed as a single area of interest (AOI) for demonstration purposes. The spatially aggregated SOCD in the AOI was calculated as the average predicted SOCD within the

region. To quantify uncertainty in the AOI's spatial aggregates, a method commonly applied in spatial environmental modeling (Araza et al., 2022; Wadoux and Heuvelink, 2023) was used, which accounts for the spatial autocorrelation of map errors within the aggregate:

$$\mathrm{sd}_{MB} = \sqrt{\frac{1}{|B|^2} \int_{s \in B} \int_{u \in B} \sigma(s) \cdot \sigma(u) \cdot \rho(|s-u|)\, ds\, du} \tag{3.3}$$

The integral in Eq.(3.3) was evaluated using a 500 m grid, larger than the pixel size, to balance computational efficiency and the ability to account for spatial autocorrelation within the AOI. The B in Eq.(3.3) represents the number of discretization points, which are randomly sampled from each discretized grid; $\sigma(s)$ and $\sigma(u)$ denote the standard deviations of the point support prediction errors at locations s and u. The $\rho(|s-u|)$ represents the correlation function of the standardized prediction error at the separation distance between s and u, derived from the variogram $\gamma(h)$ using Eq.(3.4) (Webster and Oliver, 2007, Section 4.1).

The variogram was calculated from the standardized residuals. The residuals, calculated as the differences between the map predictions and the SOCD measurements available in Germany, are standardized by dividing by the standard deviation of the map prediction error. This standardization is necessary to achieve homoscedasticity of the residuals. The standard deviation was approximated by dividing the 95% prediction interval width (P0.975–P0.025) by four, assuming a normal distribution of SOCD predictions. However, this approximation caused the variogram of standardized residuals to deviate from the expected unit sill. To address this, we applied a correction factor (the square root of the initial sill) to adjust the proxy standard deviation and ensures a unit sill upon re-standardization:

$$\rho(h) = \frac{\text{sill} - \gamma(h)}{\text{sill}} \tag{3.4}$$

Furthermore, we calculated the average and uncertainty of this AOI using design-based (DB) estimators from the available measurement samples and compared it to the AOI mean and uncertainty quantified from the map predictions; see e.g. Emick et al. (2023). The standard deviation of the mean over the NUTS3 units was calculated using Eq.(3.5), where n represents the sample size, y_i denotes SOCD in an individual sample within the collection and $\bar{y}_{DB}$ is the mean of the sampled SOCD in the AOI:

$$\mathrm{sd}_{DB} = \sqrt{\frac{1}{n(n-1)} \sum_{i \in n} (y_i - \bar{y}_{DB})^2} \tag{3.5}$$

3.3 Results

3.3.1 Feature impact analysis

RSCFI achieves comparable model performance to RFECV while significantly reducing processing time (see supplementary notebook 006_feature.selection_rfecv.rscfi in https://github.com/AI4SoilHealth/SoilHealthDataCube/). Applied to the calibration dataset using a naive RF model without parameter fine-tuning, RSCFI reduced the initial 582 features to 67, as shown in Fig. 3.5. The selected 67 feature set includes 33 climate features, 18 Landsat index features, and 13 topographic features. The remaining three are soil depth, VV radar backscatter, and the probability of peat lithology class (class code: 78). Among the climate features, 11 are long-term CHELSA BIOBLIM features, nine are precipitation-related, nine are LST features, two represent water vapor, and two indicate cloud coverage. Notably, no features related to human activities or land cover were selected. Static features were more likely to be kept, with 276 static and 306 dynamic variables in the original pool, compared to 39 static and 28 dynamic variables in the selected set.

Global feature importance was evaluated using the average absolute Shapley values for each feature across all predictions, with larger values indicating a stronger impact on model predictions. For simplicity and clarity, these are referred to as global Shapley values. Fig. 3.5 presents the global Shapley values for the 67 selected features. The `soil_depth` feature has the highest global Shapley value (0.396), and truncated in the figure for clarity. Static features not only outnumber dynamic features but also exhibit higher average global Shapley values (0.025), approximately three times that of dynamic features (0.008). Even when `soil_depth` is excluded, static features maintain a higher mean global Shapley value (0.015) compared to dynamic features.

Fig. 3.6 presents the partial dependence plots (PDPs) of Shapley values with respect to feature values for the 10 most important features, as identified by global Shapley values. The sign of a Shapley value reflects the direction of a feature's influence: positive values indicate that the feature increases the prediction relative to the average, while negative values suggest the opposite. For `soil_depth,` small values correspond to moderately high Shapley values. As `soil_depth` increases, Shapley values decrease, turning negative at depths greater than 20 cm. The negative influence of larger `soil_depth` values is particularly pronounced.

Vegetation-related features consistently exhibit a positive correlation with their corresponding Shapley values. Compared to the `p25` features—`ndvi_longterm_p25` (25th percentile of long-term NDVI values from 2000 to 2022) and `fapar_longterm_p25`—the corresponding `p50` features display a more linear slope. However, data points where these four vegetation features fall below 0.2 are not well captured by the fitted PDP curves; they correspond to negative Shapley values and do not show an obvious trend. In contrast,

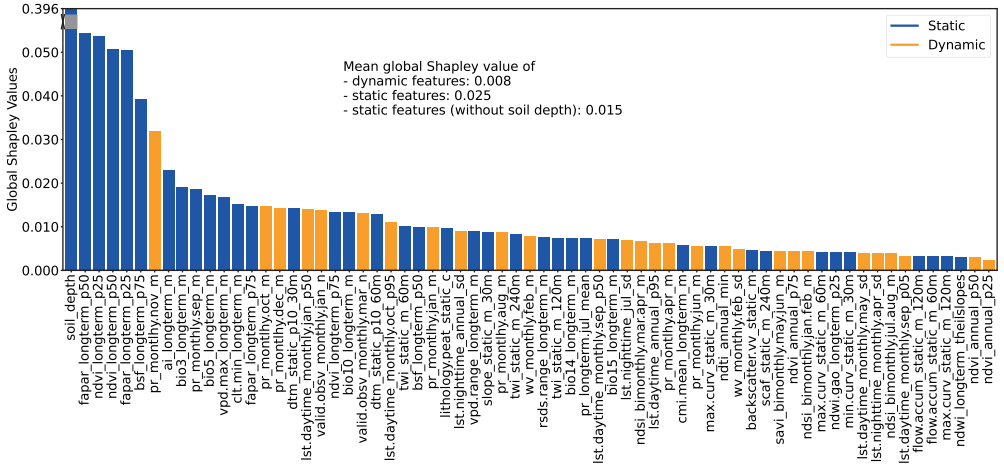


Figure 3.5: Global Shapley values for each feature, representing the feature's contribution to SOCD prediction at each location. The global Shapley value of soil_depth is substantially higher than those of other features; therefore, it has been partially rescaled (or 'squeezed') into a gray scale in the plot for visual clarity.. Feature names are displayed on the x-axis and consist of three fields: the first indicates the feature variable, the second specifies the temporal support, and the third denotes the variable type (e.g., mean, standard deviation). For topographic features, a fourth field is included to indicate the spatial resolution.

`bsf_longterm_p75` (75th percentile of long-term bare soil fraction values from 2000 to 2022), where higher values indicate barer soil surfaces, exhibits a clear negative trend in its PDP.

The last four features among the 10 most important are related to climate, three of which are associated with precipitation. For `pr_monthly.nov_m` (precipitation amount in November, kg m^{-2} $month^{-1}$), Shapley values remain close to zero when feature values are below 200 but increase steeply beyond this threshold. Another precipitation feature, `pr_monthly.sep_m` (precipitation in September), shows a different pattern: Shapley values increase with feature values up to 100, after which they begin to decline.

The `ai_longterm_mean` (aridity index, defined as the ratio of mean annual precipitation to mean annual potential evapotranspiration) displays an almost linear positive trend in its partial dependence plot (PDP) when values are below three, suggesting that less arid conditions correspond to higher predicted SOCD. However, when `ai_longterm_mean` exceeds three, the slope reverses, indicating a decreasing effect on SOCD predictions.

The temperature feature `bio3_longterm_mean` refers to isothermality, which ratio of diurnal variation to annual variation in temperatures. Its Shapley values exhibit less variation than the others. When feature value is around 30, there seems to be a clearer pattern that: when the feature values are below about 30, Shapley values tend to be

negative, indicating an decreasing effect on SOCD predictions. When it's higher than about 30, the Shaply values tend to positive. But in general, The Shapley values oscillate around zero.

Fig. 3.7 illustrates the feature contributions to the SOCD predictions at two distinct locations: (A) a forested area in Finland and (B) an agricultural area in Spain. The predicted SOCD values for these locations are 61.95 kg/m^3 (4.14 on a log1p-scale) and 8.48 kg/m^3 (2.25 on a log1p-scale), respectively, compared to observed SOCD values of 62.15 kg/m^3 (4.15 on a log1p-scale) and 4.76 kg/m^3 (1.75 on a log1p-scale).

The Shapley value patterns differ significantly between these two points. At point A, characterized by high vegetation coverage, a high probability of peat as the lithological class, shallow soil depth (10 cm), and relatively high `clt.min_longterm_m` (cloud coverage fraction, high within the test data distribution), nearly all features contribute positively to the predicted SOCD values. The only exception is `lst.daytime_monthly.oct_p95` (October daytime temperature), which decreases the predicted SOCD values. In contrast, at point B, nearly all features contribute negatively to the predicted SOCD values. This point is defined by low precipitation, low vegetation coverage, and large `rsds.range_longterm_m` (the difference between maximum and minimum monthly surface downwelling shortwave flux in air, high in the test data distribution). The only feature contributing positively to the prediction at point B is shallow `soil_depth`.

The top ten features influencing SOCD predictions at these two points mostly overlap, with both sharing `soil_depth` and vegetation features as key features. However, different climate features are important: at point A, `lst.daytime_monthly.oct_p95` and `clt.min_longterm_m` are significant, whereas at point B, `rsds.range_longterm_m` and `lst.daytime_monthly.jan_p50` play a key role. Additionally, at point A, `lithology.peat_static_c` (the probability of lithological class peat) is also an important feature.

3.3.2 SOCD model prediction accuracy

Fig. 3.8 shows that across all three evaluation components—ISIW-CV, LOYO-CV, and the independent test: the predicted values align well with the observed values. Across the three evaluations, the independent test shows the best performance, followed by ISIW-CV and then LOYO-CV. In all validation efforts, the model exhibits a tendency to underestimate high SOCD values. In the bottom panel, plotted on a logarithmic scale, it is more evident that the SOCD values are also slightly overestimated for low SOCD values. Fig. 3.9 illustrates the CCC for each year in the test dataset, highlighting substantial variation across years. And no obvious correlation between CCC and data availbility is observed.

From the top plot in Fig. 3.10, we observe that for topsoil (0–20 cm), the CCC is highest for *bare land and lichens or moss* (0.70), followed by *grassland*, *cropland*, and *woodland* (> 0.5), while the lowest CCC is observed for *shrubland* (-0.22). Depth-wise, for land covers with soil samples deeper than 20 cm, the CCC is generally higher for the 20–50 cm depth interval than for topsoil. Only *cropland* and *grassland* have sufficient data for depths beyond 50 cm. For *cropland*, the CCC decreases to 0.30 in the 50–100 cm interval but increases again at 100–200 cm. For *grassland*, the CCC peaks in the 50–100 cm interval (0.73).

The bottom plot in Fig. 3.10 shows that, in general, negative biases are more prevalent than positive ones. The most pronounced underestimation occurs in *water areas and wetlands* (-34.98), although it is important to note that this land cover includes only seven sample points in the entire dataset (Fig. 3.3). The second largest underestimation is observed in the topsoil of *grassland* (-6.25), followed by the topsoil of *woodland* (-4.14). Overestimation is most notable in the topsoil of *shrubland* (3.40), followed by the topsoil of *artificial land* (2.78). For deeper soil layers, the magnitude of bias generally decreases but begins to increase again beyond 100 cm for land cover types with available samples at these depths. Plots for additional metrics across combinations of land cover and soil depth intervals are available in the supplementary notebook 008_evaluation_test.accuracy.

3.3.3 Uncertainty estimation evaluation

The QCP plots on the left of Fig. 3.11 indicate that the model shows slight optimism in quantile prediction. QCP values tend to be slightly higher than the target quantiles when they are below 0.5 and slightly lower when target quantiles exceed 0.5. This pattern is also reflected in the PICP accuracy plot on the right of Fig. 3.11, where PICP falls below the target PIs in the middle range. Overall, the model's uncertainty estimation aligns well with expectations. In this study, we present the 95% uncertainty map using the upper and lower limits (P0.025 and P0.975), where PICP equals 91%.

When examining the 95% PI across varying levels of observed SOCD, the uncertainty model demonstrates inconsistent performance (Fig. 3.12). For SOCD values below 10 kg/m^3, the PICP is slightly below the ideal level. Between 10 and 30 kg/m^3, the PICP aligns well with the target level. However, as SOCD values exceed 30 kg/m^3, the PICP generally declines with decreasing data availability. The PIW increases steadily as the availability of data decreases with increasing SOCD values. For SOCD values above 100 kg/m^3, both PICP and PIW become unstable due to very limited data.

Fig. 3.13 presents the uncertainty estimation metrics, PIW and PICP, across combinations of soil cover and depth strata. For topsoil, PIW varies significantly across land covers: it is very high for *water areas and wetlands* (over 100 kg/m^3), high for *woodland*, *shrubland* and *grassland* (exceeding 50 kg/m^3), relatively lower for *cropland* and *bare land and lichens or*

moss (around 30 kg/m^3), and lowest for *artificial land* (around 20 kg/m^3). PIW decreases with depth for all land covers up to 100 cm.

PICP also varies across land covers in topsoil, with over-pessimistic estimates for *artificial land* and *water areas and wetlands* (PICP = 1), and slightly over-optimistic estimates for other land covers (PICP < 0.95). In *cropland*, PICP remains stable with increasing depth until the 100–200 cm interval, where it reaches 1. For *grassland*, PICP improves at depths of 50-100 cm (0.90 to 0.95) but drops sharply to 0.71 for depths greater than 100 cm. In *woodland*, PICP drops to 0.82 at depths of 20-50 cm and increases to 1 in the 50-100 cm soil depth interval.

3.3.4 Maps examination

Fig. 3.14 (top panel) presents the pan-European maps of SOCD predictions and the corresponding prediction intervals (PIs) for the period 2020–2022 for topsoil at a depth of 0 cm to 20 cm. In general, SOCD predictions are higher at higher latitudes and lower at lower latitudes, with the lowest values observed in Spain and Türkiye. In addition to the SOCD prediction maps, the P025 map (lower bound of PI: 2.5th percentile) and the P975 map (upper bound of PI: 97.5th percentile) are also shown in the top panel of Fig. 3.14. The P025 map exhibits minimal spatial variation, while the P975 map mirrors the spatial patterns of the SOCD predictions, showing higher values in regions with greater latitude and altitude. Consequently, PIs are wider in areas with higher SOCD predictions, consistent with observations in Fig. 3.12.

The second and third rows of Fig. 3.14 provide three zoomed-in examples of SOCD prediction maps for the period 2020–2022 at the topsoil level, centered on three specific locations: (A) a peatland site in the Netherlands, (B) a woodland area in Finland, and (C) a cropland site in Italy. The peatland site (A) exhibits the highest SOCD, while the cropland site (C) shows the lowest, with the woodland site (B) falling in between. The bottom row illustrates SOCD variation across different depth intervals for the three locations. A general trend of decreasing SOCD values with increasing depth is observed at all three sites, with spatial variation diminishing as depth increases. Although SOCD decreases with depth, it remains relatively high in deeper layers at the peatland site, while dropping to very low levels in the cropland. The PIW is smallest in the cropland and largest in the peatlands. For the peatland site, PIW remains relatively constant across depths, whereas for the woodland and cropland sites, it decreases with depth but increases again at the 100–200 cm interval.

Fig. 3.15 illustrates the temporal changes in SOCD on the farm scale for a location in the NUTS3 region of Unterallgäu, Bavaria, Germany, near Mindelheim Stadt, including both cropland and woodland. To emphasize the contrast in SOCD changes, a binary color map was applied with the color range compressed between 16 kg/m^3 and 36 kg/m^3. The figure shows that during the period 2004 to 2012, SOCD values decreased in the forested area

while remaining relatively stable in the cropland area. After 2012, SOCD values increased. Overall, the changes in SOCD are relatively small. Two points, `X` and `Y`, were selected from cropland and woodland, respectively, to illustrate the time series of SOCD predictions and corresponding PIs at the pixel level, as shown in the lower panel of Fig. 3.15. Over time, both points exhibited very small variations, especially compared to their respective PIs. The uncertainty level is notably high, with a PIW of approximately 15 kg/m^3 for cropland and 35 kg/m^3 for woodland.

Although pseudo-zero points are excluded from the calculation of performance metrics, we can still observe their impact in the map predictions (Fig. 3.16). In a mountainous region of the Alps, the model tends to overestimate SOCD values on rocky mountain tops when pseudo-zero points are not included.

Even with the inclusion of pseudo-zero points, the Alpine mountain regions remain among the areas highlighted by the extrapolation risk maps, as shown in Fig. 3.17. This probability map, derived using the Isolation Tree method, indicates higher values for areas with a greater likelihood of being unfamiliar to the model—implying a higher risk of extrapolation. The recommended threshold for determining whether an area is classified as extrapolated is 64%, represented by red in the figure. Other identified areas include the western mountainous regions of Norway, most of Iceland, and the eastern part of Türkiye.

3.3.5 Spatial aggregation

Fig. 3.18 illustrates the effect of spatial aggregation. The top left panel shows a satellite image depicting the landscape of the AOI, with red dots indicating the sample measurements available within this AOI. The top right panel displays the SOCD standard deviation derived from the 95% PI map. The bar plot in the bottom left panel compares the uncertainty estimated at the regional level from three approaches: spatial aggregation based on samples, spatial aggregation based on SOCD predictions, and the simple average of pixel-level standard deviations within the AOI for the period 2012–2016 (the only period with available samples). The plot demonstrates that spatial aggregation effectively reduces the uncertainty level from approximately 7.9 kg/m^3 to 2.7 kg/m^3, which is only slightly higher than the aggregated uncertainty derived from samples. The bottom right panel presents the time series of spatial aggregates and uncertainty for this AOI, with estimates from model predictions (mean: 27.5, standard deviation: 2.7) and design-based samples (mean: 27.9, standard deviation: 1.9), which are comparable.

3.4 Discussion

3.4.1 Key variables to explain SOCD variations

Our results show that `soil_depth` is the most important feature, aligning with the findings of Hengl et al. (2017) and Sothe et al. (2022). The vertical variation of SOCD and its negative correlation with soil depth have been widely reported for mineral soils (Gross and Harrison, 2019; Jobbágy and Jackson, 2000; Lorenz and Lal, 2005), and is also evident in our training data (Fig. 3.3), Shapley analysis (Fig. 3.6) and map predictions (Fig. 3.14). While this general trend is expected, the near-linear negative relationship observed in Fig. 3.6 may be influenced by the limited availability of data from deeper soil layers, which could reduce the model's ability to capture more complex patterns at depth. An exception is also observed in the deepest layer (100–200 cm) for cropland and grassland in the training data, where SOCD increases slightly—likely also a result of data scarcity introducing bias in those specific strata (Fig. 3.3). This high feature importance likely reflects `soil_depth`'s role as the only variable distinguishing samples from different soil horizon layers at the same location and year, due to the absence of depth-specific predictor layers. Most available covariates to date primarily represent surface environmental processes, such as vegetation and climate. While Helfenstein et al. (2024a) developed depth-specific peat soil maps for the Netherlands, such data are still lacking at the EU scale. Both data scarcity and potential for deep SOC storage emphasize the importance of including depth as a dimension in SOC mapping, which is critical to evaluate climate solutions and projecting terrestrial climate change feedbacks (Hicks Pries et al., 2023; Lorenz and Lal, 2005).

Following `soil_depth`, vegetation indices form the second most important feature group. The positive correlation between vegetation indices and predicted SOCD reflects the critical role of vegetation in soil formation and organic matter accumulation (Kooch et al., 2022). FAPAR, which quantifies the fraction of solar radiation absorbed by plants, emphasizes vegetation activity and demonstrates reduced saturation effects compared to NDVI (Myneni and Williams, 1994; Robinson et al., 2018). BSF, on the other hand, reflects bare soil exposure over a year. While they all rank highly in feature importance, further investigation is needed to know the extent to which they provide complementary information, given their mutual correlation. Vegetation conditions are known to influence SOCD through factors such as natural versus managed vegetation, root distributions, and the allocation of above- and below-ground biomass (Hu et al., 2018; Jobbágy and Jackson, 2000; Ramesh et al., 2019; Schneider et al., 2021). These influences are reflected in both the training dataset (Fig. 3.3) and the resulting maps (Fig. 3.14). However, the reliance on vegetation features can result in unrealistic predictions in a spatial context. As shown in Fig. 3.15, some vegetated cropland parcels exhibit higher SOCD values than woodland, driven by correspondingly high FAPAR and NDVI values in this region (see supplementary notebook 016_check_feature).

The relationship between climatic variables and SOCD is complex and often non-linear (Fig. 3.6). Numerous studies have suggested the associations between SOCD and climatic variables such as temperature, precipitation, and moisture (Gross and Harrison, 2019; Jobbágy and Jackson, 2000; Kerr and Ochsner, 2020; Luo et al., 2017; Yang et al., 2008). Beyond direct time-series data, pattern- or characteristic-based features derived from time series have shown significant value in predicting SOCD. For instance, `bio3_longterm_m`, which quantifies day-to-night temperature oscillations relative to seasonal changes, is the most important temperature-related feature, despite the inclusion of LST time series. The impact of temperature variation has also been observed in laboratory experiments (Adekanmbi et al., 2022), although further research is needed to understand this phenomenon on a continental scale in Europe. These highlight the importance of feature engineering in extracting meaningful patterns from raw data to improve model efficiency and accuracy in predicting SOCD. Understanding these interactions is crucial for anticipating how SOCD might respond to climate change (Crowther et al., 2016).

Other features that were selected but did not rank among the top ten included topographic features, `lithology.peat_static_c`, and `backscatter.vv_static_m`. As a key soil-forming factor, topography influences soil water status, as well as erosion and deposition processes (Behrens et al., 2010; Jenny, 1994). The feature `lithology.peat_static_c`, the only selected feature from the lithology group, is particularly important in areas with a high probability of being peatland (Fig. 3.7-A). Backscatter information, represented by `backscatter.vv_static_m`, ranked relatively low among the selected features. Its influence on SOCD predictions may come from its relationship with soil moisture (Ceddia et al., 2017). Although not among the most important features, these variables contribute to the prediction by providing unique information not captured by other feature groups.

In addition to examining Shapley values globally and comparing them across different locations, another valuable question is how Shapley values vary for the same point across time and depth (Fig. 3.6). Such an analysis could potentially provide insights into prediction variability over temporal and depth dimensions. However, addressing this question would require a more thorough and dedicated research effort, which is beyond the current study's focus. We expect further research to explore this important topic.

By quantifying the contribution of each feature to a model's predictions, Shapley values provide insights into the model's decision-making process. However, it is important to note that these values do not imply causality. The interpretations provided (Fig. 3.6) are conditional on the data and features used. The dataset does not fully represent continental Europe, as shown in Fig. 3.2. Additionally, while an extensive set of feature layers was selected to capture the underlying mechanisms of SOCD, these features have inherent limitations. They offer only a sampled approximation of true conditions across space and time, potentially missing critical information. Furthermore, the process of generating these

features involves errors and uncertainty. Therefore, as emphasized by Wadoux et al. (2023), these features should be regarded as proxies rather than definitive representations.

3.4.2 Interpretation of overall model performance and spatial patterns

Overall, the model shows good predictive performance, with predicted values closely aligning with observations (Fig. 3.8). Compared to the independent test results, performance metrics under ISIW-CV show only minor declines. This suggests that, despite the clustered nature of the sampling data, its representativeness is not significantly compromised—likely due to the broad spatial coverage provided by LUCAS across the EU. While the overall performance is good, the model tends to overestimate low SOCD values and underestimate the high values, consistent with findings from previous studies (Feeney et al., 2022; Hengl et al., 2017; Sanderman et al., 2018; Sothe et al., 2022). This may partly reflect the smoothing effect typical of RF, an ensemble learning algorithm.

To improve the model's performance in capturing low SOCD values, two approaches are adopted: (1) inclusion of pseudo-zero points, and (2) application of a log-transformation. In soil surveys, deeper layers and non-soil areas—such as sand dunes and bare rocks—are often overlooked due to practical constraints. This results in a lack of representation for "absence of soil" scenarios in the feature space, leading to model confusion at the lower end of the SOCD scale. To address this, pseudo-zero points were introduced. Although excluded from performance metric calculations—both to rely solely on real measurements and to avoid skewing variance-based metrics like R^2 and CCC due to their clustering at the origin—they still influence the predictions (Fig. 3.16). Note that these pseudo-zero values are only used at the surface level. Log-transformation also helps mitigate this issue, as suggested by Hengl et al. (2017) and Guevara et al. (2020). When the model is trained on the original linear scale, the overestimation of SOCD values can also be attributed to the loss function of the RF regressor, squared error, which places greater weight on extreme values and emphasizes high-variance points, often linked to higher SOCD. Log-transformation can mitigate this but may reduce accuracy for high SOCD values (see supplementary notebook 005_transformation_comparison). From a user perspective—such as for agricultural communities—accurate prediction of mineral soils is often more critical, making log-transformation a practical choice in such contexts.

When observed SOCD values exceed approximately 75 kg/m^3, the model outputs remain below this value (Fig. 3.8). This is further reflected in its uncertainty estimation. As shown in Fig. 3.12, PIW become significantly wider for high SOCD values, while PICP exhibit instability. This is likely due to the limited availability of datasets with high SOCD training points in certain years (Fig. 3.2). Many legacy point datasets (e.g., Batjes et al., 2017) over-represent agricultural or mineral soils, which can introduce systematic bias. This issue becomes less pronounced when more high-SOCD data are available (Helfenstein et al., 2024b; Sothe et al., 2022). Agricultural and mineral soils behave differently from

organic soils, as emphasized by de Brogniez et al. (2015) and Feeney et al. (2022). While Helfenstein et al. (2024b) utilized peat occurrence across different times and depths as a feature to effectively distinguish between mineral and organic soils, such a layer is still lacking for continental Europe. Data quality may also contribute to this issue. Ziche et al. (2022) reported inconsistencies in the LUCAS 2015 dataset for forest soils, where some samples mix organic layers with mineral soils, potentially leading to underestimation. While this study uses LUCAS 2018 data instead of LUCAS 2015, further investigation is necessary to rule out similar problems.

The spatial patterns of both SOCD predictions and PI estimates generally align with the results of Poggio et al. (2021) and Wesemael et al. (2024). However, discrepancies are observed in regions such as part of the Alps, the northwest of Scotland, the mountainous areas of western Norway, and Iceland. These regions are characterized by bare surfaces, snow coverage and high altitudes. The discrepancies are likely attributable to the inclusion of pseudo-zero points.

The quantified PI appears to be primarily influenced by the corresponding SOCD values, with higher predicted SOCD values associated with wider PIs (Fig. 3.12). This relationship is also evident in Fig. 3.14, where the P975 map approximately mirrors the spatial variation of SOCD predictions, while the P025 map remains relatively low across most regions. In contrast, the spatial pattern of extrapolation risk probability (Fig. 3.17) differs and aligns more closely with the data availability distribution (Fig. 3.2) with highest risk of extrapolation likely happening in high altituded. Regions with limited data availability and high mountains, such as Türkiye, Norway, Iceland, and Switzerland, exhibit relatively high extrapolation risk probabilities. Interestingly, Ukraine, despite missing data, shows a low extrapolation risk probability, possibly because its feature space is similar to other well-represented regions, such as Poland or Germany. More analysis is needed to confirm this hypothesis. A simple correlation analysis quantitatively confirms the weak correlation between PIW and extrapolation risk probability (see supplementary notebook 011_aoa_isolation.tree), a finding that has also been reported by Hateffard et al. (2024). Further investigation is required to provide an explanation, which is beyond the scope of this study.

3.4.3 Model performance across land covers and soil depths

Overestimation and underestimation issues are particularly pronounced for the most under-represented land covers, *water areas and wetlands* and *artificial land*. Exhibiting the highest average SOCD values, fewest observations are available for *water areas and wetlands*, and the model significantly underestimates these values. Conversely, the model overestimates SOCD values for *artificial land*. Their extremely limited data availability make the other performance metrics less informative.

Shrubland and *bare land and lichens or moss* also have limited data availability, but not spare enough to exlude CCC analysis. *Shrubland* shows poor prediction accuracy and high model uncertainty, while *bare land and lichens or moss* achieve the best prediction accuracy and minimal model uncertainty among all land covers. This contrast may be explained by differences in the feature space. *Shrubland* is similar to other vegetated land covers, such as *grassland*, in terms of above-surface characteristics—despite differences in rooting systems and below-ground SOCD distribution (Jobbágy and Jackson, 2000; Lorenz and Lal, 2005). However, our feature set lacks information that captures these below-ground distinctions. In contrast, *bare land and lichens or moss* can be more easily distinguished due to its distinct and detectable surface features.

Cropland, *grassland*, and *woodland* are the three land covers with sufficient data at deeper soil layers. All show moderate prediction accuracy at the topsoil level (0–20 cm) and improved accuracy at the 20–50 cm soil depth interval, as reflected in narrower PIs. This improvement may result from reduced spatial variability in deeper soils—allowing limited data to better represent the underlying patterns—or it may simply reflect that the reduced data availability at these depths has already lost its informative value. Only *grassland* and *cropland* have sufficient data for analysis beyond the 50 cm soil depth. However, the variation in their prediction accuracy across depths still requires more detailed analysis and additional data to fully understand.

Compared to the model's performance on the entire test dataset, its performance across specific land cover types and soil depth intervals is suboptimal for both prediction and uncertainty estimation. This is expected, as global model performance often declines when evaluated on stratified subsets—particularly those with limited data availability. As shown in Fig. 3.2 and Fig. 3.17, additional data from underrepresented regions, such as Norway, Switzerland, and Türkiye, are needed to fill current gaps and avoid extrapolation risk. Improved data availability for deeper soil layers, particularly for the 100–200 cm interval, is essential to reduce uncertainty at these depths. This could involve both collecting more measurements and inserting pseudo-zero values derived from horizon surveys, which are often recorded separately from SOC measurements. Furthermore, increasing the representation of high-SOCD samples would be highly beneficial.

3.4.4 Reduced uncertainty at larger spatial support

In this work, model-predicted uncertainties are presented as PI maps along with SOCD predictions. We consider this necessary for several reasons:

1. it prevents misleading summary metrics by highlighting prediction uncertainty (Veronesi and Schillaci, 2019);
2. it identifies conditions where the model may fail or produce unreliable predictions, aiding responsible decision-making (Breure et al., 2022; Wadoux et al., 2020b);

3. it enables more thorough model comparisons, especially when a model with reliable uncertainty estimates is preferred (Szatmári and Pásztor, 2019; Veronesi and Schillaci, 2019);

4. it allows for uncertainty propagation when the soil map is used in models (Heuvelink, 1998; Schmidinger and Heuvelink, 2023).

Although a 90% probability PI is commonly used (e.g. Arrouays et al. (2020b), Malone et al. (2017), Odgers et al. (2015), and Schmidinger and Heuvelink (2023)), we opt for a 95% PI because it corresponds to ±2 standard deviations. This choice simplifies communication and comparison, although a correction factor may still be required if the data distribution deviates from normality (see the supplementary notebook 013_variogram_spatial.aggregate).

The uncertainty level (95% PI) is high at the pixel level, especially for pixels with high SOCD values. In some cases, the PIW even exceeds the corresponding predicted SOCD values (see Figs. 3.14 and 3.15). However, this uncertainty can be significantly reduced by spatial aggregation on larger supports, as shown in Emick et al. (2023), Heuvelink et al. (2020), Szatmári et al. (2021), and Wadoux and Heuvelink (2023) and Szatmári et al. (2024b). Larger spatial supports result in smaller uncertainties associated with aggregated predictions due to the weak spatial autocorrelation of standardized map errors. This approach is valuable for revealing temporal changes and regional comparisons using a larger support, which are crucial for policy and decision-making.

The regional averages and uncertainties derived from the model predictions are consistent with those estimated from the samples, although not necessarily more accurate. However, sample-based estimation depends on the availability of sufficient measurements, which is not always feasible (Emick et al., 2023). For example, as shown in Fig. 3.18, measurement samples are available only for the 2012–2016 period, limiting the formation of long-term time series. In such scenarios, model predictions serve as a viable alternative.

This study demonstrates the method and impact of spatial aggregation using the maps presented at the NUTS3 regional level for a single region. For a more comprehensive evaluation of the impact of spatial aggregation, it would be beneficial to test different spatial supports, as done by Szatmári et al. (2021) and Szatmári et al. (2024b). In addition, exploring the aggregation in various regional classifications, such as soil regions, represents a valuable direction for future research.

3.4.5 Challenges in detecting temporal SOCD changes

In our study, static features were found to have a stronger influence on model predictions than dynamic ones, aligning with findings from previous research (Heuvelink et al., 2020; Szatmári et al., 2021; Szatmári et al., 2024b). This dominance is also reflected in the largely stable map predictions over time (Fig. 3.15). It can be partly explained by the

nature of many key features—such as topography—which change very slowly over time and are typically only available in static form. Even when both static and dynamic versions of a variable were available, the model often favored the long-term (static) representations. This preference for static features may reflect the slow-changing nature of SOCD (Poeplau and Don, 2013), whose changes may be too subtle to stand out over a 20-year period. At the same time, spatial variability in SOCD across continental Europe is strong and consistent, which can overshadow subtle temporal dynamics. For example, while high NDVI values are generally associated with high SOCD across space, changes in NDVI can occur rapidly, whereas SOCD responds much more gradually and may not show immediate change. This indicates that predictions are driven primarily by spatial patterns rather than temporal variation. This suggests that while the model performs well at capturing the spatial distribution of SOCD, its ability to detect changes over time—especially subtle or short-term changes—is more limited.

Furthermore, in our spatiotemporal model, each observation—defined by a unique combination of time, location, and depth—is treated independently. This means the model does not explicitly capture temporal correlation. Instead, it can only infer temporal patterns from the values of dynamic predictors. This may further limit its sensitivity to temporal dynamics and cause it to over-focus on patterns learned from broad spatial scales. Temporal correlation includes not only the autocorrelation of SOCD itself but also its relationships with the time series of environmental predictors. Some studies have attempted to incorporate such relationships in DSM to distinguish subtle temporal changes from dominant spatial patterns and to account for the delayed response of SOCD to environmental drivers. For example, Heuvelink et al. (2020) introduced a decayed NDVI feature to account for a 36-year lag, while Helfenstein et al. (2024b) used historical land-use maps to incorporate the most frequently occurring land-use types from years prior to the prediction point in the Netherlands. However, such approaches remain rare, largely due to the limited availability of repeated sample measurements, the intensive effort needed to compile long-term datasets, and the lack of established methods for integrating temporal correlation into model structures.

Among the dynamic features—which introduce the temporal dimension into the models and resulting maps—climate-related variables generally showed higher importance than vegetation-related ones. While climate parameters are traditionally defined as long-term summaries of weather conditions, in this context, "climate-related features" also include higher-resolution time series, such as monthly precipitation and LST, which ranked higher than vegetation indices. This pattern may reflect two factors: first, the underlying environmental dynamics, as variables like precipitation and LST tend to fluctuate more over time than vegetation cover; and second, the higher temporal resolution at which climate-related data are typically available, offering richer input for the model. Notably, precipitation was the most important among all dynamic features, possibly due to its more direct influence on soil properties. Another factor is that climate features

often lack consistent long-term counterparts, while vegetation indices typically exist at multiple temporal resolutions. These observations also highlight the need for caution when interpreting feature importance. Although it can offer insights into model behavior and potential environmental processes, feature importance is also shaped by the way input layers are structured and the quality of the underlying data. As such, it should not be overinterpreted as a direct reflection of ecological or pedological processes, but rather understood as a product of both environmental relevance and data representation.

The model's temporal transferability was assessed using LOYO-CV, which resulted in lower performance metrics compared to the independent test set. This decline is likely due to the exclusion of specific datasets when individual years are omitted from the training data. For instance, the LUCAS SOCD dataset, which is only available for 2018, is entirely excluded when this year is left out, resulting in more spatially clustered training data. Consequently, the evaluation of temporal transferability is partly influenced by the challenges associated with spatial clustering. The challenge of evaluating temporal transferability is further illustrated in Fig. 3.9. The figure shows that low training data availability does not consistently lead to poor performance. Instead, performance appears to be more influenced by the proportion of data points across cropland and non-cropland areas. Since the temporal dimension is introduced primarily through dynamic features, we argue that model performance depends more on representativeness in the feature space than on the temporal coverage itself.

Consistent, repeated soil sampling is critical for assessing model performance in temporal dimensions, analyzing temporal correlations in SOCD and enabling their integration into modeling approaches. While Helfenstein et al. (2024b) assessed model's accuracy across time using datasets spanning 1952 to 2018 in the Netherlands, extending such analyses to larger spatial extents remains challenging due to the lack of such validation data with repeated SOCD measurements. Unlike established networks such as Fluxnet for flux measurements (https://www.europe-fluxdata.eu/home/sites-list) or the International Soil Moisture Network (https://ismn.earth/en/), there is currently no extensive long-term monitoring network for soil, apart from LUCAS. To date, only three rounds of repeated SOC measurements are available. Thus, in this study, the temporal model uncertainty and the ability to predict SOCD changes were mainly analyzed through estimated PIs, assuming that model predictions are valid and their uncertainty is effectively captured by the PIs, which serve as a reference to identify potential SOCD changes. As discussed in the previous section, pixel-level PIs can be quite high, making it challenging to reliably identify areas with statistically significant SOCD changes. Aggregating predictions over larger spatial supports reduces prediction uncertainty, facilitating the identification of statistically significant SOCD changes over larger spatial extent (Szatmári et al., 2021; Szatmári et al., 2024b).

In summary, given the current time frame of 2000–2022, the lack of repeated samples, the absence of temporal correlation in the model, and the predominance of spatial patterns over temporal ones in the model, drawing reliable conclusions about SOCD changes from model predictions remains challenging. As highlighted by (Broeg et al., 2024b) and Helfenstein et al. (2024b), even though the spatio-temporal model demonstrates good general performance, this does not necessarily ensure comparable temporal accuracy in predicting SOCD changes or trends. Moreover, the estimated uncertainty of SOCD predictions cannot be directly translated into the uncertainty of SOCD changes, as this requires accounting for cross- and spatial correlations in prediction errors (Helfenstein et al., 2024b). Consequently, using such a model for temporal monitoring or detecting SOCD change should be approached with caution. Although PIs can help assess the reliability of predicted SOCD changes, areas where changes exceed the PIs—or where consistent PI patterns align with the direction of change—should be interpreted as indicators of potential change, not as confirmation that real-world change has occurred.

3.4.6 Future prospects

This mapping framework provides a baseline time series of spatially continuous, high-resolution SOCD maps for continental Europe. However, as briefly covered in previous sections, several potential improvements have been identified, some of which are already included in our implementation plan:

- **Include more representative soil data to fill spatial and temporal gaps:** More frequent sampling over time, along with additional data to improve representativeness in both spatial coverage and feature space, is essential for increasing the model's generalization ability (Yuzugullu et al., 2024). Continued support for, and expansion of, initiatives like LUCAS is critical to establishing a consistent, long-term monitoring framework capable of providing such data. The forthcoming LUCAS 2022 dataset, providing a fourth round across the EU, marks an important step in this direction.
- **Pay equally attention to BD and coarse fragment data:** While SOC content data are relatively abundant, corresponding BD and coarse fragment data are often overlooked, , limiting the ability to accurately quantify SOCD—especially in the context of climate change and carbon farming (Poeplau et al., 2017).
- **Introduce more relevant, appropriately developed features:** Addressing process gaps requires the inclusion of representative features—such as those capturing deeper soil conditions, distinguishing organic soils, and introducing temporal correlations.
- **Develop a map for depth to bedrock:** This would enable more complete estimation of SOCD across the full soil profile and help users interpret SOCD maps

at different depths. However, achieving this will require significant effort and data integration.

- **Explore of new model structures and approaches**: Experiments with alternative model approaches or structures, such as training separate models (de Brogniez et al., 2015; Wesemael et al., 2024) or hierarchical model (Nussbaum et al., 2023) for each soil type and subsequently assessing their accuracy, could potentially improve model performance.

- **Maintain and update the maps:** The maps are planned to be updated regularly using a *'re-analysis'* approach, as is common in climate science (e.g., Dee et al. 2014). We recommend that users base their analyses on the latest version of the data. Although large-scale changes between versions are unlikely, regional updates may occur where new data begin to fill gaps, potentially revealing new insights. This update could also be challenging, as many feature layers used in mapping are no longer updated. For example, CHELSA climatic features, which rank high in importance, is no longer updated beyond 2018 (https://chelsa-climate.org/timeseries/). Suitable alternatives need to be identified.

- **Optimize computational efficiency:** Deriving PIs and extrapolation risk probability is essential for ensuring map reliability but significantly increases computational costs, particularly for finer spatial or temporal resolutions. To minimize compromises in map accuracy, efforts should be directed toward improving computational efficiency.

- **Seek for ways to validate block predictions:** Validating block predictions over multi-year periods and depth intervals is challenging. Direct sampling for such validation is infeasible, so legacy data and CV within the spacetime block of interest must be leveraged. Distinguishing between point predictions (e.g., for 2022 at a depth of 10 cm) and block predictions (e.g., for 2020–2024 at a depth interval of 0–20 cm) is critical. Validation of block predictions would require composite samples over multi-year periods, which may not always be feasible.

- **Improve the communication of map uncertainties to the users**: For specific locations or use cases, the integration of SOCD maps, uncertainty maps, extrapolation risk maps, and local knowledge is essential for mitigating potential risks associated with the use of this map product and similar tools in soil carbon farming, as well as for addressing credibility concerns (Moinet et al., 2023; Paul et al., 2023). Efforts have been made to provide users with as thorough accuracy-related information as possible. While we emphasize the importance of transparency and thoroughness regarding model and map accuracies, we also recognize that this level of detail may pose challenges for users (Padarian and McBratney, 2023). Therefore, there is a need to develop improved methods for effectively communicating this information to users.

3.5 Conclusions

We developed a comprehensive framework for automated predictive soil mapping using RF in a spatiotemporal context, utilizing EO data and a harmonized SOCD dataset as its foundation. This framework maps soil carbon dynamics over an extended period (2000–2022+) at a fine spatial resolution (30 m). Quantitative evaluation demonstrated good overall prediction accuracy, though model performance declined for specific land covers and depth intervals. Validation of PI estimates confirmed that the models effectively capture uncertainty, despite the reduced accuracy for higher soil carbon density values. Exploratory analysis using Shapley values identifies soil depth as the most important feature, followed by vegetation (Landsat bimonthly biophysical indices), long-term bioclimatic variables, and topographic features. While pixel-level PIs are notably wide, spatial aggregation reduces uncertainty significantly, enabling the detection of SOCD changes at larger spatial supports. However, given the current time frame of 2000–2022, the absence of consistent, repeated samples, and the predominance of spatial patterns over temporal ones in the model, drawing reliable conclusions about SOCD changes is still challenging.

Potential applications of the data include: (1) providing a resource for spatiotemporal analysis of land degradation and restoration hotspots, and supporting soil health assessments with caution regarding its limitations; (2) offering a basis for optimizing sampling designs by considering prediction uncertainty; and (3) exploring future soil carbon potential through model extrapolation under different land use or climate scenarios. We aimed to keep the SOCD maps as complete as possible to maximize their usability. However, map performance varies. We recommend that users interpret the maps with caution, considering local knowledge, specific use-case requirements, and the accompanying uncertainty and extrapolation risk layers. Based on accuracy metrics, the maps are most reliable for topsoil applications in cropland, grassland, and woodland. They can still offer valuable insights under other conditions, but should be applied with awareness of their limitations. We plan to update predictions regularly with new data, and also ensure it's up to date. This work was made possible by the LUCAS soil initiative (Orgiazzi et al., 2018) and contributions from several European countries. Continued collaboration among European Commission bodies, national soil agencies, and research institutes is essential to advancing open, high-quality soil data across Europe through the sharing of point data under appropriate agreements.

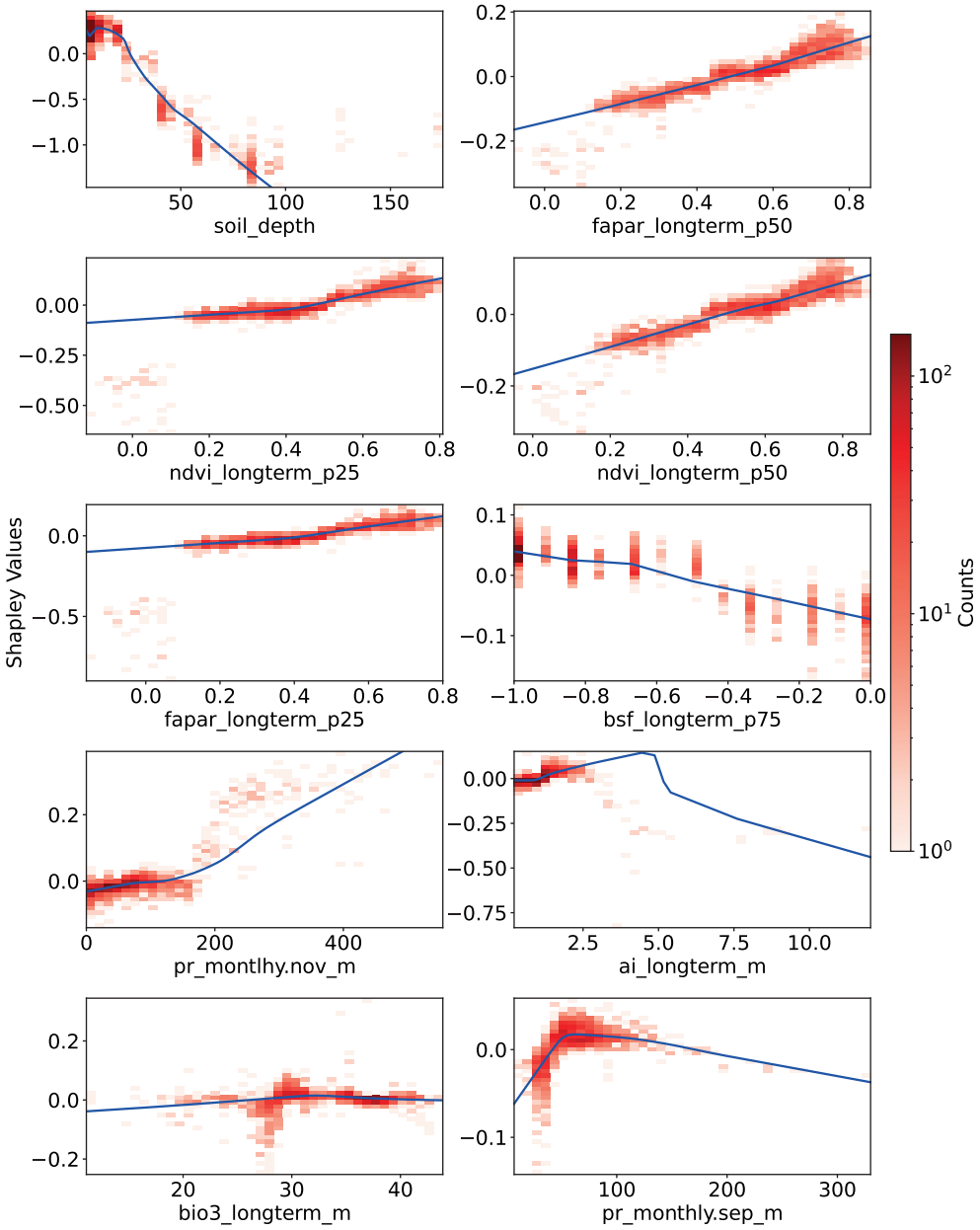


Figure 3.6: These plots represent the relative contribution of each feature to SOCD predictions in the test dataset. The x-axis indicates feature values, while the y-axis represents Shapley values for the respective feature. The blue line illustrates a smoothed curve fitted to the Shapley values for visualization purposes.

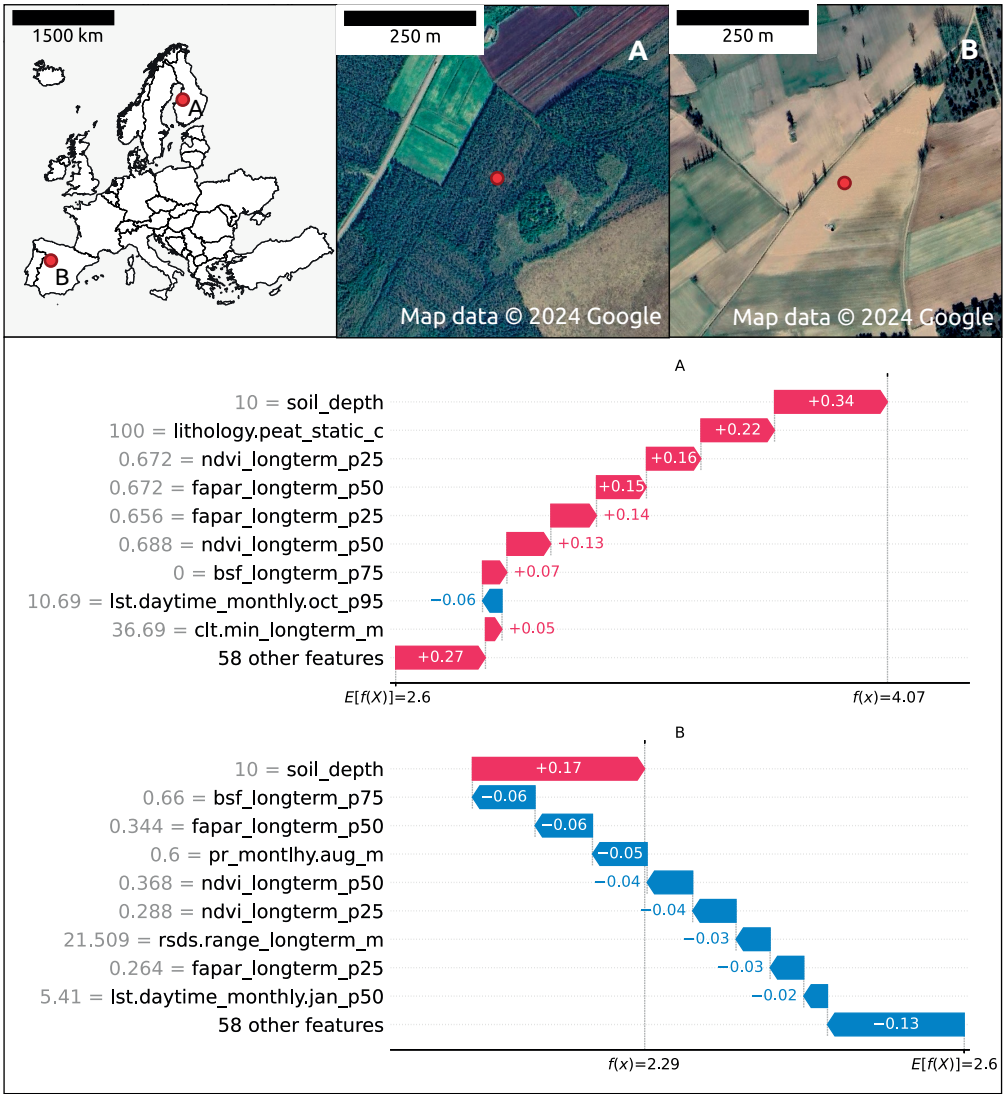


Figure 3.7: Contribution of the top 10 features to SOCD prediction at two spatial locations: (A) forested area in Finland and (B) a an agricultural area in Spain. Red indicates a positive contribution to the SOCD prediction, while blue indicates a negative contribution. Feature names and values are displayed on the y-axis for each location's prediction. The decision track of SOCD predictions is shown on the x-axis, illustrating the progression from the starting point (average SOCD value in the test set) to the final predicted value. Satellite images from © Google Maps (2024, CNES/Airbus, Maxar Technologies), available through https://www.google.com/maps/, last accessed: 22 August 2024.

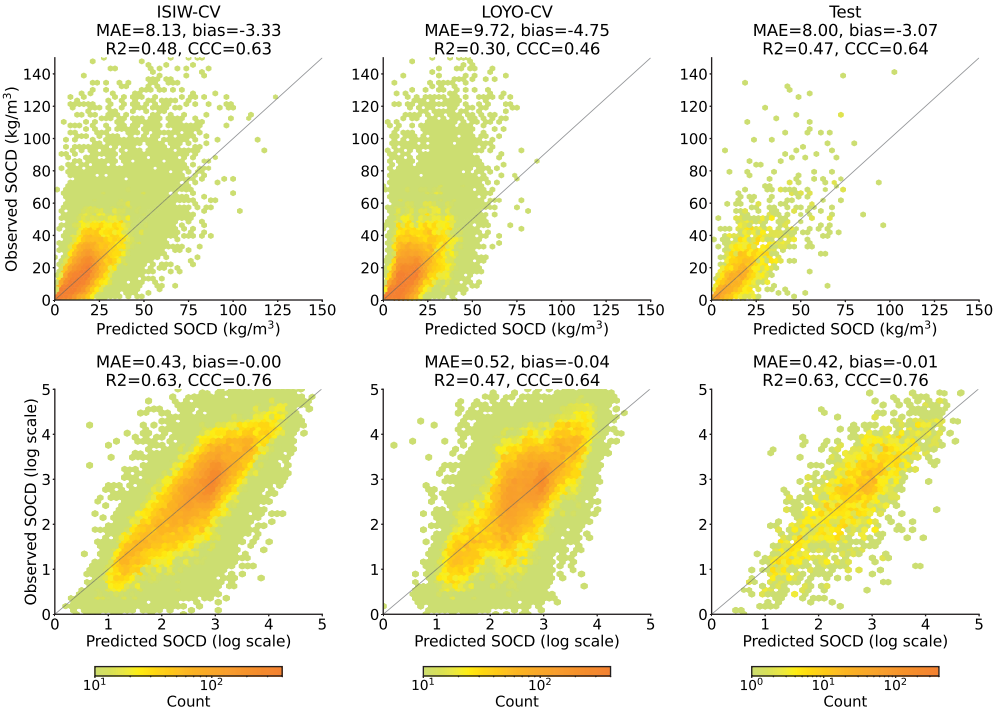


Figure 3.8: Accuracy plots in the original scale (top panel) generated using ISIW-CV on the training set (left), LOYO-CV on the training set (middle), and the independent test set (right), with corresponding performance metrics displayed in the titles. The bottom panel presents the same results in logarithmic scale. Note that SOCD values are truncated at 150 kg/m^3 in the plots for visualization purposes; however, all performance metrics are calculated using the full range of predicted and observed SOCD values (up to 400 kg/m^3 as shown in Fig.3.2), excluding the pseudo-zero points. The same applies to the remaining figures.

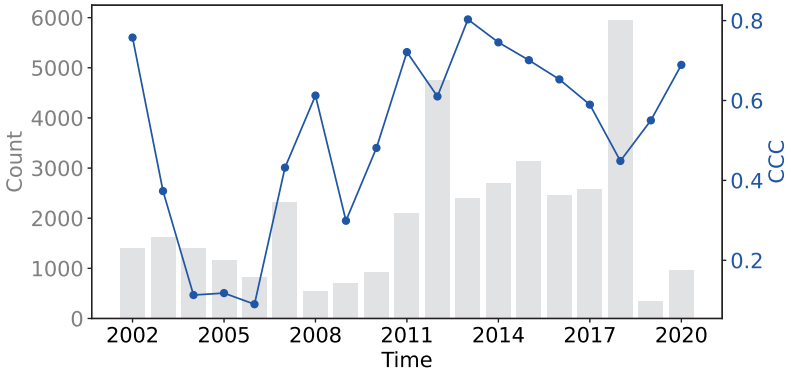


Figure 3.9: The CCC from the test dataset across different years, displayed alongside the data availability in the training dataset.

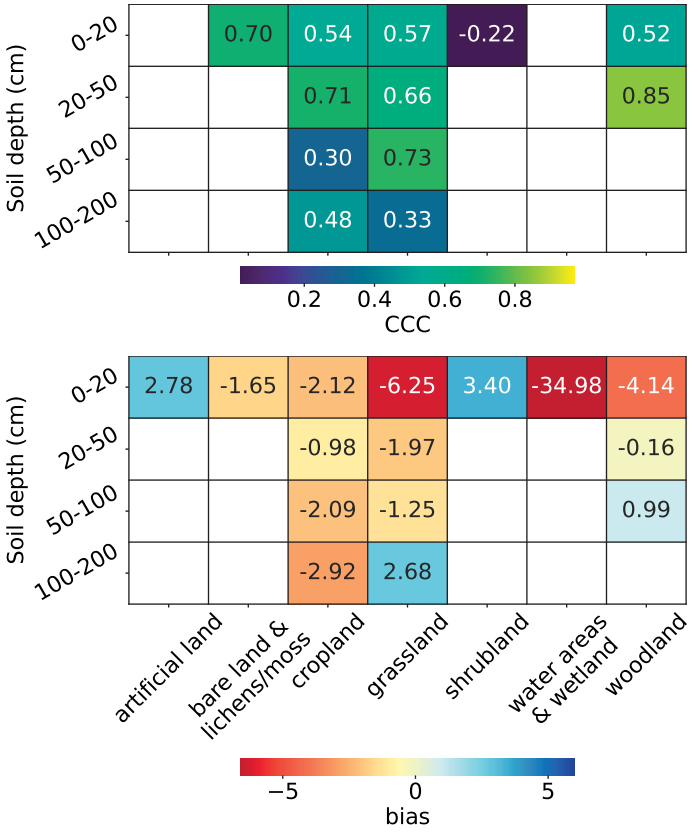


Figure 3.10: Accuracy metrics: CCC (top) and bias (bottom) for each combination of land cover and soil depth strata. Note that CCC is not shown for the 0-20 cm soil depth interval in "water areas and wetland" and *artificial land*, and for the 50-100 cm soil depth interval in *woodland* due to limited data ($n < 5$) in the test set.

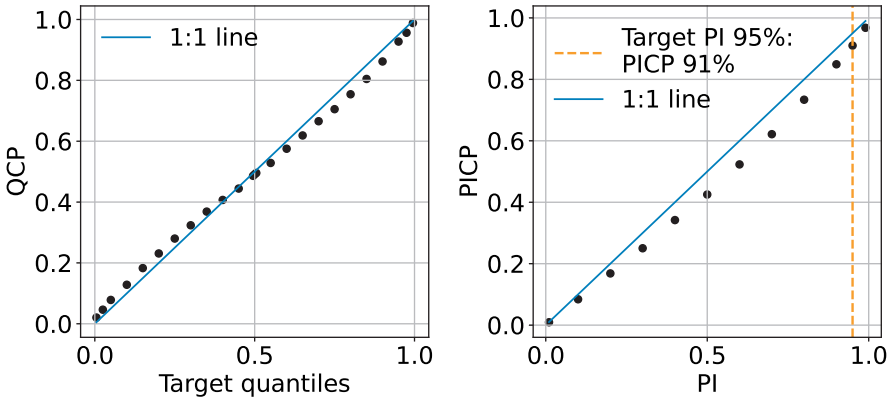


Figure 3.11: Mean QCP reliability plot (left) and mean PICP reliability plot (right) generated in the independent testing from the test set.

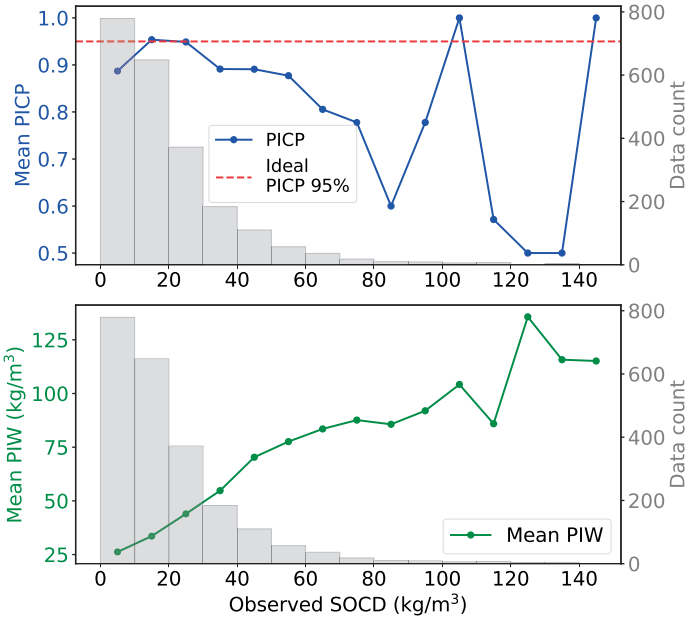


Figure 3.12: Mean PIW and PICP across different observed SOCD intervals, with the corresponding histogram.

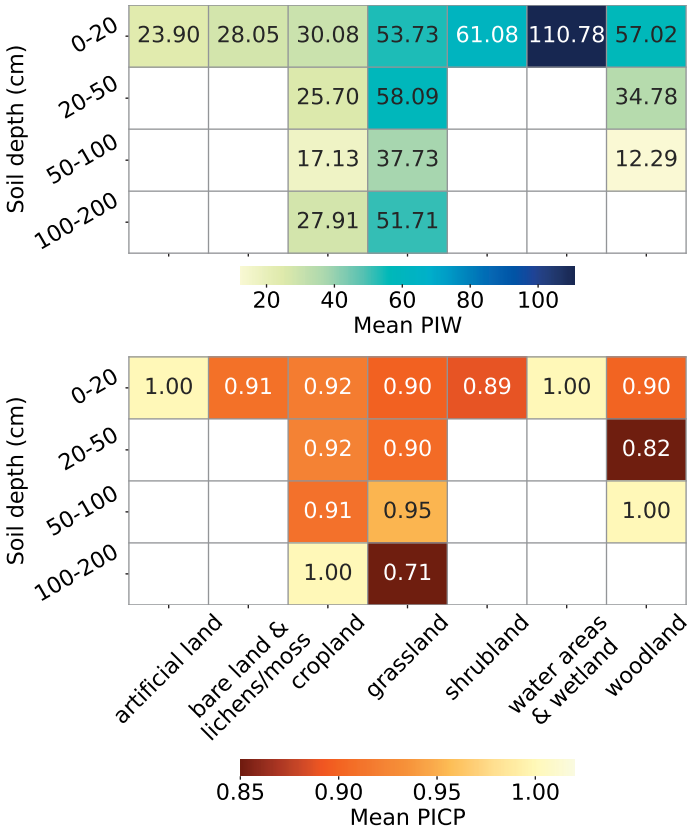


Figure 3.13: Mean PIW (left) and mean PICP (right) across combinations of land covers and soil depth intervals.

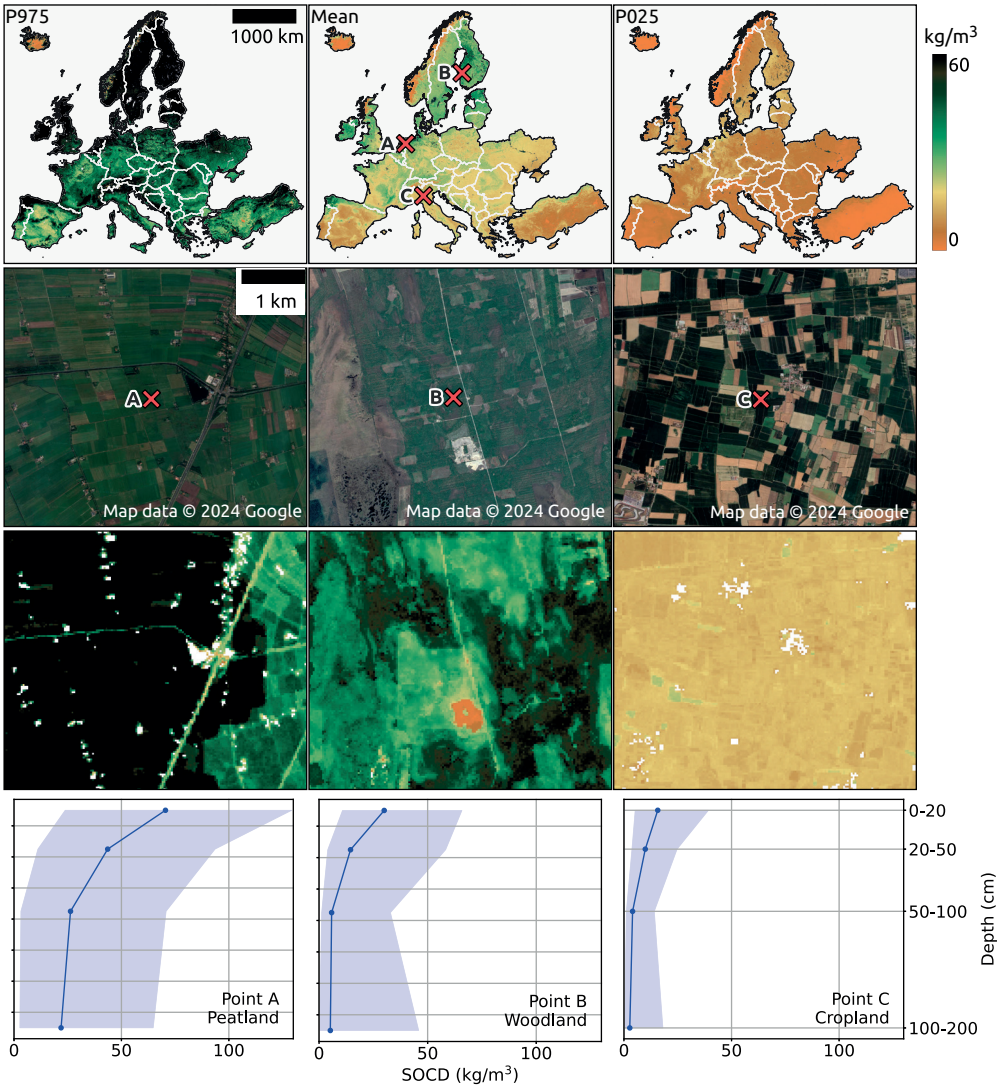


Figure 3.14: SOCD predictions for the topsoil (0–20 cm) across continental Europe from 2020 to 2022 (available at https://doi.org/10.5281/zenodo.13754343). The top row presents an overview of the predictions (middle), including the upper bound (P0.975 at left) and lower bound (P0.025 at right). The middle two rows provide zoomed-in SOCD predictions and corresponding satellite images for the same period at three specific locations: (A) a peatland site in the Netherlands (lon: 6.177, lat:52.583), (B) a woodland area in Finland (lon: 22.531, lat: 62.351), and (C) a cropland site in Italy (lon: 10.339, lat: 45.101). The bottom row illustrates the variation of SOCD predictions with depth at these three points, with the pale blue shading indicating the corresponding PI. Google Maps (2024, CNES/Airbus, Maxar Technologies), available through https://www.google.com/maps/, last accessed: 30 August 2024.

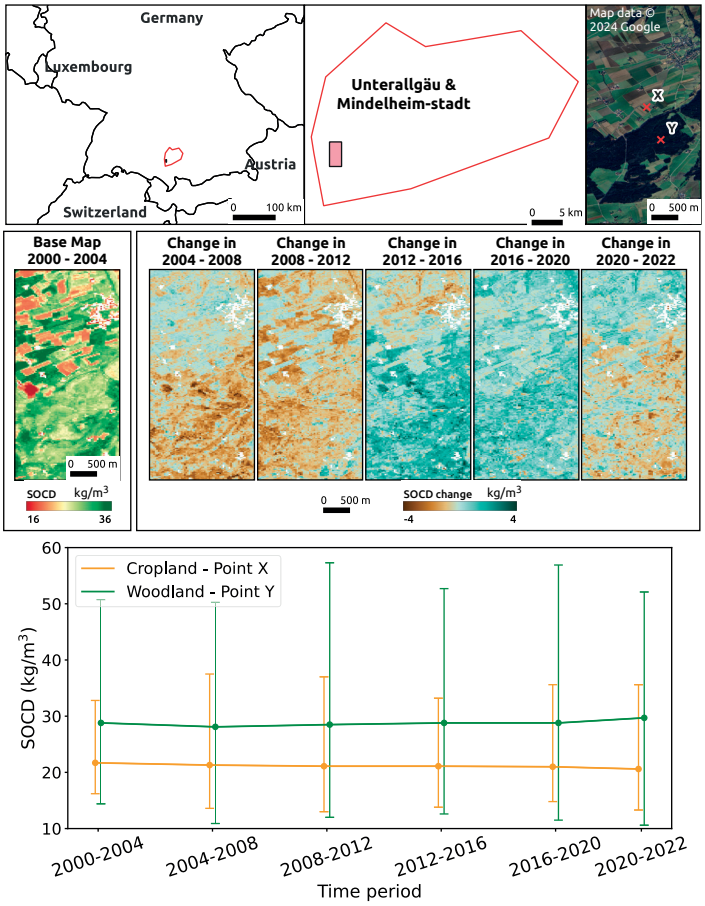


Figure 3.15: Temporal changes in SOCD at the farm scale. The top panel displays the location of the study area in the NUTS3 region of Unterallgäu, Bavaria, Germany, near Mindelheim Stadt, along with its landscape from satellite images. The middle panel presents SOCD predictions for 2000–2004 as the baseline map, followed by maps showing changes in SOCD for each subsequent 4-year interval (e.g., 2004–2008, 2008–2012), calculated as the difference from the preceding interval. The middle panel presents SOCD predictions in 2000–2004 as a base map, and show the changes happened in the consecutive years compared to previous year intervals from 2004 to 2022. The bottom panel shows the time series of SOCD predictions and corresponding 95% PI for two points: `X` (cropland, long: 10.159, lat: 47.937) and `Y` (woodland, lon: 10.163, lat: 47.929). Google Maps (2024, CNES/Airbus, Maxar Technologies), available through https://www.google.com/maps/, last accessed: 30 August 2024.

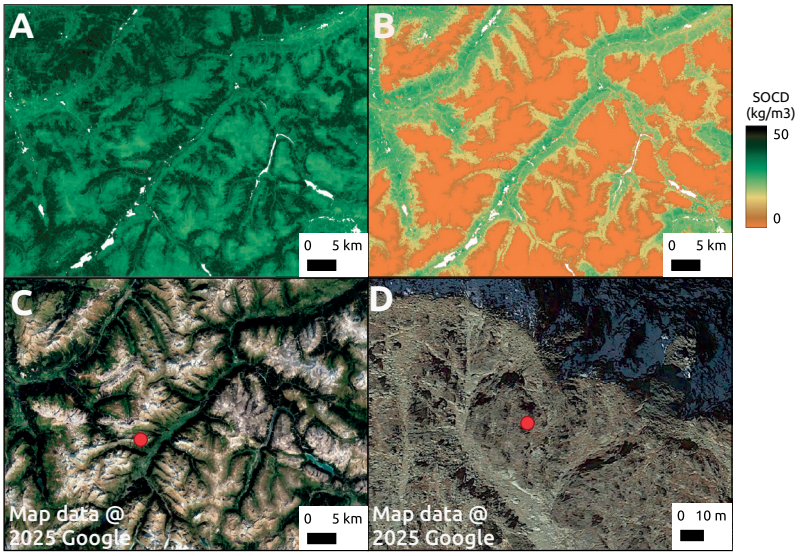


Figure 3.16: Comparison of model predictions without (A) and with (B) the inclusion of pseudo-zero points during training. The comparison is shown for a region in the Alpine mountains (C), with a zoomed-in view of a rocky mountain peak area (D). Google Maps (2024, CNES/Airbus, Maxar Technologies), available through https://www.google.com/maps/, last accessed: 16 April 2025.

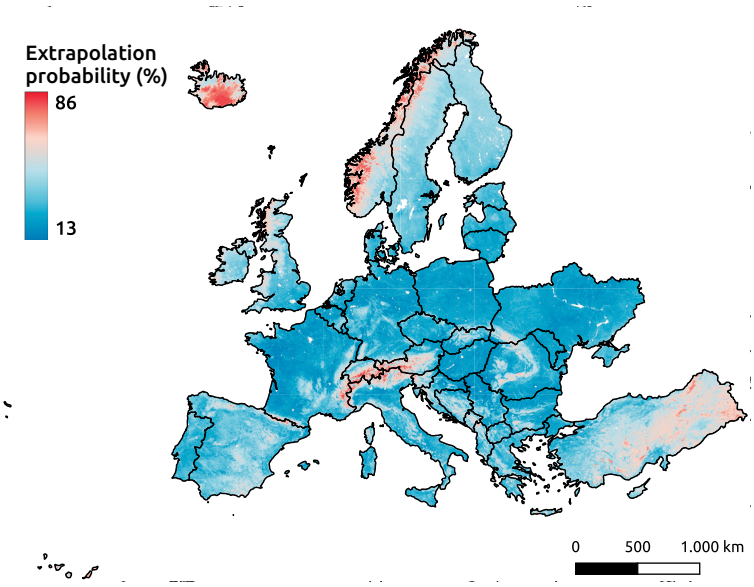


Figure 3.17: Extrapolation risk probability map for topsoil during the period 2020–2022.

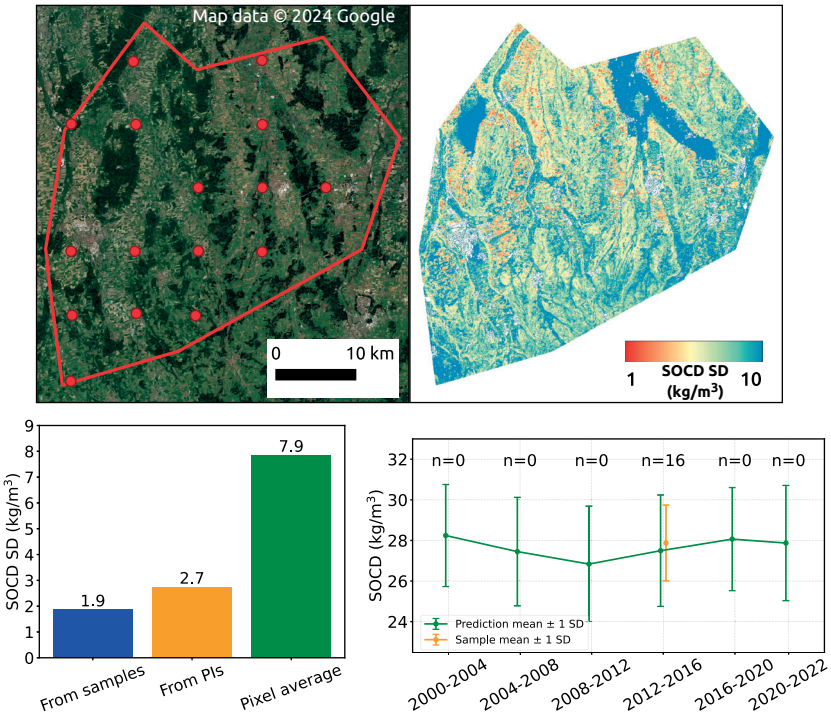


Figure 3.18: SOCD observation samples within the AOI (top left); SOCD standard deviation map of the AOI (top right); Comparison of SOCD uncertainty in AOI estimated from model-based maps, design-based samples, and the average of pixel-level uncertainty (bottom left); time series of spatial aggregates and associated uncertainty (bottom right). Google Maps (2024, CNES/Airbus, Maxar Technologies), available through https://www.google.com/maps/, last accessed: 30 August 2024.

Chapter 4

Modeling soil organic carbon changes using signal-to-noise analysis: a case study using European soil survey datasets

This chapter is based on:

X. Tian, S. de Bruin, F. Schneider, M. Herold, and K. de Beurs (2026b). "Modeling Soil Organic Carbon Changes Using Signal-To-Noise Analysis: A Case Study Using European Soil Survey Datasets". *Global Change Biology* 32.4, e70813. DOI: 10.1111/gcb.70813

Abstract

Soil Organic Carbon (SOC) is a key indicator of soil health and a crucial component of climate mitigation, making its reliable monitoring increasingly important. While Digital Soil Mapping (DSM) based on Machine Learning (ML) and Earth Observation (EO) data enables the generation of time series of spatially explicit SOC predictions, detecting temporal changes from these model predictions remains challenging due to the relatively large associated uncertainties. Although prediction uncertainties are now commonly reported, few studies have explicitly accounted for them when assessing SOC change. This study introduces a model-based signal-to-noise ratio (SNR) framework to assess the detectability of SOC change using both the *state-first* approach—modeling SOC states at each time point and then deriving change—and the *change-first* approach—modeling SOC change directly from repeated measurements. Signal-to-Noise Ratio (SNR) is defined as the ratio of predicted SOC (concentration, g/kg) change to its modeled uncertainty, enabling evaluation of change-model reliability at pixel levels. Applied to repeated SOC observations from the pan-European Land Use/Cover Area Frame Statistical Survey, this framework assesses the reliability of SOC change modeling across multiple land cover types using Random Forest and Quantile Regression Forests (QRF). At the site level, prediction accuracy was poor and SNR values were consistently low (<1), highlighting the limitations of current data for modeling SOC dynamics. An illustrative aggregation analysis showed that spatial averaging improved SNR, supporting SOC change assessments at broader scales. However, further work is needed to incorporate land use and management information and to systematically examine how different aggregation schemes affect the results in various contexts, ensuring that aggregated outcomes remain meaningful and policy-relevant. As an internal metric based solely on model predictions and their estimated uncertainty, SNR provides a practical diagnostic of change-model confidence, especially when repeated ground-truth SOC measurements are not available. We advocate for routine SNR reporting to enhance the transparency and credibility of DSM-based SOC change monitoring as EO datasets and ground-measurement surveys continue to expand in time and quality.

4.1 Introduction

Soil Organic Carbon (SOC) is widely recognized as a key indicator of soil health and a critical component of climate mitigation, owing to its role in sequestering atmospheric carbon (Don et al., 2024; Lal, 2004; Lehmann et al., 2020; Stockmann et al., 2013). Reliable information on SOC dynamics is therefore increasingly sought by scientists, policymakers, and land managers, particularly for monitoring, reporting, and verification (MRV) purposes (Lehmann et al., 2020; Smith et al., 2020).

To meet this growing need for reliable SOC information, Digital Soil Mapping (DSM) has gained momentum in recent years, particularly through approaches combining Machine Learning (ML) and satellite-based Earth Observation (EO) data (Wadoux et al., 2020b). ML captures complex, nonlinear relationships among predictors, while EO provides consistent, high-resolution, and regularly updated surface information. Together with the expansion of soil survey databases, these advances have driven a surge in spatiotemporal SOC mapping (Hengl et al., 2018; Wadoux et al., 2020b). Notable examples include the biannual Worldsoils global maps (Wesemael et al., 2024), the SoilHealthDataCube for Europe (Tian et al., 2025a), and national-scale time series for Argentina, the Netherlands, South Africa, Hungary, and others (Helfenstein et al., 2024b; Heuvelink et al., 2021; Szatmári et al., 2024a; Venter et al., 2021a). Beyond reconstructing past SOC dynamics, several studies have also extended similar approaches to project future trends (Soils Revealed, 2025; Wang et al., 2022; Yigini and Panagos, 2016; Zhang et al., 2023c). From such SOC maps, SOC change is derived by first predicting SOC levels at each time step and then calculating their difference or fitting a trend line (Chen et al., 2019; Helfenstein et al., 2024b; Li et al., 2022; Meng et al., 2024; Szatmári et al., 2024b; Yang et al., 2022), an approach here referred to as the *state-first* strategy.

In contrast, the *change-first* approach derives SOC changes directly from repeated observations and models those changes as the target variable. By focusing on change rather than absolute states, this strategy can mitigate correlated noise between time steps (De Rosa et al., 2024). It remains less common due to the scarcity of long-term, repeated SOC measurements. Practical mapping applications are still rare, with De Rosa et al. (2024)'s analysis of repeated Land Use/Cover Area Frame Statistical Survey (LUCAS, Orgiazzi et al. 2018) topsoil samples being one of the few examples.

Uncertainty in SOC prediction can arise from multiple sources, including soil sampling variability, EO data quality, and model assumptions (Even et al., 2025; Heuvelink et al., 2021; Zhang et al., 2024b). As a result, accounting for prediction uncertainty has become an integral part of DSM (Heuvelink, 2018; Heuvelink, 2014). Various methods have been developed to quantify model uncertainty, among which Quantile Regression Forests (QRF) are widely applied (Vaysse and Lagacherie, 2017; Wadoux et al., 2020b). Consequently, most recent DSM products now report uncertainty alongside predicted values, typically

as prediction intervals derived from model distributions (Heuvelink, 2014; Kasraei et al., 2021; Schmidinger and Heuvelink, 2023).

Although uncertainty is particularly relevant when expected SOC changes are small relative to prediction uncertainty (Paustian et al., 2019; Poeplau and Don, 2013), few studies have quantitatively tested whether modeled SOC changes exceed their associated uncertainty. In the *state-first* approach, uncertainty is often visualized through prediction intervals alongside SOC time series (Tian et al., 2025a), which provides useful context but remains qualitative. A rare quantitative example is Szatmári et al. (2024b), who considered uncertainty in SOC change but focused only on trend direction (positive vs. negative). For the *change-first* approach, applications are even rarer; in De Rosa et al. (2024)'s study, uncertainty was shown alongside predicted SOC change, but only at an aggregated longitudinal level.

To address this gap, we introduce a model-based signal-to-noise ratio (SNR) framework to assess whether predicted SOC changes are distinguishable from model uncertainty. Signal-to-Noise Ratio (SNR) is defined as the ratio between the magnitude of the predicted SOC change (the signal) and its modeled prediction uncertainty (the noise). A high SNR (SNR > 1) indicates that the modeled change exceeds its uncertainty, providing a straightforward internal measure of model confidence—particularly valuable when independent repeated soil samples for external validation are scarce.

We demonstrate the approach with repeated SOC measurements from the pan-European Land Use/Cover Area Frame Statistical Survey (LUCAS) topsoil (0-20 cm) dataset, which enables evaluation across diverse land cover types under both the *state-first* and *change-first* approaches. Although implemented at the pixel level, the SNR framework can also be applied to aggregated spatial units (e.g., administrative boundaries or regular grids), which is particularly relevant for MRV applications at different spatial levels. Spatial aggregation further improves SNR by dampening localized noise (Szatmári et al., 2024b; Tian et al., 2025a; Wadoux and Heuvelink, 2023), a point we demonstrate through case studies in this paper. Although SOC stock (e.g., Mg/ha) is more relevant for climate change mitigation, repeated measurements of bulk density, coarse fragments and depth to bedrock are even scarcer than those of SOC concentration (e.g., g/kg). Hence, this study focuses on SOC concentration, with "SOC" hereafter referring to concentration. While SOC concentration is used in this study due to current data limitations, the proposed SNR framework can in principle be extended to SOC stock when the necessary auxiliary data become available.

Through this SNR-based evaluation, we address the following research questions, which together aim to assess the reliability and detectability of SOC change modeling across spatial scales:

- **RQ1**: How do SOC change prediction capabilities across Europe compare to those reported for the Netherlands (Helfenstein et al., 2024b) and German croplands (Broeg et al., 2024b)?
- **RQ2**: How do the *change-first* and *state-first* strategies differ in predictive accuracy and SNR when modeling SOC change?
- **RQ3**: Which factors influence SNR in SOC change modeling, such as land cover, monitoring duration, sampling frequency and spatial scales of interest?
- **RQ4**: How does SNR relate to prediction accuracy in SOC change modeling?

4.2 Materials and Methods

4.2.1 Definitions and notation for SOC and its change

We denote the measured SOC (g/kg) at location s and time t_i as c_{s,t_i}. SOC change can be described in two main ways: (i) as the net difference between paired observations over a defined interval ($\delta_{s,t_1 \to t_2}$; g/kg) (Harbo et al., 2023; Lettens et al., 2005; Poeplau and Don, 2013), and (ii) as a temporal trend estimated through regression models that relate SOC to time. In the absence of detailed knowledge about temporal SOC dynamics, it is commonly assumed that SOC changes linearly with time, and linear models are fitted across multiple time points. The resulting slope parameters, denoted $\beta_{s,t_1 \to t_n}$ (g/kg/year) (Bellamy et al., 2005; Harbo et al., 2025), represent the average rate of SOC change over the time window $t_1 \to t_n$ for spatial unit s. All SOC change metrics are derived from repeated measurements at the same location (c_{s,t_i}).

For simplicity, spatial and temporal indices (e.g., s, t) are omitted throughout the manuscript unless explicitly required. Accordingly, SOC measurements and their derived metrics are denoted as c, δ, and β, representing the observed SOC concentration, net change, and temporal trend, respectively. Their corresponding model-predicted values are denoted as $\hat{c}$, $\hat{\delta}$, and $\hat{\beta}$.

4.2.2 Data preparation

In this study, we primarily used soil data from the LUCAS soil database, supplemented with three national soil datasets. Two of these are from Spain: Parcelas COS (top 30 cm depth) and Parcelas INES (top 10 cm) (Serrano et al., 2022). The third dataset, BZE-LW, is from Germany and represents the core dataset of the first German agricultural soil inventory (Poeplau et al., 2020b), containing measurements across multiple soil depths. As the only database providing repeated topsoil samples (0–20 cm) with pan-European coverage, LUCAS served as the reference dataset and foundation of this study. The analysis focused on topsoil samples within the 0–20 cm depth range. For sites with measurements at multiple

depths, SOC values were interpolated to a standardized depth of 10 cm—representing the midpoint of the target layer. To preserve the shape of data and avoid overshooting, we used the Piecewise Cubic Hermite Interpolating Polynomial method (PCHIP) (Fritsch and Butland, 1984), which achieved a high coefficient of determination (0.8) in reconstructing SOC values at missing depths in an internal accuracy assessment (Supplementary Notebook 02e). The final harmonized dataset includes 87,684 c observations, from all over Europe (Fig. 4.1-A), covering year 2002 to 2019 (Fig. 4.1-C). Summary statistics are provided in Table 4.1.

Table 4.1: Summary statistics of point SOC (g/kg) per dataset.

Source	Year	N sites	N samples	mean	median	std
LUCAS	2009, 2012, 2015, 2018	27737	62437	45.98	20.70	82.03
Parcelas INES	2002-2018	21502	21502	36.63	23.53	38.67
BZE-LW	2011-2018	2953	2953	30.07	18.37	49.22
Parcelas COS	2019-2020	789	789	40.78	27.17	34.64

Among the LUCAS observations, a subset includes repeated measurements taken at the same locations over three sampling rounds: 2009/2012, 2015, and 2018, enabling the analysis of SOC dynamics over time. To ensure data consistency and quality, a filtering procedure was applied to keep only consistent SOC time series. The following criteria were used: (1) each time series must contain at least three observations, representing the longest possible duration; (2) the maximum spatial distance between repeated measurements must not exceed 30 m, a practical threshold aligning with the resolution of our previous SOC maps (Tian et al., 2025a), under the simplifying assumption that points within 30 m fall within the same map pixel; and (3) the standard deviation of the carbon-to-nitrogen (C/N) ratio across the time series must be less than 3.8, based on the observed bimodal distribution of C/N differences, where 3.8 marks a local minimum between two peaks (see Supplementary Notebook 02b for details, Ziche et al. 2022). This filtering yielded 10,204 SOC time series from the LUCAS dataset, covering most of Europe (Fig. 4.1-B), with observations primarily from the years 2009, 2015, and 2018 (Fig. 4.1-C). A small subset of time series includes measurements from 2009, 2012, and 2018, though this group is not clearly visible due to its limited size.

The change (δ) was calculated for all valid pairs within each time series, including step-to-step (e.g., $t_1 \rightarrow t_2$, $t_2 \rightarrow t_3$) and longer-interval (e.g., $t_1 \rightarrow t_3$) differences. This resulted in a total of 30,612 δ values across the dataset. We estimated the rate of change at each location using the Theil–Sen estimator—a non-parametric, robust linear regression method well-suited for small sample sizes and resistant to outliers (Sen, 1968; Theil, 1950). Slopes were derived from the full SOC time series at each site, yielding 10,204 β values. This was implemented using `TheilSenRegressor` from `sklearn.linear_model`.

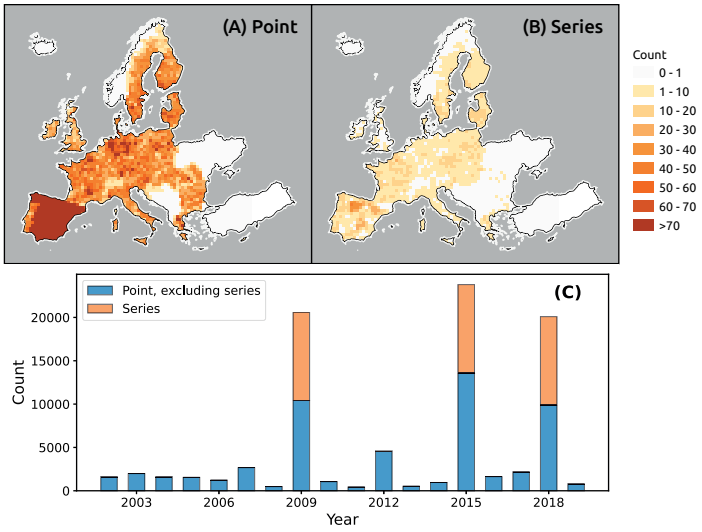


Figure 4.1: Spatial density of all point measurements (A) and filtered time-series observations (B) aggregated to 50 km grid cells across Europe, together with their temporal distribution across survey years (C).

An independent test set of 1,800 SOC time series (about 20% of the dataset, corresponding to 5,400 point observations and 5,400 time-step pairs) was randomly selected from the filtered LUCAS dataset. This size was chosen to ensure a sufficiently large and diverse test set for evaluation, while leaving enough data for training and exploration. This test set was used for evaluating prediction accuracy, assessing uncertainty estimates, and conducting SNR analysis.

4.2.3 Covariates

A wide range of covariates (also referred to as *predictors* or *features*) was included to represent environmental factors influencing soil formation, as described by the SCORPAN model. Specifically, SCORPAN refers to s: soil properties; c: climate; o: organisms (vegetation); r: topography; p: parent material (lithology); a: age; and n: spatial position (McBratney et al., 2003). Topographic variables (Ho et al., 2025) and parent lithology (Isik et al., 2024) were included to represent terrain-driven controls on hydrology, erosion, and soil formation processes. Plant functional type (Harper et al., 2023) and vegetation cover fraction (Sun et al., 2023) maps were used to characterize vegetation composition and cover, which are directly related to carbon inputs to the soil (Smith, 2008). Climate variables include precipitation (Karger et al., 2021), temperature (Wan, 2006), and water vapor (Lyapustin and Wang, 2018), as well as bioclimatic variables (Karger et al., 2017) that represent biologically meaningful climate constraints relevant to vegetation growth and

soil carbon processes. Thematic peatland (Widyastuti et al., 2024) and cropland (Potapov et al., 2022) datasets were included to represent persistently waterlogged peat soils with high SOC contents and intensively managed croplands with direct human intervention. Soil moisture (Bauer-Marschallinger et al., 2018) was included to indicate water availability and aeration conditions that regulate organic matter decomposition and SOC stabilization. In addition, less-processed remote sensing signals were incorporated, including Landsat reflectance and derived spectral indices (Tian et al., 2025b) and bare-soil composites from Sentinel imagery (Rogge et al., 2018). Two Synthetic Aperture Radar (SAR) backscatter signals at different wavelengths (Shimada and Ohtaki, 2010; Wagner et al., 2021) were also included. These signals are sensitive to surface properties and vegetation structure and have been shown to be informative for modelling SOC in superficial soil layers (Baghdadi et al., 2008).

Some covariates were available as time series, providing multiple observations for the same location over time, including vegetation-related variables, climate variables, cropland information, and soil moisture. Other covariates provided only a single value per location, including static variables or long-term summaries. Static layers include slowly varying properties such as topography and lithology, as well as variables treated as static as a pragmatic compromise given data availability and consistency at the continental scale, including Synthetic Aperture Radar (SAR) backscatter and peatland cover. Long-term summaries were derived from multi-year time series for selected covariates and include statistical descriptors such as the mean, percentiles, and standard deviation, resulting in a single value per location. In this study, "long-term" refers to the period from 2000 to 2022, corresponding to the time span over which most dynamic covariates are consistently available.

The resulting initial feature pool consisted of 582 individual covariates, organized into 15 covariate groups, which in turn represent seven broader SCORPAN classes (Table 4.2). The full list of covariates is not included in the main text but is available in the supplementary materials at our Supplemetary Repository. Various preprocessing steps were applied to the original datasets, including temporal aggregation, gap-filling of missing values, and reprojection to ensure consistent coordinate reference systems, spatial coverage, and temporal alignment across all layers.

These covariate layers were overlaid with c observations based on spatial coordinates (latitude and longitude) and temporal information (year) to construct the training matrix for the RF_c. To reduce dimensionality and improve interpretability, we applied the Repeated Subsampling-Based Cumulative Feature Importance (RSCFI) method (Tian et al., 2025a). This approach, conceptually similar to recursive feature elimination with cross-validation, iteratively removes features with low cumulative importance scores, thereby retaining only the most relevant predictors while balancing model performance

Table 4.2: Covariate groups used in this study. Note that each entry represents a group of covariate layers, which may include multiple datasets with different temporal and spatial resolutions.

Class	Covariate group	Source	Temporal resolution	Spatial resolution
Topography	Digital terrain model and derived land surface parameters	Ho et al., 2025	Static	30, 60, 120, 240, 480, and 960 m
Lithology	Lithology type probability	Isik et al., 2024	Static	250 m
Vegetation	Plant functional type	Harper et al., 2023	Annual	300 m
Vegetation	Vegetation cover fraction	Sun et al., 2023	Annual	500 m
Climate	Precipitation	Karger et al., 2021	Annual	1 km
Climate	Land surface temperature	Wan, 2006	Annual	1 km
Climate	Water vapor	Lyapustin and Wang, 2018	Annual	1 km
Climate	Bioclimate	Karger et al., 2017	Long-term	1 km
Imagery	Landsat spectral index	Tian et al., 2025b	Annual, long-term	30 m
Imagery	Bare soil composite	Rogge et al., 2018	Long-term	30 m
Imagery	Sentinel-1 SAR backscatter	Wagner et al., 2021	Static	25 m
Imagery	PALSAR backscatter	Shimada and Ohtaki, 2010	Static	25 m
Land cover	Peatland	Widyastuti et al., 2024	Static	1 km
Land cover	Cropland	Potapov et al., 2022	Annual	30 m
Soil	Soil moisture	Bauer-Marschallinger et al., 2018	Annual	1 km

and computational efficiency. Details of the RSCFI implementation are provided in Supplermentary Notebook 07a.

The covariates identified as important by RF_c were then organized to reflect environmental changes over time, supporting the construction of training matrix for RF_δ and RF_β. Following Yang et al. (2022), we applied time-series feature engineering to convert state-based covariates into change-based representations. Each covariate time series was partitioned into three periods: (i) the change interval (between start and end years), (ii) the years preceding the change (from 2000—the earliest common year—up to the start year), and (iii) the full period from 2000 to the end year. From each period, we derived summary statistics of feature values—mean, standard deviation, and linear slope (as a proxy for trend)—yielding a structured set of dynamic predictors aligned with the temporal spans of δ and β. These were then used to train RF_δ and RF_β.

4.2.4 SOC model approaches

All models are implemented using Random Forest (RF) and QRF, two widely used methods in DSM for SOC prediction and uncertainty quantification (Hengl et al., 2018; Wadoux et al., 2020b). We used a modified version of the RF algorithm to capture full prediction distributions by storing values from all valid nodes across trees. This mimics the behavior of QRF, enabling simultaneous estimation of the target variable (via the mean) and associated uncertainty (via the standard deviation). This was implemented by modifying the `RandomForestRegressor` in `scikit-learn` to keep node-level outputs. Three separate RF models were constructed using the same workflow, including feature selection, hyperparameter tuning, and model fitting, to predict c, δ and β, respectively.

- *state-first*: This approach relies on a single model, RF_c, to generate spatial-temporal $\hat{c}$, from which $\hat{\delta}$ and $\hat{\beta}$ are subsequently derived.
  - **Model of spatial-temporal point SOC values (RF_c)**: The training matrix was constructed by aligning SOC observations with the corresponding environmental features, drawn from 15 covariate groups (Table 4.2), based on spatial coordinates (latitude and longitude) and temporal information (year).
- *change-first*: In this approach, the target variable directly corresponds to the modeled output. Temporal dynamics are incorporated through feature engineering of covariates, aiming to reflect underlying processes rather than static states (see next section for details).
  - **Model of paired SOC change (RF_δ)**: The training matrix was constructed by matching δ values with features that reflect environmental changes over the corresponding time period at the same location.

– **Model of SOC trend slope (RF_β)**: As with the RF_δ, covariates were engineered to reflect environmental processes across the time span of each series.

Performance metrics—including mean absolute error (MAE), bias, and concordance correlation coefficient (CCC)—were calculated on an independent test set to evaluate the prediction accuracy of the three models: RF_c, RF_δ, and RF_β, used under both the *state-first* and *change-first* approaches. This resulted in five prediction scenarios (Table 4.3).

Table 4.3: Overview of RF models used in the *state-first* and *change-first* approaches and their corresponding target variables.

Approach	Model	Target	Predicted
State-first	RF_c	c	$\hat{c}$
		δ	$\hat{\delta}$ derived as temporal differences of $\hat{c}$
		β	$\hat{\beta}$ derived as temporal slopes fitted to $\hat{c}$
Change-first	RF_δ	δ	$\hat{\delta}$
	RF_β	β	$\hat{\beta}$

We used the standard deviation of RF node outputs as a proxy for prediction noise in the SNR calculation—an approach valid only if this spread reliably reflects uncertainty. Therefore, the models were also evaluated for the validity of their uncertainty estimates using two metrics: the Prediction Interval Coverage Probability (PICP), which measures the proportion of observed values falling within the predicted intervals; and the Quantile Coverage Probability (QCP), which assesses the symmetry of the uncertainty distribution by evaluating single-quantile predictions. Accuracy plots of PICP and QCP—scatter plots comparing observed and expected coverage rates across a range of expected prediction intervals—were used to visualize and evaluate local uncertainty estimation performance, as recommended by Schmidinger and Heuvelink (2023).

4.2.5 SNR Analysis

The SNR was defined as the ratio between the model's predicted value and its associated prediction uncertainty. SNR thus quantifies the extent to which a modeled change exceeds its own uncertainty, providing an internal diagnostic of model confidence. Its specific computation varies by model, target variable, and approach, as shown in Table 4.4, where the numerator represents the *signal*—the absolute magnitude of the predicted target ($|\hat{c}|$, $|\hat{\delta}|$, or $|\hat{\beta}|$). Although the calculation of SNR varies across models, all rely on the prediction distribution—that is, the set of terminal node values from the QRF. This conditional distribution is assumed to approximate model uncertainty within the local feature space.

In the *change-first* approach, noise is quantified as the standard deviation of this prediction distribution. While this is similar to QRF, which use the same prediction set to estimate conditional quantiles (Meinshausen, 2006), our method focuses on capturing uncertainty level via spread. This procedure is applied consistently across the three direct models (RF_c, RF_δ, and RF_β), corresponding to the diagonal entries in Table 4.4.

For the *state-first* approach, the RF_c first predicts $\hat{c}$ values, from which both $\hat{\delta}$ and $\hat{\beta}$ are derived (Table 4.3). To estimate noise in $\hat{\delta}$, we assumed the two $\hat{c}$ estimates to be independent, as RF models treat samples independently. Under this assumption, the noise for $\hat{\delta}$ is computed as the square root of the sum of variances from both time steps. To estimate noise for $\hat{\beta}$, we implemented a Monte Carlo simulation. For each iteration, $\hat{c}$ values were randomly sampled from the full prediction distribution (i.e., all valid leaf node outputs) at each time step, and a linear regression was fitted to estimate the corresponding slope. This process was repeated 200 times, generating a distribution of $\hat{\beta}$ values. The standard deviation of this distribution was used as the noise estimate for $\hat{\beta}$.

Table 4.4: SNR calculation formulas for each model and target variable.

Target \ Model	RF_c $R = \{\hat{c}^{(1)}, \hat{c}^{(2)}, \ldots, \hat{c}^{(n)}\}$	RF_δ $R = \{\hat{\delta}^{(1)}, \hat{\delta}^{(2)}, \ldots, \hat{\delta}^{(n)}\}$	RF_β $R = \{\hat{\beta}^{(1)}, \hat{\beta}^{(2)}, \ldots, \hat{\beta}^{(n)}\}$
c	$\frac{\hat{c}}{std(R)}$	-	-
δ	$\frac{\lvert\delta(\hat{c}_{t_0},\hat{c}_{t_1})\rvert}{\sqrt{std(R_{t_0})^2+std(R_{t_1})^2}}$ *	$\frac{\lvert\hat{\delta}\rvert}{std(R)}$ #	-
β	$\frac{\lvert\beta(\hat{c}_{t_0},\hat{c}_{t_1},\hat{c}_{t_2})\rvert}{std(\beta(R_{t_0},R_{t_1},R_{t_2}))}$ *	-	$\frac{\lvert\hat{\beta}\rvert}{std(R)}$ #

Notes:
* indicates the *state-first* approach; # indicates the *change-first* approach.
R denotes the prediction distribution, i.e., all prediction instances for a target variable obtained from the full valid terminal node values of the RF.

To explore the relationship between internal model confidence (SNR) and actual predictive accuracy, we compared SNR with relative absolute error (RAE) for each $\hat{\delta}$ and $\hat{\beta}$ in the validation dataset. RAE was used to evaluate prediction accuracy across all target variables (c, δ, and β). It is defined as the absolute difference between the predicted and observed values, normalized by the observed value.

4.2.6 Spatial aggregation

While spatial averages can be calculated by averaging predictions over the area of interest (AOI), quantifying the corresponding uncertainty is less straightforward—particularly

owing to the spatial dependence of prediction errors. In this study, we adopted an approach similar to that used by Araza et al. (2022) and Wadoux and Heuvelink (2023), which estimates the uncertainty of spatial aggregates using model-based point-support uncertainty-prediction standard deviation in our case-derived from discretization points within the AOI. This method avoids the numerical complexity of traditional statistical techniques, making it especially suitable for large-scale applications. Specifically, it is implemented as follows:

$$\mathrm{sd}_{AOI} = \sqrt{\frac{1}{|B|^2} \int_{s \in B} \int_{u \in B} \mathrm{sd}(s) \cdot \mathrm{sd}(u) \cdot \rho(|s-u|)\, ds\, du} \tag{4.1}$$

The B in Eq.(4.1) represents the number of discretization points, generated from the AOI; $\mathrm{sd}(s)$ and $\mathrm{sd}(u)$ denote the standard deviations of the point support prediction residuals at locations s and u, assuming the prediction errors are proportional to the prediction uncertainty (standard deviation of prediction distribution from QRF) at point support. The $\rho(|s-u|)$ represents the correlation function of the standardized prediction error at the separation distance between s and u. Standardized prediction errors and corresponding uncertainty were derived through cross-validation, where predictions were made on samples excluded from model training. Errors were calculated as the difference between observed and predicted values, then standardized by dividing by the prediction uncertainty. We assumed second-order stationarity of the standardized errors—that is, their spatial correlation depends only on the Euclidean distance between locations. The correlation structure was estimated by fitting a variogram $\gamma(h)$ using:

$$\rho(h) = \frac{\mathrm{sill} - \gamma(h)}{\mathrm{sill}}, \tag{4.2}$$

where sill denotes the fitted sill of the variogram.

To assess the effect of spatial aggregation on SNR, SOC prediction maps at 1 km^2 resolution were first generated, which served as the basis for subsequent aggregation. We considered four levels of spatial support, a concept from geostatistics describing the area over which a variable is aggregated: 2.5, 10, 40, and 200 km^2. These scales, spanning fine to coarse resolutions, were chosen to demonstrate the effect of spatial aggregation on SNR and are not intended as prescriptive values. The analysis focused on Germany and Spain, where additional data is available. Each country was divided into square grid units corresponding to the four support sizes. Within each grid unit, map pixels were used as discretization points for aggregation, and B denotes the number of pixels contained in each grid unit at the respective spatial support.

Spatial aggregation was applied only to the RF_c model under the *state-first* approach. This restriction was mainly due to the limited density of available SOC time series measurements, which must be sufficient to fit standardized error variograms for aggregated uncertainty estimation. Moreover, as *state-first* approach remains the standard in SOC change modeling, results are broadly applicable to current SOC time series products.

4.3 Results

4.3.1 Prediction accuracy and uncertainty estimation evaluation

The prediction accuracy of $\hat{c}$ was first evaluated to justify the use of the *state-first* approach. RF_c achieves a CCC of 0.344 in the original scale (Fig. 4.2-A) and 0.705 in the log-transformed scale (Fig. 4.2-B). These values are comparable to those reported in previously published continental-scale DSM models (Tian et al., 2025a; Wesemael et al., 2024). While the model shows a slight positive bias—tending to overestimate SOC—it consistently underestimates the highest observed SOC values.

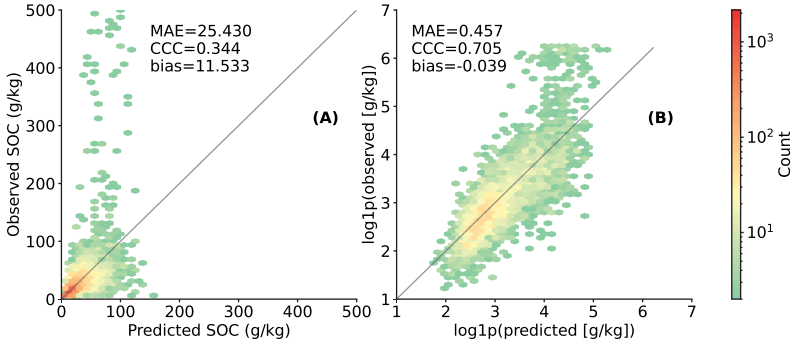


Figure 4.2: Prediction accuracy of c using RF_c. Panel (A) shows the density plot of observed versus predicted values on the original scale (g/kg), and panel (B) shows the same results with log-transformed values (log1p transformed g/kg) for greater detail at lower concentrations. Both panels are based on the same validation dataset. Performance metrics are calculated and presented in each panel.

Figure 4.3 shows the accuracy of SOC change predictions—$\hat{\delta}$ (top) and $\hat{\beta}$ (bottom)—using both *state-first* (left) and *change-first* (right) approaches. Unlike the good performance for $\hat{c}$, $\hat{\delta}$ and $\hat{\beta}$ show poor accuracy across all models and approaches, with all models yielding near-zero CCC and high MAE relative to the scale of the target variables. For *state-first* (Fig.4.3-A and C), predictions are weakly correlated with observations. Accuracy does not improve under *change-first* (Fig.4.3-B and D). Both approaches struggle to reliably capture the correct direction of change. The *state-first* and *change-first* approaches yield conflicting biases for both $\hat{\delta}$ and $\hat{\beta}$.

Despite overall poor performance, *state-first* approach (Fig.4.3-A) shows a wider spread of $\hat{\delta}$ predictions than *change-first* (Fig.4.3-B), where predictions cluster around zero. This contrast is less pronounced for $\hat{\beta}$ (Fig. 4.3-C and D).

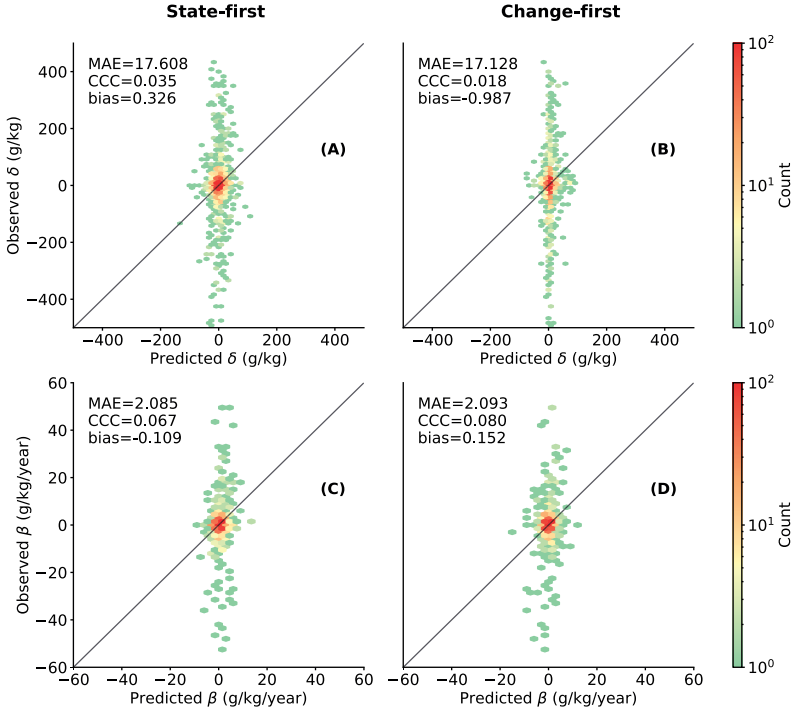


Figure 4.3: Prediction accuracy of δ and β using two approaches: *state-first* based on predictions from RF_c (A and C), and *change-first* using RF_δ (B) and RF_β (D). Performance metrics are shown in the upper-left corner of each subplot.

All three models demonstrated reasonable performance in estimating uncertainty when applied to their direct target under the *change-first* approach (Fig. 4.4). A slight underestimation—indicative of over-optimism—was observed in the PICP of the RF_δ and RF_β (Fig. 4.4-D and F) when the expected prediction interval was around 50%. When assessing the sharpness of prediction intervals, the results (see Supplementary Notebook 15) show that the uncertainties estimated by RF_δ and RF_β under the *change-first* approach are generally narrower but exhibit less variation.

4.3.2 SNR analysis of SOC and its change modelling

Model-based SNR provides an internal diagnostic of SOC change detectability, especially when repeated ground-truth data are unavailable. The SNR analysis for $\hat{c}$ is presented in Fig. 4.5, stratified by land cover type. All Level 1 land cover classes from the LUCAS

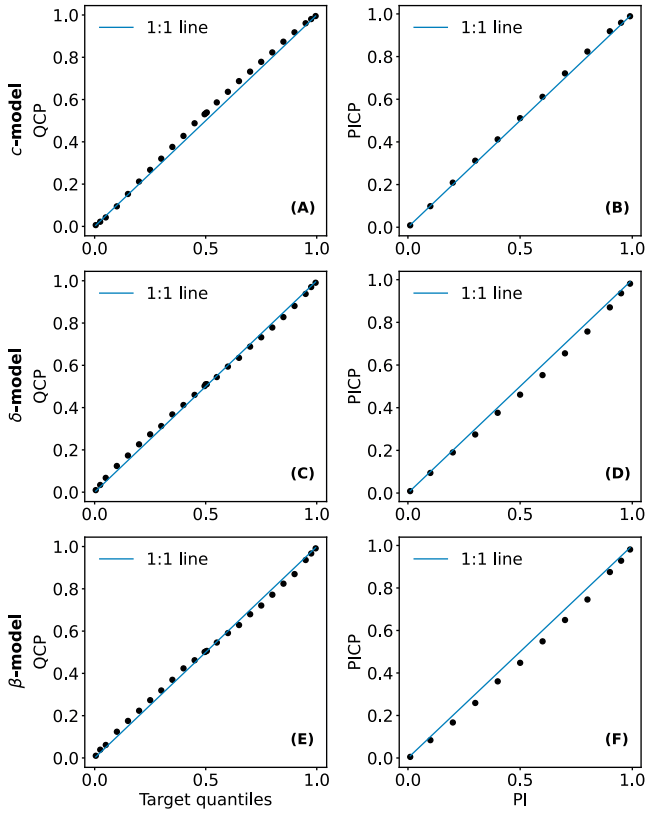


Figure 4.4: Accuracy plots of QCP (left column) and PICP (right column) used to evaluate the RF_c (top), RF_δ (middle), and RF_β (bottom)—in estimating uncertainty for their respective target variables. All plots present scatter points comparing observed coverage rates on the y-axis with expected coverage rates on the x-axis across a range of prediction intervals.

survey are represented in the study, except for *Wetland and Water Areas* and *Artificial Land*, which were excluded due to insufficient sample sizes (fewer than 10 observations in the test set). As shown in Fig. 4.5, the histograms of observed and predicted SOC generally align for *Bareland*, *Cropland*, *Grassland*, and *Shrubland*, while clear mismatch is observed for *Woodland*. In *Woodland*, the predicted SOC values are shifted toward higher concentrations and appear compressed, with less extreme values at both ends.

As the independent test set is a random subset of the full dataset, the sample counts reflect the original data distribution, with the majority of observations coming from *Cropland*, followed by *Woodland* and *Grassland*. *Bareland* and *Shrubland* are minimally represented. The signal component increases from *Bareland* to *Woodland*, consistent with increasing vegetation content. Both prediction errors and noise estimates also increase along this

gradient. In terms of SNR, *Cropland* shows the highest values, while *Woodland* has the lowest, despite having the second-highest number of available samples. SNR values are mostly above one for *Bareland* and *Cropland*, close to one for *Grassland* and *Shrubland*, and predominantly below one for *Woodland*.

Fig. 4.6 presents the SNR of $\hat{\delta}$ (bottom panel) for both the *state-first* and *change-first* approaches, along with the corresponding signal, noise, and the distribution of observed δ for reference. Observed δ values tend to be larger and more variable in land-cover types with higher vegetation. For example, δ is larger when *Woodland* is involved—either when the land cover remained as *Woodland* across both sampling years or transitioned to or from it. In contrast, transitions involving *Cropland* show lower magnitudes and narrower spreads. Although the exact values of δ and $\hat{\delta}$ do not match at the individual prediction level (Fig. 4.3), both the *state-first* and *change-first* approaches reproduce the overall trend of δ in $\hat{\delta}$ across land-cover types, with higher predicted signal magnitudes for transitions involving vegetation-rich classes.

Overall, the *change-first* approach yields lower SNR values than the *state-first* approach, except for transitions involving *Woodland* (e.g., WW, SW, GW). The variation in SNR across land-cover types is more pronounced for the *change-first* approach. This approach reproduces the signal pattern across land-cover classes—where vegetation-rich covers correspond to larger signals—while maintaining relatively stable noise estimates. Consequently, transitions involving high-vegetation land covers show higher SNR values and greater variation across classes. In contrast, under the *state-first* approach, although the strongest and most variable signals also occur in high-vegetation classes such as *Woodland*, the noise magnitudes follow similar land-cover patterns as the signal and δ. As a result, SNR values remain relatively consistent across transitions and generally below 0.4.

Compared to $\hat{\delta}$, the SNR estimates for $\hat{\beta}$ from both approaches show stronger agreement, with more similar magnitudes and variation across land-cover types, although the overall SNR patterns for β remain similar to those for δ. Observed β values are higher and more variable for transition types involving vegetation-rich land covers, a pattern mirrored in the predicted signals from both the *state-first* and *change-first* approaches. For the *change-first* approach, noise estimates remain relatively stable, resulting in larger SNR variations across land-cover transitions compared with the *state-first* approach.

For both $\hat{\delta}$ and $\hat{\beta}$, only a weak positive correlation was observed between SNR and RAE (Spearman's rank correlation coefficient was 0.17 in both cases for the *state-first* approach and close to zero for the *change-first* approach.; see Supplementary Notebook 12). The majority of predictions fall into the "high RAE and low SNR" quadrant, indicating that large prediction errors are generally associated with low detectability of modeled SOC changes.

4.3.3 The influence of time span and sampling times

Table 4.5 shows no substantial differences in the mean values of signal, noise, or SNR for $\hat{\delta}$ across different time intervals, regardless of the approach used. Notably, SNR values based on nine-year SOC observation pairs are not higher than those derived from three-year intervals. This is also observed in the δ reported by LUCAS (Supplementary Notebook 04b).

Table 4.5: Summary statistics of signal, noise, and SNR for δ across different time intervals, based on predictions from *state-first* and *change-first* approach.

Category	$\Delta T = 3$		$\Delta T = 6$		$\Delta T = 9$	
	state-first	*change-first*	*state-first*	*change-first*	*state-first*	*change-first*
Count	1811	1811	1800	1800	1789	1789
Signal	5.12	2.47	4.97	3.34	5.56	3.23
Noise	63.09	31.74	64.35	32.21	64.30	32.17
SNR	0.09	0.06	0.09	0.08	0.10	0.08

Compared to δ, which is derived from paired SOC observations between 2009 and 2018, the estimation of β over the same period incorporates an additional intermediate SOC observation. When comparing the SNR of $\hat{\delta}$ and $\hat{\beta}$ across this nine-year span, $\hat{\beta}$ (0.11 for the *state-first* and 0.16 for the *change-first* approach) consistently exhibits slightly higher SNR values than $\hat{\delta}$ (0.10 for the *state-first* and 0.08 for the *change-first* approach), regardless of the approach used.

4.3.4 SNR at larger spatial support

The correlation function of the standardized errors from $\hat{c}$ was estimated using variograms to support the calculation of uncertainty in spatial aggregates. Separate variograms were derived for Germany and Spain (Fig. 4.8). While both exhibit similar nugget values (variance at zero distance; often attributed to measurement error), the variogram for Germany has a noticeably larger spatial range parameter (about 10.60 km versus 5.99 km for Spain), indicating that prediction errors in Germany remain spatially correlated over longer distances.

Fig 4.9-left panel shows that for $\hat{\delta}$, both the signal and the noise generally decline as the spatial aggregation support increases, except for a slight signal increase in Spain between 40 km and 200 km. Overall, the reduction in noise is greater than the reduction in signal, resulting in a steady increase in SNR in both countries. The spread of both signal and noise also narrows with increasing aggregation size, whereas the spread in SNR widens. The SNR increase with spatial aggregation is less pronounced in Germany than in Spain.

For $\hat{\beta}$, both the signal and noise also decrease as the aggregation size increases. However, the rates of decrease are similar, so the SNR improves only marginally. In Germany, the SNR remains relatively constant across all aggregation levels, while in Spain it increases only slightly.

As the aggregation unit size increases, $\hat{\delta}$ values tend to converge toward zero, yet many grid cells in both countries exhibit increasing SNR values (Fig. 4.10). The spatial distribution of $\hat{\delta}$ and its SNR from 2009 to 2018 indicates that the SNR increase with larger spatial supports is more pronounced in Spain, where several grid cells exceed an SNR of one. In contrast, although SNR also increases in Germany, most grid cells remain below one.

4.4 Discussion

4.4.1 Poor accuracy in predicting SOC change

We evaluated two approaches to modeling SOC change—*state-first* and *change-first*—with δ and β as target variables. Across all four scenarios, prediction accuracy for SOC change remained low (Fig. 4.3).

Most existing DSM SOC time series adopt the *state-first* approach (Gasch et al., 2015; Helfenstein et al., 2024b; Hengl et al., 2017; Heuvelink et al., 2021; Li et al., 2022; Poggio et al., 2021; Szatmári et al., 2024b; Tian et al., 2025a; Wesemael et al., 2024), which are often used for change detection and driver analysis. However, consistent with findings from the Netherlands (Helfenstein et al., 2024b), the accuracy of SOC change modelling remains poor despite high accuracy in $\hat{c}$.

A key limitation of the *state-first* approach is that it treats predictions for different time instances as independent variables, ignoring temporal autocorrelation in both SOC and environmental covariates (Poggio et al., 2021; Szatmári et al., 2024b; Tian et al., 2025a). As a critical component of SCORPAN (McBratney et al., 2003), the time factor is therefore not well integrated into current *state-first* space–time DSM frameworks. Efforts to address this gap include incorporating the age of observations as a covariate (Batjes et al., 2020), using lagged vegetation indices that weight past conditions to reflect delayed SOC responses (Heuvelink et al., 2021), or explicitly incorporating past land use changes in the input covariates (Helfenstein et al., 2024b). However, constructing such covariates is often impractical at large scales. Moreover, spatial variability in SOC typically outweighs subtle temporal variation, limiting the benefits of such preparation (Poggio et al., 2021; Tian et al., 2025a). This pattern is also reflected in the exploratory feature contribution analysis (see Supplementary Notebook 14), where static features exert a stronger overall influence on model predictions than temporal variables. As a result, models tend to overfit spatial patterns while underrepresenting temporal dynamics, projecting spatial

relationships—such as higher SOC under denser vegetation—onto the temporal dimension, even when actual SOC changes occur more slowly than vegetation.

The *change-first* approach takes a different strategy by modeling SOC change directly. In theory, this offers two advantages: (1) correlated noise in SOC values may cancel out when computing change, and (2) delayed environmental effects may be better captured by directly linking environmental dynamics to SOC change. However, as shown in Fig. 4.3, SOC change prediction accuracy did not improve as expected. Although the predictive accuracy of $\hat{\beta}$ was slightly higher than that of $\hat{\delta}$, both appeared largely random when validated against repeated measurements, with predictions clustering around zero. This reflects not only the model's limited ability to resolve change, but also the inherent nature of SOC as a slow-changing property, for which true changes occur over long time scales and are typically small over short intervals (Gubler et al., 2019; Poeplau and Don, 2013; Smith, 2004). This finding aligns with Broeg et al. (2024b), who attempted to predict β using a much longer (35-year) cropland SOC time series in Bavaria. While their model struggled to accurately estimate β, it succeeded in classifying SOC change direction (increase, decrease, stable). In contrast, with only the nine-year time span of LUCAS observations, our study shows that $\hat{\beta}$ predictions could not even resolve the direction of change.

4.4.2 SNR as a predicted SOC change assessment metric

Several studies have applied the concept of SNR to describe challenges in detecting SOC change. For instance, Croft et al. (2012) and Stevens et al. (2008) reported low SNR in EO reflectance data caused by illumination, terrain, and atmospheric effects, while Paustian et al. (2019) highlighted that SOC changes are often masked by natural spatial variability. Broeg et al. (2024b) quantified SNR by defining the signal as temporal variation in measured SOC over a defined time interval and the noise as the corresponding spatial prediction error. Although the specific definitions and calculations of SNR differ across studies, our findings are consistent with those of Broeg et al. (2024b). In both cases, SNR values rarely exceed one over short time spans, surpassing this threshold only after approximately 25 years of observation, and SNR is higher for measurements with higher SOC contents (>15 g/kg). This convergence underscores the robustness of SNR as a diagnostic metric and the importance of long-term monitoring for capturing SOC change.

Large-scale SOC change assessments are often based on prediction time series using the *state-first* approach, as repeated measurements are typically unavailable. In such contexts, SNR provides a practical and interpretable means of quantifying the detectability of predicted changes when direct validation is infeasible. It can therefore serve as a readily available indicator of the reliability of SOC change estimates in widely used SOC mapping products. When repeated measurements are available, *change-first* approaches become feasible, allowing overall error metrics to be computed by directly comparing predicted

and observed changes. However, these metrics remain global summaries and are limited to locations where measurements exist. By contrast, because SNR can be computed for every individual prediction, it can function as a spatially explicit quality indicator across the full prediction domain.

In this sense, SNR complements traditional accuracy metrics by providing additional insight into the reliability and interpretability of modeled SOC change, but it should not be viewed as a replacement for predictive accuracy. Accuracy metrics evaluate model performance against observations, whereas SNR serves as an internal diagnostic that indicates whether predicted changes exceed the model's own uncertainty. A key prerequisite for meaningful SNR assessment is valid uncertainty estimation. In this study, all three models—RF_c, RF_δ, and RF_β—produced reasonably calibrated uncertainty intervals, as demonstrated by both coverage (Fig. 4.4) and sharpness (Supplementary Notebook 15). Under this condition, SNR provides information on whether the model itself considers a predicted change to be distinguishable from its associated uncertainty, independent of whether the prediction is ultimately accurate.

We further compared SNR with predictive accuracy at the point level (Supplementary Notebook 12). Only weak correlations were observed between SNR and relative absolute error (Spearman's $\rho = 0.17$) for both $\hat{\delta}$ and $\hat{\beta}$. At first glance, this weak relationship, together with the generally poor predictive accuracy, may raise questions about the validity of the proposed SNR framework. However, across all modelling strategies, we consistently observe both low predictive accuracy and low SNR. These outcomes are mutually consistent, indicating that SOC change signals are dominated by noise and therefore offer limited interpretive value. As monitoring networks such as LUCAS expand, national soil surveys continue, and EO and ancillary datasets improve, longer and denser time series are expected to support more robust models and higher signal-to-noise ratios. Under such conditions, SNR may align more closely with external accuracy measures and provide a richer basis for investigating where, why, and under which conditions SOC change becomes more detectable.

4.4.3 Factors influencing SNR

When comparing SNR between $\hat{\delta}$ and $\hat{\beta}$ for the same time span, we observed higher SNR values for $\hat{\beta}$, which incorporates a third SOC observation. In contrast, signal, noise, and SNR of δ remained largely stable despite increasing intervals between paired two observations (Table 4.5). This highlights the importance of not only long-term monitoring, but also adequate temporal sampling density. Using data from Denmark's long-term Soil Monitoring Network, Harbo et al. (2023) showed that short-term fluctuations can mask long-term SOC trends, making it risky to estimate δ from only two observations. Similar short-term fluctuations are also evident in the LUCAS repeated measurements Supplementary Notebook 11b, which can be partially smoothed during trend fitting when

multiple time steps are available. Consistently, Broeg et al. (2024b) showed that SNR increases with longer observation periods. The availability of LUCAS 2022 data will allow further assessment of this effect.

Compared to the *state-first* approach, *change-first* models generally produced more variable SNR values across land-cover series for both δ and β. While both approaches yield variation in predicted changes across land-cover contexts, the *change-first* approach exhibits more stable uncertainty estimates across land-cover types, suggesting the presence of a baseline noise level in its predictions. As a result, differences in SNR across land-cover series are more strongly driven by variation in the predicted signal than by changes in uncertainty.

For specific land-cover contexts—such as *Woodland* for δ (Fig. 4.6) and both *Woodland* and *Grassland* for β (Fig. 4.7)—the *change-first* approach generally yields higher SNR values than the *state-first* approach, whereas lower SNR values are observed for other land-cover settings. Notably, these patterns appear to be more closely related to the presence of *Woodland* or *Grassland* in the land-cover history than to the occurrence of land-cover transitions themselves. For example, sites classified as *Grassland* throughout the entire observation period show higher SNR values than sites experiencing *Grassland–Cropland* or *Cropland–Grassland* transitions (Fig. 4.7), despite the latter being expected to exhibit larger SOC changes. Although increasing vegetation cover can mask soil signals and reduce detectability from a sensing perspective, dense vegetation is often associated with higher SOC levels, which may counterbalance this effect. This observation is consistent with the findings of Broeg et al. (2024b), who reported that higher SNR is associated with larger background SOC signals, which are commonly found in *Woodland* and *Grassland* systems.

4.4.4 Spatial aggregation improves SNR, but not always

The results confirms that spatial aggregation can improve SNR (Fig 4.9). For δ, noise decreased more rapidly than signal with increasing aggregation size, resulting in a clear increase in SNR. The reduction in noise is primarily a statistical effect: when aggregating over larger spatial units, random and spatially uncorrelated errors tend to cancel out (Smith, 2004; Wadoux and Heuvelink, 2023; Wadoux and Heuvelink, 2025). In contrast, the aggregated signal becomes smaller and less variable as it is dominated by the prevalence of small or near-zero changes within the aggregation unit. For β, both signal and noise decrease with increasing aggregation size at relatively similar rates, leading to largely stable SNR values, with only a slight increase at the largest aggregation scale (200 km^2). This difference likely reflects the fact that β is estimated from multiple time steps, which already suppresses part of the short-term variability and measurement noise during the trend-fitting stage. As a result, spatial aggregation provides less additional noise reduction for $\hat{\beta}$ than for $\hat{\delta}$. Despite the weaker aggregation effect, $\hat{\beta}$ generally exhibits higher SNR

values than $\hat{\delta}$, indicating greater detectability of trend-based change estimates across spatial scales.

Another factor influencing SNR behavior is the spatial domain being mapped—specifically its environmental conditions, the underlying pattern of pixel-level SNR, and the spatial structure of model errors. In our study, the improvement was more pronounced in Spain (Figs. 4.9, 4.10). Although SNR also increased with aggregation in Germany, the improvement was smaller and insufficient to exceed one, even at 200 km^2 support. This contrast primarily reflects environmental and land-cover differences: Spain's more sparsely vegetated landscapes produced smaller signal magnitudes but also lower uncertainties than Germany (Figs. 4.6–4.7). Furthermore, the longer variogram range observed in Germany suggests that spatially correlated errors persist over larger distances, limiting the benefits of aggregation (Wadoux and Heuvelink, 2023).

The choice of aggregation size and unit also influences SNR behavior through the Modifiable Areal Unit Problem (MAUP) (Dark and Bram, 2007; Openshaw and Taylor, 1979). The results revealed the scale effect of the MAUP, where analytical outcomes vary with aggregation level. While SNR generally increased with coarser resolution, the trend was not strictly monotonic—for instance, in parts of northeastern Spain, SNR rose at 40 km^2 but declined again at 200 km^2 as low-SNR areas became dominant within the larger grids. The zonal effect of MAUP—where different zoning schemes (e.g., aggregation by regular grids or administrative units) yield distinct spatial patterns—also needs consideration, although it was not examined in this study. When pixel-level SNR is homogeneous within a unit, aggregation reduces noise and preserves the signal, thereby improving SNR, as observed in parts of northwestern and central Spain. In contrast, in heterogeneous areas, strong local signals may be masked by surrounding low-SNR regions, limiting the benefits of aggregation.

Aggregation strategies should also consider their practical relevance for MRV, which depends on contextual factors such as land use, management practices, and local regulations that vary spatially (Gocke et al., 2023; Smith, 2008). For instance, management may differ considerably between neighboring farms, yet in some areas, practices are relatively homogeneous due to shared traditions or policy requirements. In this study, we used regular grids for demonstration purposes, but in practice, aggregation units should reflect local land-use and management contexts. For example, in the Netherlands, farmers are legally required to grow cover crops after maize cultivation (Fan et al., 2020), resulting in relatively uniform management across large areas. In such cases, aggregation at administrative or farm levels may be more meaningful than uniform grid cells.

However, this improvement comes with a trade-off: increasing aggregation enhances detectability but reduces spatial detail and local interpretability. Although higher SNR with aggregation is statistically expected, it does not necessarily imply improved model skill. Depending on the aggregation scheme, aggregation may yield little benefit or lead

to apparent improvements less meaningful or representative. Ultimately, incorporating domain knowledge, local context, and intended use into aggregation design is necessary to ensure the interpretability and policy relevance of SOC change assessments.

4.4.5 Strengths and challenges of using LUCAS for SOC change modeling

LUCAS forms the cornerstone of this study. As of this writing, three survey rounds (2009/2012, 2015, and 2018) are publicly available, covering a total of nine years—still a relatively short period for reliable modeling SOC change using ML and EO data alone. In Denmark, decadal sampling has been suggested as sufficient to capture long-term SOC trends, as short-term fluctuations may not reflect meaningful change (Harbo et al., 2023). The upcoming release of LUCAS 2022 will extend the record beyond a decade and provide an additional time step for trend analysis, increasing the potential for detecting long-term SOC dynamics.

While LUCAS is an exceptional resource, several inherent characteristics limit its suitability for SOC change modeling. Designed to represent all of Europe with roughly 20,000 sites (Orgiazzi et al., 2018), it effectively captures broad spatial patterns. However, this continental scope can mask subtle local or temporal variations in SOC, particularly in large-scale mapping frameworks where a single generalized model is applied across all locations and time steps (Lemercier et al., 2022). In addition, the relatively low sampling density of such continental surveys restricts their suitability for fine-scale spatial aggregation, as fitting spatial variograms requires sufficient site density at short distances. For this reason, denser national surveys were used in this study to derive standardized error variograms for the spatial aggregation analysis.

Coordinating sampling across many member states and survey years also poses challenges for maintaining consistency. Although all samples are analyzed using standardized methods, variations in both laboratory and field implementations may still introduce non-negligible differences—particularly given the subtle SOC changes being monitored. Laboratory-related inconsistencies, for example, have been reported in alkaline soils, where higher analytical uncertainties have been observed in SOC measurements (Schneider et al., 2021). Field-related inconsistencies are also reflected in the extreme apparent SOC changes observed (Fig. 4.3). In some cases, δ reaches 400 g/kg and β up to 50 $g/kg/year$—values that far exceed realistic expectations (Gubler et al., 2019; Poeplau and Don, 2013). Many of these outliers occur in *Woodland* areas, where inconsistent separation of organic litter from mineral soil layers is suspected (Hiederer, 2020; Ziche et al., 2022). Another likely source of such anomalies is the destructive nature of soil sampling: exact resampling of the same point is impossible. Although LUCAS mitigates this through composite sampling, the official data evaluation report (Hiederer, 2020) emphasizes that repeated samples should be interpreted as representative of a small area rather than a fixed point. Minor spatial displacement between survey rounds can therefore generate apparent SOC

differences unrelated to temporal change, particularly where spatial heterogeneity exceeds temporal variability (Broeg et al., 2024b; Poeplau et al., 2022).

Another source of uncertainty lies in the sampling method itself. In LUCAS, soils are sampled with a spade. While this approach facilitates continent-wide soil surveys at relatively low cost, it raises concerns about whether spade-based samples accurately represent the upper 0–20 cm layer in certain land uses. Comparative studies from Switzerland indicate that hand-auger sampling provides greater precision, particularly in *Grassland* and forest soils with pronounced SOC depth variation (Fernández-Ugalde et al., 2020). Furthermore, repeated LUCAS soil measurements are currently limited to the 0–20 cm topsoil layer. Although this is where most SOC change typically occurs, management effects in croplands have been reported to extend to depths of up to 50 cm (Skadell et al., 2023).

Several refinements could further enhance LUCAS's capacity for SOC change modeling and monitoring. Permanent site marking or more detailed and standardized location referencing could help reduce relocation uncertainty. Likewise, replacing spade-based sampling with hand augers would improve depth control and the representativeness of samples within the 0–20 cm layer. LUCAS already provides detailed survey-level quality reports—an excellent practice that promotes transparency. Extending this to the individual record level, for example by including per-sample quality flags (analogous to pixel-level quality indicators in remote-sensing datasets), would further enhance data consistency and interoperability. National soil surveys—such as those from Spain and Germany used here—complement LUCAS by adding spatial detail and sampling depth, and stronger coordination with national monitoring networks and EO datasets could further improve temporal consistency and model robustness (Froger et al., 2024). While some refinements could increase precision, they would also raise costs. Therefore, we recommend prioritizing denser and more frequent resampling of existing LUCAS sites to build consistent long-term time series, supported by improved and standardized metadata for uncertainty assessment.

Taken together, these recommendations are not criticisms of LUCAS but reflect the challenges of applying a continental-scale survey to high-resolution SOC change modeling. LUCAS was designed to provide harmonized, long-term environmental data rather than detailed modeling inputs and already achieves a strong balance between spatial coverage, cost, and feasibility. The program has continuously evolved based on lessons from previous rounds, and the recent inclusion of subsurface layers (e.g., 20–30 cm) marks a valuable step toward capturing vertical SOC trends. Its continuation remains vital for SOC dynamics modeling and will further strengthen the empirical basis for model calibration and improve SNR.

4.4.6 Implications for SOC monitoring and DSM applications

Our findings support the conclusions of Broeg et al. (2024b): reliable SOC change detection at fine spatial resolutions (e.g., point, site, or pixel level) remains difficult when relying on EO data and ML methods alone. High prediction uncertainty, limited time-series length, and the inherently slow dynamics of SOC all contribute to this limitation. Making these constraints explicit and measurable is a necessary step toward more transparent interpretation of DSM-based SOC change products. In this context, SNR provides a quantitative diagnostic by explicitly relating predicted change magnitude to its associated uncertainty. Low SNR values should therefore be interpreted as a warning against over-interpretation of modelled SOC change, indicating that predictions are likely dominated by noise. SNR can nevertheless be used in spatial aggregation analyses, where aggregation may improve detectability. In addition, spatial patterns of SNR help identify regions with relatively higher or lower detectability, and locations or aggregation units with higher SNR may serve as starting points for investigating environmental or land-use conditions associated with more detectable SOC change signals.

Improving SOC change detectability will depend heavily on the quality of data. To achieve higher SNR, longer, consistent observational datasets—like those from the LUCAS program—are needed to improve the training, validation, and understanding of SOC change models. Beyond extending temporal depth or increasing spatial support, additional strategies can improve SNR, including the use of more relevant, higher-quality covariates and advances in model architectures (Heuvelink et al., 2021; Tian et al., 2025a). Hybrid approaches that combine the strengths of ML models and domain-specific process knowledge have emerged as a promising direction (Reichstein et al., 2019). Notable examples include constraining numerical models with domain knowledge (De Rosa et al., 2024), coupling process-based and ML models through loss functions (Zhang et al., 2024a), replacing sub-modules of complex process-based models with ML components (Liu et al., 2024), and augmenting limited temporal SOC datasets with synthetic data generated by process-based models (Zhang et al., 2024a).

That said, time series of high-resolution DSM products remain valuable. In fact, $\hat{c}$ often yield SNR larger than one, demonstrating their value in capturing spatial SOC patterns. Moreover, this SNR evaluation framework can be adapted to work with typical DSM outputs. These predictions can also be aggregated to coarser spatial units, where the resulting SOC changes tend to exhibit higher SNR, indicating more certain and interpretable trends. Given the limited availability of repeated SOC measurements, especially in regions without long-term monitoring programs like LUCAS, the *state-first* approach will likely remain the most viable strategy for detecting and reporting SOC change in the near future. As long as corresponding SNR values are reported alongside SOC change estimates, this approach still support reliable interpretation of model outputs and help identify areas

of meaningful change. We advocate for more inclusion of SNR metrics in SOC change estimate efforts to better account for uncertainty and guide data-informed decisions.

4.4.7 Limitations

The SNR analysis framework developed in this study evaluates the reliability of SOC dynamics modeling by comparing predicted SOC change magnitudes (signal) against model-derived uncertainty (noise). Since both components are produced by the ML model, the validity of the resulting SNR values depends on the accuracy of both the predictions and the associated uncertainty estimates. In this study, we employed RF and QRF, which are commonly used in ML-based DSM to generate predictions with uncertainty (Wadoux et al., 2020b). While we expect our conclusions to hold for other models with robust uncertainty quantification, further work is needed to confirm their generalizability.

Additionally, while we included the trend-based metric β in our analysis, it was estimated from only three SOC measurements over time. Such limited temporal depth introduces uncertainty in trend estimation. Given current data availability, this remains the most feasible approach. However, the forthcoming LUCAS 2022 dataset is expected to provide an additional time step, enabling more reliable trend estimates and reinforcing the findings of this study.

Although predicted changes inherently contain both magnitude and direction, the SNR metric, as implemented in this study, quantifies magnitude-based detectability. With respect to direction, under the assumption of symmetric prediction uncertainty, a sufficient but not necessary condition for directional inference is an SNR value exceeding 0.5, which implies that the prediction distribution lies entirely on one side of zero change. In this study, however, SNR values remain below this level for the vast majority of predictions, substantially limiting the practical relevance of directional inference under current data availability. When the conditional prediction distribution is asymmetric, directionality cannot be inferred from a single summary measure of uncertainty; upper and lower uncertainty bounds must be considered separately. Under such conditions, change direction may still be resolvable even when overall SNR remains low. While direction of change is an important aspect of SOC dynamics, its reliable assessment requires a dedicated methodological framework, potentially involving classification-based models and explicit treatment of directional uncertainty, which which lies beyond the scope of the present study and could be addressed in future work.

While this study provides a quantitative demonstration of how spatial aggregation can improve the SNR, the analysis was intentionally simplified to illustrate the concept rather than to exhaustively evaluate all possible aggregation scenarios. The results should therefore be viewed as a demonstration of the general potential to identify SOC trends more reliably at larger spatial scales, rather than as an attempt to quantify national or regional SOC changes (e.g., in Germany or Spain). Land use and management—key

drivers of SOC dynamics—were not explicitly considered in this analysis due to the lack of harmonized, temporally resolved management data at the continental scale. Land-use data at large scales, although increasingly available through ML-based EO products, remain subject to uncertainty—just like to that of SOC maps derived from DSM. Data on management practices are even scarcer: existing datasets are typically static in time (Sandström et al., 2023) or available only at coarse spatial resolution (Lampach et al., 2025). The omission of these factors means that part of the observed noise, particularly at the point scale, may reflect unaccounted management heterogeneity. Moreover, the spatial aggregation analysis was restricted to the *state-first* approach using the RF_c model. Extending the framework to RF_δ or RF_β models would in principle be possible, but would require denser temporal datasets to support reliable variogram fitting—data that are not yet available.

Future work should aim to integrate contextual information—such as land use and management practices—into both SOC dynamic modeling and spatial aggregation analyses. Incorporating these factors would strengthen the mechanistic understanding of SOC change and enhance the policy relevance of results. Further research should also systematically evaluate how different aggregation schemes—varying in both scale and zoning design—affect SNR behavior across contexts (e.g., farm management, policy reporting, and global monitoring). Such analyses would help clarify the mechanisms by which aggregation enhances or diminishes detectability and provide empirical guidance for designing effective MRV systems. Without detailed land-use and management information or a systematic assessment of aggregation effects, the interpretability of aggregated SNR values—and thus their usefulness for understanding underlying SOC processes—remains limited.

4.5 Conclusions

With SNR consistently below one and low prediction accuracy, our study demonstrates that, under current data availability, modeling SOC dynamics at the site level using ML and EO data alone remains infeasible. The limited number and frequency of repeated SOC observations constrain the ability to capture meaningful temporal trends. Both the *state-first* and *change-first* approaches exhibit similarly poor predictive performance, although the *change-first* method provides more consistent yet typically narrower uncertainty estimates. While a longer monitoring period does not necessarily increase SNR, incorporating more time steps into SOC change quantification tends to improve it. Higher SNR values are also observed in land covers with greater vegetation content, such as *Woodland* and *Grassland*.

SNR, as an internal metric for assessing the detectability of modeled SOC changes, cannot replace independent validation. It offers a valuable indicator of change detectability given the scarcity of repeated SOC observations and can be applied to evaluate the prevailing practice of mapping SOC dynamics through the *state-first* approach. When repeated SOC

measurements become available, SNR could complement traditional performance metrics by providing spatially explicit insights into the reliability of modeled SOC changes. We argue that SNR should be routinely reported when interpreting or claiming SOC change from model predictions. Without explicitly accounting for uncertainty, such maps risk overstating their reliability for MRV purposes.

Finally, our analysis confirms that spatial aggregation improves SNR, with the degree of improvement depending on both the spatial structure of model predictions and the choice of aggregation units. This suggests that, while site-level assessments remain challenging, reliable SOC change modeling is still achievable at larger scales. This offers a promising pathway for regional-scale monitoring and reporting. Nonetheless, careful selection of aggregation scales and units is required to ensure that reported SNR values remain meaningful and representative.

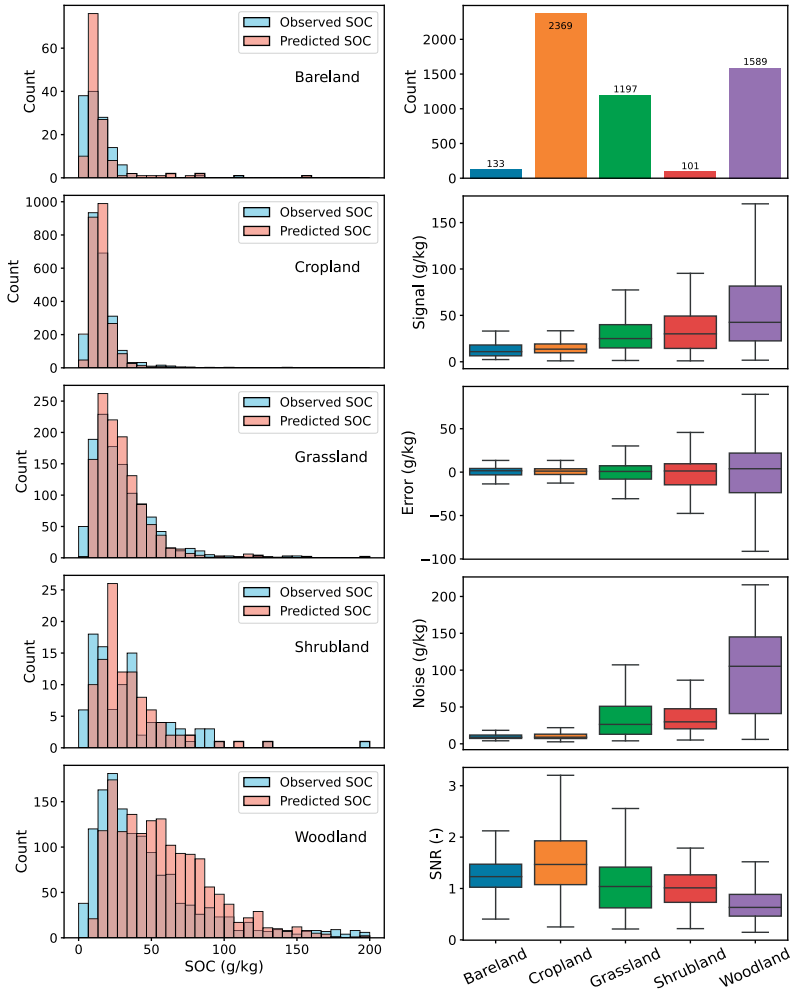


Figure 4.5: Analysis of the SNR for $\hat{c}$ by RF_c. Left: Histograms of observed and predicted SOC values across different land cover types. Right: Summary statistics of sample count, signal ($\hat{c}$), error, noise, and SNR derived from the independent test set for each land cover type. In the right panel, boxes indicate the interquartile range, and whiskers represent the remaining distribution excluding outliers. Note that both plots are based solely on the independent test set.

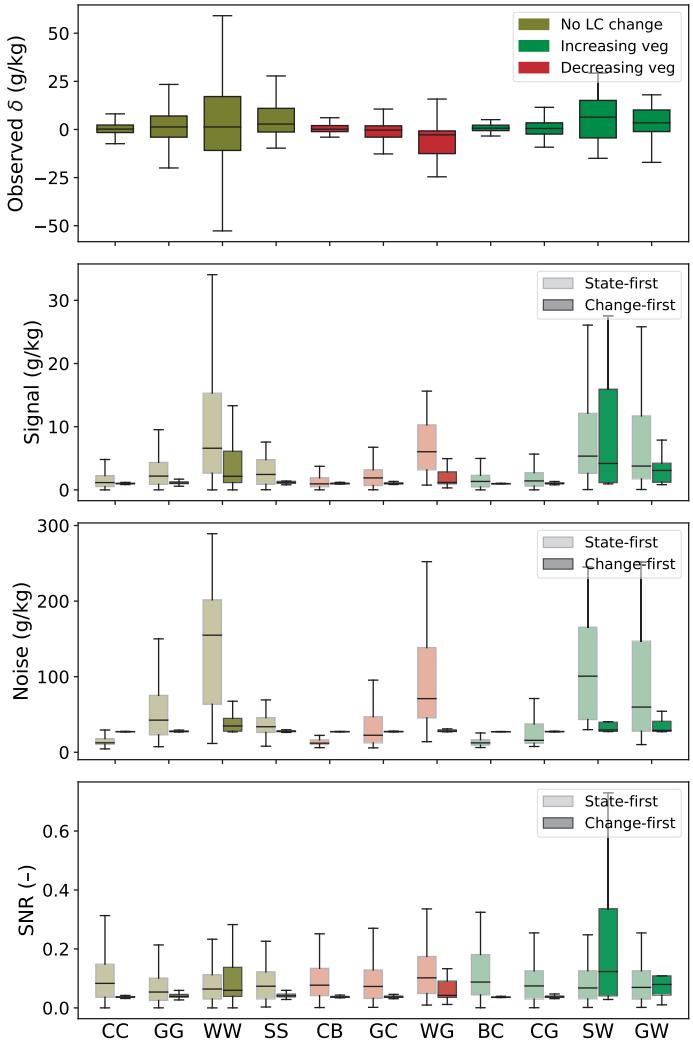


Figure 4.6: Summary statistics of observed δ, as well as signal, noise, and SNR of $\hat{\delta}$, comparing results from the RF_δ under the *change-first* approach and the RF_c under the *state-first* approach, grouped by land cover transition pairs. Bars represent median values, and whiskers indicate the 5th to 95th percentile range. Land cover codes are as follows: B = Bareland, C = Cropland, G = Grassland, S = Shrubland, and W = Woodland. For example, "CC" indicates that the location was classified as Cropland in both sampling years. Based on the general increase in vegetation content along the gradient B → C → G → S → W, land cover transitions are grouped into three categories shown in the top panel: no change, increasing vegetation content, and decreasing vegetation content.

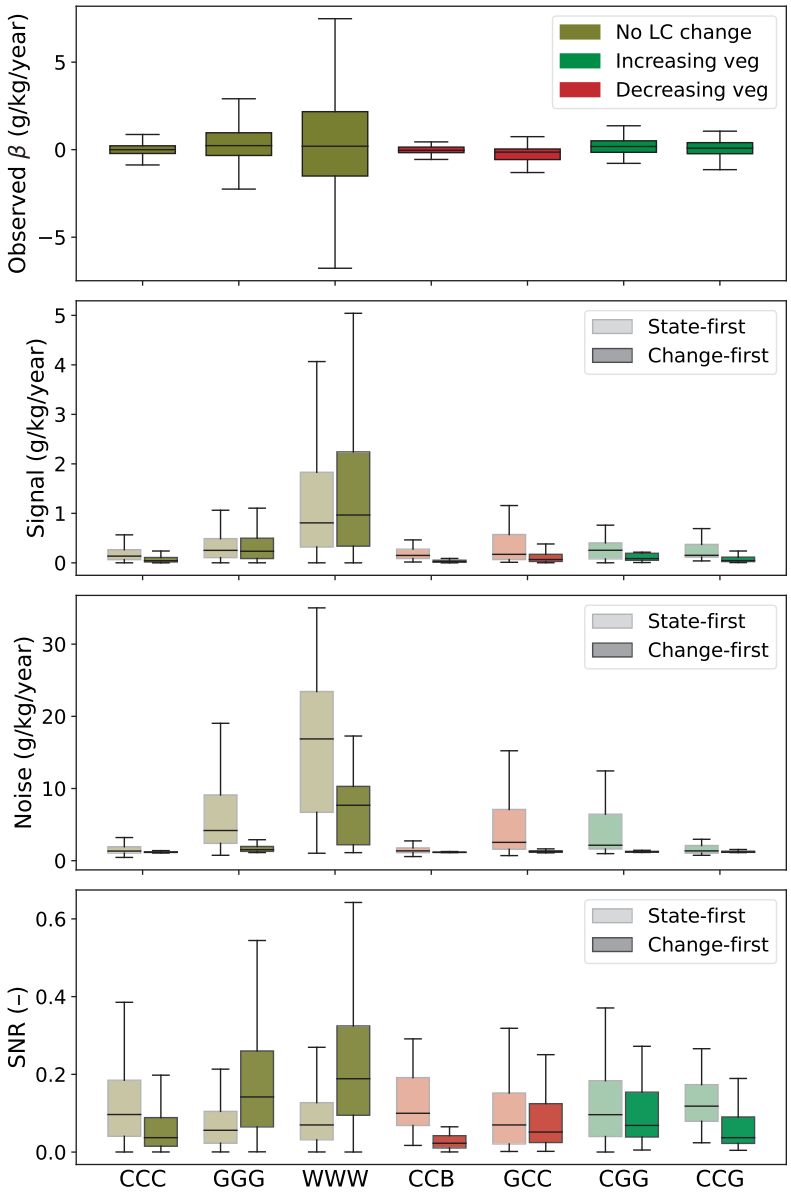


Figure 4.7: Summary statistics of signal (top panel), noise (middle panel), and SNR (bottom panel) for β, comparing results from the β-model under the *change-first* approach and the c-model under the *state-first* approach, grouped by land cover time series. Bars represent median values, and whiskers indicate the 5th to 95th percentile range. Land cover codes follow those used in Fig. 4.6, and time series are grouped into three categories: no change, increasing vegetation content, and decreasing vegetation content.

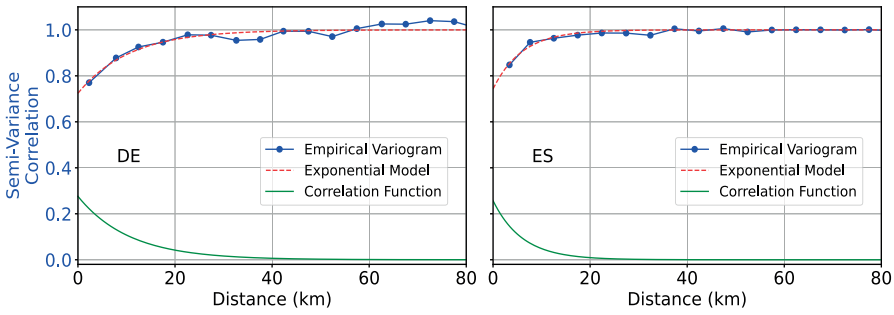


Figure 4.8: Variograms and corresponding correlation functions of standardized prediction errors for c in Germany (left, nugget: 0.72, range: 10.60) and Spain (right, nugget: 0.74, range: 5.99).

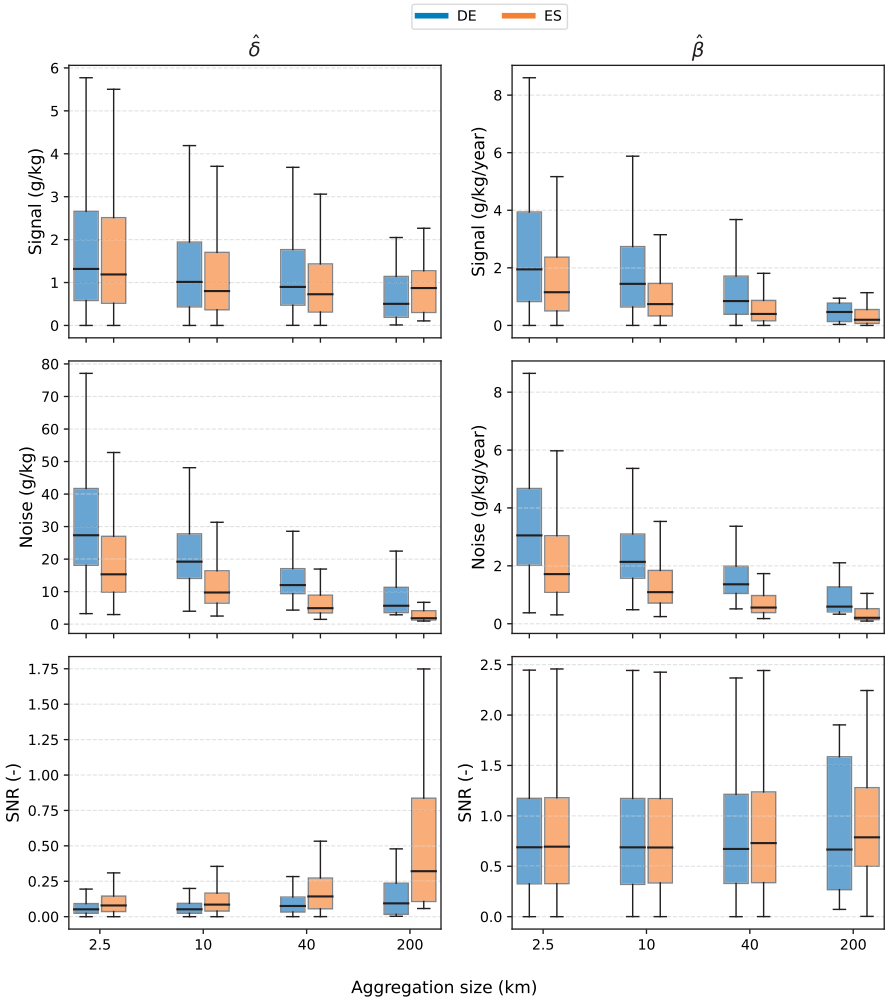


Figure 4.9: Box plots of signal, noise, and SNR for $\hat{\delta}$ and $\hat{\beta}$ in Germany and Spain for spatial aggregation unit sizes of 2.5 km, 10 km, 40 km and 200 km. Boxes represent the interquartile range, and whiskers indicate the remainder of the distribution. Note the difference in the y-axes.

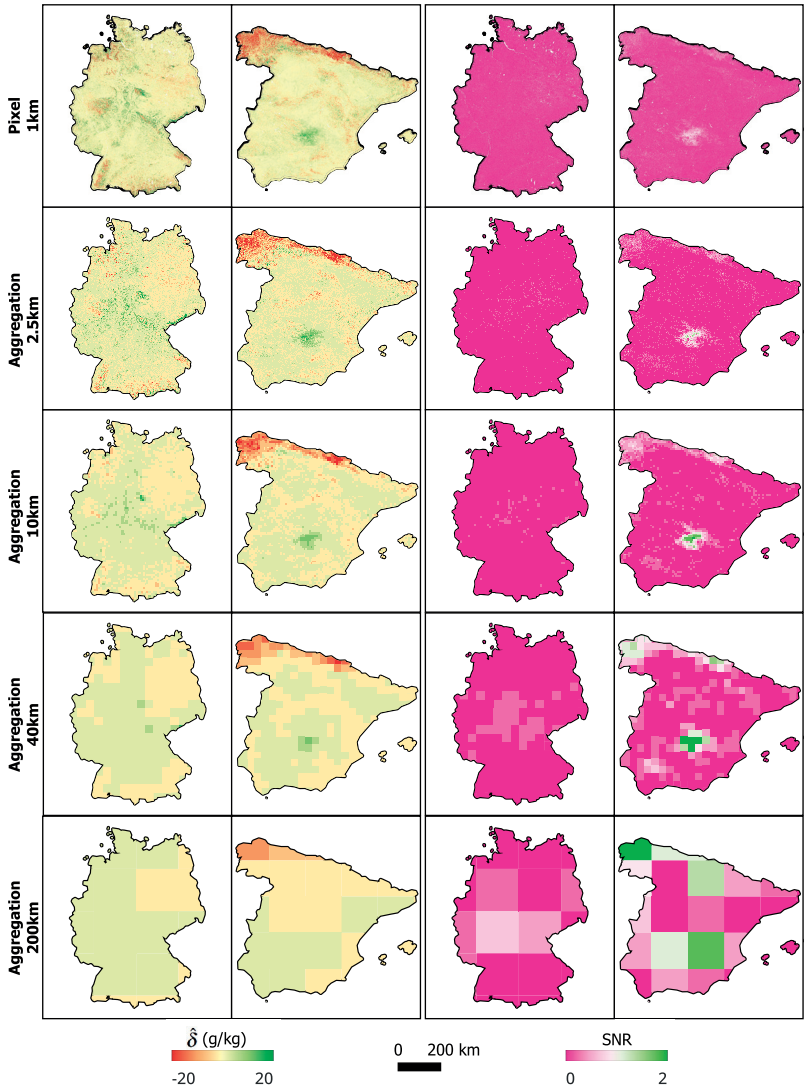


Figure 4.10: Spatial distribution of $\hat{\delta}$ (left two columns) and corresponding SNR (right two columns) from 2009 to 2018 in Germany and Spain, quantified using the *state-first* approach. Results are shown across increasing spatial aggregation scales: 1km pixel-level predictions, and aggregated units of 2.5 km, 10 km, 40 km and 200 km.

Chapter 5

Soil science-informed neural networks for soil organic carbon density modelling under scarce bulk density data

This chapter is based on:

X. Tian, B. Ahrens, L. Rossdeutscher, L. Alonso, and L. Parente (2026a). "Soil science-informed neural networks for soil organic carbon density modelling under scarce bulk density data". *EGUsphere* 2026, 1–22. DOI: 10.5194/egusphere-2026-229

Abstract

Soil Organic Carbon (SOC) density is a key variable for quantifying soil carbon stocks, yet its modelling is challenged by sparse and inconsistent measurements of Bulk Density (BD) and coarse fragments relative to SOC content. Conventional Digital Soil Mapping (DSM) approaches typically model SOC density as a single target variable, thereby underutilising abundant SOC content data and overlooking physical relationships among soil properties. This study evaluates a soil science-informed neural network for SOC density prediction that explicitly constrains the SOC–BD relationship, and compares it with univariate and multivariate neural network architectures. Across sparsely sampled target variables, including SOC density, BD, and coarse fragments, the soil science-informed model achieves comparable or slightly improved prediction accuracy relative to multivariate and univariate models. Although it yields lower accuracy for SOC content, the soil science-informed model better preserves physically plausible SOC–BD joint distributions and generates smoother, more temporally stable SOC density trajectories. Overall, the results demonstrate that incorporating soil physical constraints into Machine Learning (ML) models adds value beyond univariate accuracy, improving robustness, plausibility, and temporal coherence of SOC density predictions under sparse data conditions. Moreover, the latent parameters inferred by the soil science-informed model improve model interpretability and offer additional soil science relevant insights beyond predictive accuracy.

5.1 Introduction

Soil Organic Carbon (SOC) is fundamental for soil health and is increasingly recognised for its critical role in climate-change mitigation (Lal, 2004; Lehmann et al., 2020). For practical applications, stakeholders are primarily interested in SOC density ($kg\,m^{-3}$) rather than SOC content ($g\,kg^{-1}$), as SOC density directly determines SOC stocks. Digital Soil Mapping (DSM) provides a means to generate spatially explicit information on soil properties, including SOC. The increasing availability of SOC measurements and Earth Observation (EO) data has stimulated a rapid growth of Machine Learning (ML) applications in DSM, resulting in numerous SOC map products at global, continental and national scales (Wadoux et al., 2020b).

In DSM, SOC density is typically modelled in a univariate way: SOC density is calculated from observed SOC content, Bulk Density (BD, $g\,cm^{-3}$), and Coarse Fragments (CF, unitless fraction between 0 and 1) (Hengl and MacMillan, 2019), and then learned as a single target variable. However, SOC content is usually measured far more frequently than Bulk Density (BD) and CF, and even when BD and Coarse Fragment (CF) are available (Holmes et al., 2021), inconsistencies in measurement protocols can make them incompatible for direct SOC density calculation (e.g., BD measured for fine earth only while CF reported as mass fraction; Poeplau et al. 2017). For example, in the Land Use/Cover Area Frame Statistical Survey (LUCAS) Topsoil survey, SOC content is available for 2009/2012, 2015 and 2018, with only SOC content available through all three surveys (Orgiazzi et al., 2018). This mismatch leads to a "waste" of SOC content data and limits our ability to model SOC density dynamics.

Focusing solely on SOC density may also obscure the intrinsic physical relationships among SOC content, BD, and coarse fragments. Different combinations of SOC content, BD and CF can yield similar SOC density values, yet correspond to distinct environmental and soil conditions represented by different regions of the feature space. Ignoring these relationships effectively discards valuable information and may limit the model's ability to learn physically meaningful patterns, ultimately undermining predictive performance and robustness.

Multivariate modelling has long been used in DSM to address such interdependencies, ranging from co-kriging (Heuvelink et al., 2016) and structural equation modelling (Angelini et al., 2017) to more recent machine-learning approaches. Multivariate ML has been applied to jointly model SOC, total nitrogen and C:N ratio (Van Der Westhuizen et al., 2023), soil properties at multiple depths (Taghizadeh-Mehrjardi et al., 2020; Wadoux, 2019b), multiple soil attributes (Ng et al., 2019; Padarian et al., 2019), and the three texture components (Ließ and Sakhaee, 2024; Taghizadeh-Mehrjardi et al., 2020). However, to our knowledge, it has not yet been applied to jointly model SOC concentration, BD, coarse fragments and SOC density within a single modelling framework.

Beyond multivariate modelling, another way to account for relationships among soil variables is to explicitly incorporate knowledge about relationships among soil properties into machine-learning models, which has been referred to as soil science–informed ML (Minasny et al., 2024). Embedding domain knowledge can constrain model behaviour and prevent predictions that violate known soil processes. In univariate SOC dynamics modelling, purely data-driven ML approaches have been shown to produce noisy SOC time series, sometimes exhibiting implausibly large changes over short time spans (Tian et al., 2026b). Hybrid approaches that combine soil process understanding with ML have been reported to yield more realistic predictions than those derived from purely data-driven models (Zhang et al., 2024a; Zhang et al., 2023b). Although the mechanistic relationship between SOC concentration and BD has been well documented in soil science (Adams, 1973; Federer et al., 1993; Robinson et al., 2022; Stewart et al., 1970), its integration into DSM frameworks remains limited.

The aim of this study is to compare multivariate and soil relation–informed modelling strategies for predicting SOC density under sparse data conditions while explicitly accounting for inter-variable relationships. Among ML approaches, neural networks (NNs), compared to more conventional models such as Random Forests, Cubist, and partial least squares regression, are particularly well suited for multivariate soil property mapping. This is due to ease-of-use in representing complex dependencies and their flexibility to accommodate data sparsity, where not all observations contain measurements for all target variables (Ng et al., 2019; Padarian et al., 2019). However, explicitly incorporating expert soil knowledge into ML-based DSM remains challenging (Ma et al., 2019; Wadoux, 2019b), as ML models are inherently data-driven and the soil processes they represent are implicitly inferred from the training data (Wadoux et al., 2020b). The `EasyHybrid.jl` package developed by Alonso et al. (2025) provides a practical framework for embedding mechanistic constraints directly into NN architectures. Therefore, in this study, all three modelling strategies are implemented using neural networks.

Using the LUCAS topsoil dataset, three model structures are compared: (i) a univariate neural network (UniNN), (ii) a multivariate neural network (MultiNN), and (iii) a soil relation–informed neural network (SiNN). The models are evaluated with respect to their ability to predict SOC density, as well as the soil properties from which SOC density is derived, namely SOC content, BD and CF. Using the subset of LUCAS 2018 with complete measurements for all four properties, we further examine whether MultiNN and SiNN can exploit the abundant SOC concentration data from 2009/2012, 2015, and the incomplete part of 2018 to reconstruct SOC density under sparse BD availability. Inter-variable relationships are also examined. As CF are excluded from SOC content and fine-earth BD determinations in this study, the assessment focuses primarily on the SOC–BD relationship. In addition, we assess whether the hybrid SiNN model produces more temporally stable and physically realistic SOC density trajectories. Finally, plausibility checks are performed

on the latent representations of BD components inferred by the SiNN to evaluate the soil-science consistency of the model outputs.

5.2 Materials and Methods

5.2.1 Soil data

This analysis uses LUCAS soil data from the 2009/2012, 2015, and 2018 campaigns. Besides SOC content $g\,kg^{-1}$, the key properties are the volumetric form of CF (fraction, unitless between 0-1) and the fine-earth BD ($g\,cm^{-3}$), excluding coarse fragments larger than 2 mm, taken from the curated dataset of Pacini et al. (2023). When all three variables are available, SOC density ($kg\,m^{-3}$) is calculated as:

$$\mathrm{SOC}_{\mathrm{density}} = \mathrm{SOC}_{\mathrm{content}} \cdot \mathrm{BD}_{\mathrm{fe}} \cdot (1 - \mathrm{CF}_{\mathrm{vol}}) \tag{5.1}$$

This study focuses on topsoil (0–20 cm); thus, LUCAS 2018 measurements at 0–10 and 20–30 cm were excluded. For sites with three repeated measurements, we applied a quality filter based on temporal consistency of SOC content: assuming SOC changes at less than 0.5 $g\,kg^{-1}\,yr^{-1}$, we used a conservative threshold of 50 $g\,kg^{-1}\,yr^{-1}$ for the maximum absolute difference across measurements (Gubler et al., 2019; Poeplau et al., 2011). After removing time series exceeding this threshold and discarding records lacking required covariates, 56,117 measurements remained. All of them include SOC content, among them 5,194 also contain CF and BD. Regarding temporal replication, 11,690 sites have three measurements, 7,309 have two, and 6,429 have one.

Table 5.1: Summary statistics of soil properties: SOC content ($g\,kg^{-1}$), BD ($g\,cm^{-3}$), CF (fraction) and calculated SOC density ($kg\,m^{-3}$) per year.

Year	N sites	SOC content	BD	CF	SOC density
2009	17817	38.266 ± 72.672	-	-	-
2012	1777	19.517 ± 9.528	-	-	-
2015	19780	32.440 ± 53.596	-	-	-
2018	16743	34.836 ± 57.591	1.032 ± 0.302	0.061 ± 0.055	23.411 ± 17.764

5.2.2 Covariates

A wide range of covariates (also referred to as *predictors* or *features*) was included to represent environmental factors influencing SOC dynamics. Some covariates were available as time series (e.g., precipitation), providing multiple values for the same location across time. Others were static or long-term summaries (e.g., topography), offering a single value per location. All covariates were rescaled to the [0,1] range to ensure their suitability in

the neural network, and variables lacking informative values within Europe were excluded from the analysis. This results in a covariate set with 362 individual covariates, from 15 covariate groups (Table 5.2). The full list of covariates is not included in the main text but is available in the supplementary materials at our Supplemetary Repository.

Table 5.2: Covariate groups used in this study. Note that each entry represents a group of covariate layers, which may include multiple data layers with different temporal and spatial resolutions.

Covariate group	Source	Temporal resolution	Spatial resolution
Digital terrain model and derived land surface parameters	Ho et al., 2025	Static	30, 60, 120, 240, 480, and 960 m
Lithology type probability	Isik et al., 2024	Static	250 m
Plant functional type	Harper et al., 2023	Annual	300 m
Vegetation cover fraction	Sun et al., 2023	Annual	500 m
Precipitation	Karger et al., 2021	Annual	1 km
Land surface temperature	Wan, 2006	Annual	1 km
Water vapor	Lyapustin and Wang, 2018	Annual	1 km
Bioclimate	Karger et al., 2017	Long-term	1 km
Landsat spectral index	Tian et al., 2025b	Annual, long-term	30 m
Bare surface reflectance	Rogge et al., 2018	Long-term	30 m
Sentinel backscatter	Wagner et al., 2021	Long-term	25 m
PALSAR backscatter	Shimada and Ohtaki, 2010	Static	25 m
Peatland indicator	Widyastuti et al., 2024	Long-term	1 km
Cropland indicator	Potapov et al., 2022	Annual	30 m
Soil moisture	Bauer-Marschallinger et al., 2018	Annual	1 km

5.2.3 Model architectures

Three NN architectures are evaluated and compared in this study: UniNN, MultiNN, and SiNN (Fig. 5.1). To improve training stability and predictive performance of the NN, all four target variables—BD, SOC content, CF, and SOC density—were transformed

to reduce skewness and to constrain their values to the $[0, 1]$ range. This was achieved through log transformation and scaling using a standard scaler.

The UniNN model serves as a baseline that reflects the conventional approach in DSM, where each soil property is modelled independently using a separate NN. The MultiNN architecture advances this idea by training a single NN to simultaneously learn and predict four targets. All targets are treated as purely data-driven outputs in both UniNN and MultiNN. The SiNN architecture incorporates physical constraints directly into the learning process. In SiNN, SOC content and CF are predicted by the NN, while BD and SOC density are computed within the model using mechanistic equations. Here, BD is expressed as a function of SOC content and two latent BD parameters—organic BD (oBD) and mineral BD (mBD), following the well-established relationship of Federer et al. (1993):

$$BD = \frac{1}{\frac{SOM}{oBD} + \frac{1-SOM}{mBD}} \tag{5.2}$$

where the soil organic matter (SOM) could be converted from SOC content with a factor 1.724 ($SOM = 1.724 \cdot SOC_{\text{content}}$). To constrain the latent parameters oBD and mBD and to facilitate model convergence, prior ranges and recommended initial values are specified. Following Robinson et al. (2022), oBD is constrained to the range 0.05–0.40 $g\,cm^{-3}$, with an initial value of 0.20 $g\,cm^{-3}$. Mineral BD (mBD) is constrained to the range 0.75–2.00 $g\,cm^{-3}$, with an initial value of 1.20 $g\,cm^{-3}$. SOC density is subsequently derived from SOC content, BD, and CF using the mechanistic formulation in Eq. 5.1.

Across all three architectures, the supervised training targets include the four soil variables (SOC content, BD, CF, SOC density), and the loss function jointly constrains the predictions for all variables. In the SiNN model, the latent parameters oBD and mBD are also learned during training and can be output to provide additional interpretability regarding soil physical properties. The model experiments are implemented using the EasyHybrid.jl package (Alonso et al., 2025).

5.2.4 Model evaluation

All models were trained and evaluated using a five-fold cross-validation scheme to ensure objective assessment while maximising the use of available data. In each iteration, three folds are used for hyperparameter optimisation and model training, one fold serves as the validation set for selecting the best-performing model, and the remaining fold is held out as the test set for final evaluation. The folds are constructed through random partitioning of the dataset. By iterating across all five folds, full-coverage predictions for all samples are obtained for subsequent analysis.

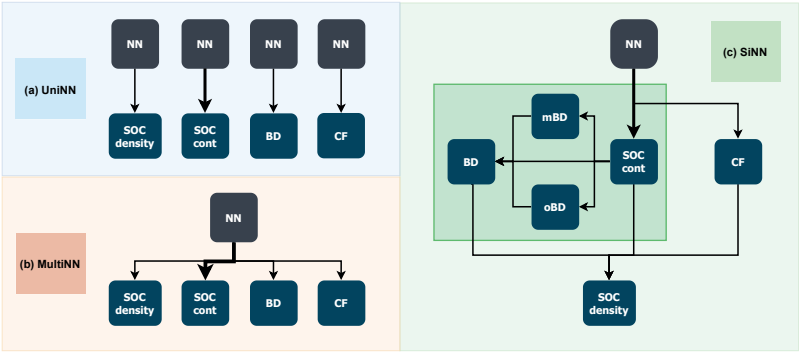


Figure 5.1: Schematic illustration of the three NN architectures evaluated in this study: (a) UniNN, in which a separate NN is trained for each soil variable; (b) MultiNN, in which a single NN is trained to predict all soil variables simultaneously; and (c) SiNN, in which SOC–BD relationships are explicitly embedded in the model structure. Dark grey boxes denote NN models, while dark blue boxes denote target variables and intermediate variables that are output by the model.

Model evaluation is performed at two levels. First, the predictive accuracy of each individual target variable is assessed using all available observations for that variable. Second, joint evaluation is conducted on the complete subset of samples where all four target variables are present. Both global accuracy and stratified accuracy across land-cover classes are examined. The metrics considered include the coefficient of determination (R^2), mean squared error (MSE) and bias, defined as the mean difference between predictions and observations.

Additionally, to assess the plausibility and consistency of BD–SOC content relationships, the joint distributions of predicted SOC content and BD were compared across the three neural network architectures and against observed data. Finally, the latent parameters oBD and mBD predicted by the SiNN model are examined with respect to soil texture, land cover, SOC content, and other soil properties to assess whether their spatial patterns and magnitudes are physically reasonable.

5.2.5 Temporal performance

The three models are evaluated in terms of temporal performance from two complementary perspectives: temporal plausibility and temporal transferability. Temporal plausibility refers to the coherence and stability of predicted SOC density time series, and is assessed using SOC density predictions obtained from cross-validation across multiple survey years. The underlying expectation is that the SiNN produces more coherent and less noisy SOC density trajectories than purely data-driven models.

Temporal transferability, in contrast, evaluates how well a model trained on one time period generalises to other time periods. This is assessed by training the models on data from 2018 and predicting SOC density for the remaining survey years. However, as SOC density measurements in the current dataset are available for only a single survey year, direct evaluation of temporal transferability for SOC density is not feasible. Therefore, an indirect assessment is conducted using SOC content, which is consistently measured across multiple survey years, as a proxy to evaluate temporal transferability.

5.3 Results

5.3.1 Model accuracy

Fig 5.2 shows the cross-validation performance of SOC density predictions for the three models. Overall, SiNN exhibits the highest predictive performance, followed by MultiNN and UniNN, as reflected by R^2, MSE, and bias. Among the three models, SiNN produces predictions that most closely align with the 1:1 line relative to the observations. However, the performance differences among the models are relatively small.

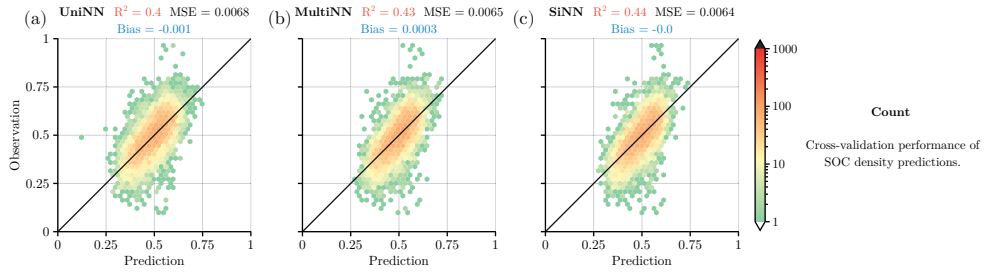


Figure 5.2: Accuracy of cross-validated SOC density predictions for the three models: UniNN, MultiNN and SiNN. All results are shown in the transformed target space used during model training.

These modest overall performance differences are also reflected in the land-cover-specific accuracy metrics, shown in Table 5.3. SiNN does not consistently outperform the other models across all land-cover classes or evaluation metrics. Based on R^2 and MSE, SiNN shows better performance in land covers typically associated with higher SOC levels, including *Grassland*, *Woodland*, and *Wetland*. Comparable performance is observed for *Cropland* and *Bareland*, whereas SiNN underperforms in *Artificial land*.

In terms of bias, UniNN tends to overestimate SOC density across most land-cover classes, with the exception of *Grassland* and *Shrubland*, resulting in an overall positive bias. In contrast, both MultiNN and SiNN exhibit overestimation in *Bareland*, *Cropland*, and *Wetland*, while showing slight underestimation in the rest land-cover classes.

Table 5.3: SOC density prediction performance by land cover and model; sample sizes for each land-cover class are reported in the last row.

Metric	Model	Artificial	Bareland	Cropland	Grassland	Shrubland	Woodland	Wetland
R^2	UniNN	0.30	0.45	0.40	0.39	0.08	0.15	0.53
R^2	MultiNN	0.26	0.44	0.41	0.41	0.21	0.19	0.84
R^2	SiNN	0.18	0.44	0.41	0.43	0.17	0.21	0.90
MSE	UniNN	0.0094	0.0057	0.0050	0.0068	0.0115	0.0094	0.0239
MSE	MultiNN	0.0098	0.0058	0.0049	0.0065	0.0098	0.0090	0.0079
MSE	SiNN	0.0110	0.0058	0.0049	0.0064	0.0103	0.0088	0.0049
Bias	UniNN	0.0025	0.0149	0.0070	-0.0033	-0.0163	0.0046	0.0148
Bias	MultiNN	-0.0086	0.0138	0.0034	-0.0044	-0.0183	-0.0008	0.0197
Bias	SiNN	-0.0066	0.0164	0.0033	-0.0042	-0.0204	-0.0024	0.0360
N	–	24	203	2339	1159	161	1306	2

5.3.2 Joint distribution of SOC content and BD

Figure 5.3 compares the observed and predicted joint distributions of SOC content and BD. The observations exhibit a clear wedge-shaped support, characterised by high SOC values occurring predominantly at low BD and an overall negative SOC–BD dependence. In the observations, BD spans a wide range (approximately 0–1.8 $g\,cm^{-3}$). In contrast, none of the three models fully reproduces the extreme ends of the BD distribution, with predicted values rarely falling below 0.2 $\mathrm{g\,cm^{-3}}$ or exceeding 1.6 $\mathrm{g\,cm^{-3}}$, except for two isolated cases produced by UniNN.

The negative SOC–BD relationship observed in the data is qualitatively reproduced by all three models. However, this relationship is more clearly expressed and more strongly constrained in the MultiNN and SiNN predictions, whereas UniNN exhibits a weaker coupling between SOC content and BD.

In the observations, SOC content at very low BD (<0.2 $g\,cm^{-3}$) shows a bimodal pattern, with both very high (400-500 $g\,kg^{-1}$) and low SOC values (<100 $g\,kg^{-1}$). While UniNN produces SOC predictions exceeding 200 $g\,kg^{-1}$, these occur mostly at moderate BD values rather than at low BD. MultiNN shows a reduced occurrence of such implausible combinations, while SiNN largely avoids them. Overall, the predicted SOC–BD distributions become progressively more constrained from UniNN to MultiNN to SiNN, with extreme and weakly supported combinations occurring less frequently.

5.3.3 Temporal performance

Temporal performance is assessed from two perspectives: temporal plausibility and temporal transferability. Plausiblity is assessed based on the stability of SOC density time series

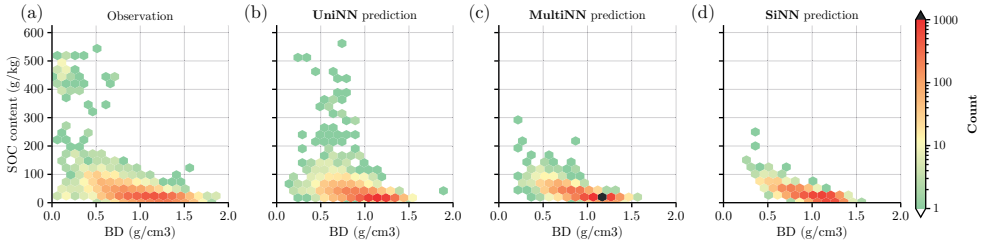


Figure 5.3: Joint distribution of SOC content ($g\,kg^{-1}$) and BD ($g\,cm^{-3}$) for observations and model predictions.

derived from cross-validation predictions, whereas transferability is evaluated by training the models on 2018 data and predicting other years.

The stability of predicted SOC density time series is quantified using the temporal range, defined as the maximum absolute change in SOC density within each predicted time series across survey years. The resulting range distributions (Fig. 5.4) are long-tailed for all three models, with the majority of values concentrated around a change magnitude of approximately 3 $kg\,m^{-3}$. Despite similar median ranges, clear differences emerge in the upper tails of the distributions. UniNN exhibits the largest extreme changes, followed by MultiNN, whereas SiNN shows a much reduced upper outliers. Overall, the SiNN produces smoother and more temporally stable SOC density trajectories.

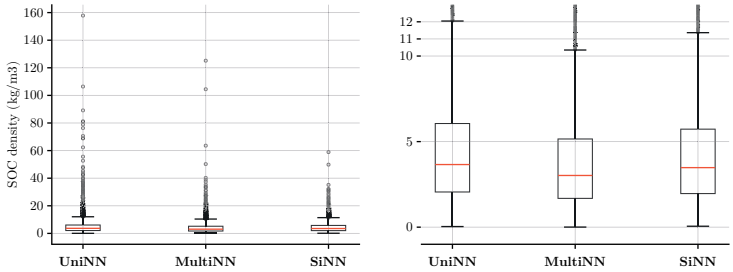


Figure 5.4: Temporal range of predicted SOC density time series across LUCAS survey years for the three models (left), with a zoomed-in view for detailed inspection (right). Quantile statistics of the temporal ranges are shown for UniNN (5% = 0.76, 50% = 3.66, 95% = 12.35), MultiNN (5% = 0.63, 50% = 3.02, 95% = 10.66), and SiNN (5% = 0.74, 50% = 3.48, 95% = 10.39).

Temporal transferability is assessed indirectly using SOC content. Among the three models, UniNN achieves the highest prediction accuracy when transferred to other survey years, followed by SiNN, while MultiNN shows the lowest accuracy. However, when these results

are compared with the cross-validation performance reported in Appendix X, a larger performance drop is observed for UniNN (from $R^2 = 0.59$ to $R^2 = 0.49$) and MultiNN (from $R^2 = 0.49$ to $R^2 = 0.42$) than for SiNN (from $R^2 = 0.48$ to $R^2 = 0.46$). This smaller degradation in performance indicates that the SiNN model exhibits greater temporal robustness.

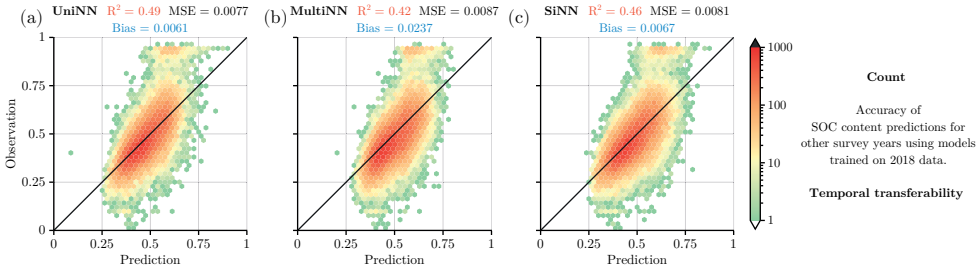


Figure 5.5: Accuracy of SOC content predictions for other survey years using models trained on 2018 data, assessing temporal transferability. All results are shown in the transformed target space used during model training.

5.3.4 Plausibility of oBD and mBD

An additional advantage of SiNN is that, beyond predicting the target soil properties, it simultaneously estimates the latent parameters oBD and mBD. Figure 5.6 shows the distributions of the latent oBD and mBD components across seven land cover classes. Overall, predicted oBD values range from approximately 0.1 to 0.3 $g\,cm^{-3}$ and exhibit a clearer land cover relevant pattern than mBD. From *Bareland* and *Cropland* through *Grassland* and *Shrubland* to *Woodland*, oBD decreases gradually from around 0.22 to approximately 0.15 $g\,cm^{-3}$, consistent with increasing vegetation cover. The oBD of *Artificial land* lies between that of *Cropland* and *Grassland* (around 0.2 $g\,cm^{-3}$), while *Wetland* exhibits the lowest mean oBD and largest spread among all land cover classes. Assuming peat as an organic end member, with reported peat BD values of approximately 0.1–0.2 $g\,cm^{-3}$ (Huat et al., 2011; Päivänen, 1969), and organic forest soils exhibiting oBD values around 0.11 $g\,cm^{-3}$ (Perie and Ouimet, 2008), the inferred oBD range appears physically plausible (Robinson et al., 2022).

In contrast, mBD values range from approximately 1.0 to 1.5 $g\,cm^{-3}$ and show much weaker differentiation across land covers. The mBD distributions for *Artificial land*, *Bareland*, *Cropland*, *Grassland*, and *Wetland* are centered around similar values (approximately 1.3 $g\,cm^{-3}$), although *Wetland* shows a slightly larger spread. *Shrubland* and *Woodland* exhibit slightly lower mBD values, with mean values below approximately 1.25 $g\,cm^{-3}$. Compared with mBD values estimated from soil data collected in Wales (1.9 $g\,cm^{-3}$), the

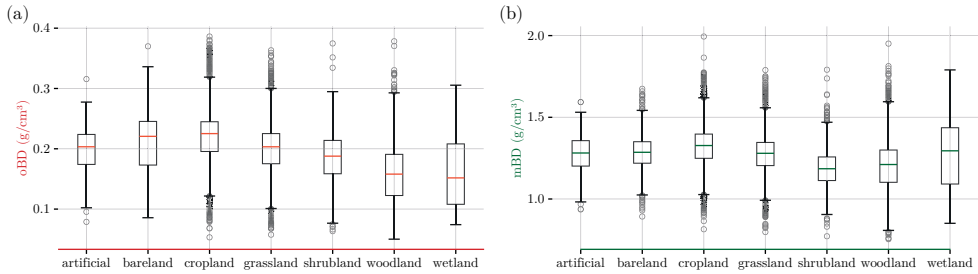


Figure 5.6: Distribution of latent parameters oBD (a) and mBD (b) across different land covers.

values inferred here are lower, but are comparable to those suggested by Stewart et al. (1970) (1.4 $g\,cm^{-3}$).

Fig 5.7 shows the distributions of the estimated oBD and mBD across the soil texture space. Both oBD and mBD exhibit lower values in fine-textured soils, particularly where sand content is below 10 %, corresponding to silty clay and clay textures. In contrast, higher oBD and mBD values are observed in coarse-textured soils characterised by high sand content (above 70 %) and low clay content (below 40 %), corresponding to sandy clay loam and sandy loam textures. In other texture regions, the patterns are more mixed.

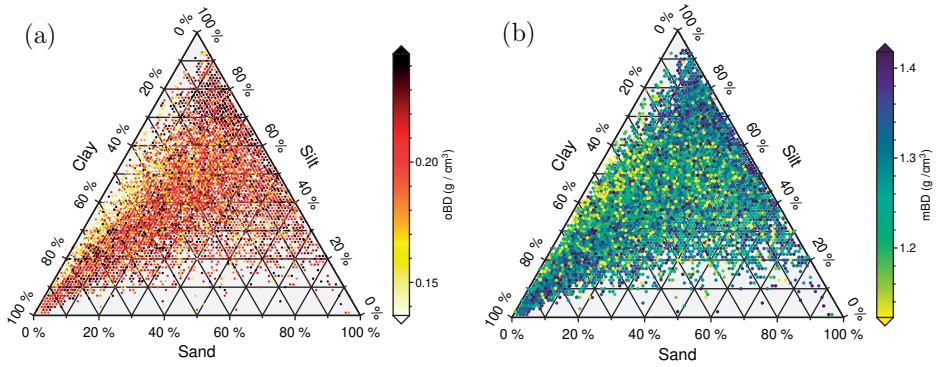


Figure 5.7: Distribution of latent parameters oBD (a) and mBD (b) across textures.

Fig 5.8 presents example maps of these latent parameters generated by SiNN. Consistent with the land-cover analysis shown in Fig. 5.6, clear land-cover and land-use patterns are evident: higher oBD and mBD values are observed in cropland, followed by artificial land associated with built-up areas, and lower values in woodland regions, which are indicated by dense vegetation (dark green) in the satellite imagery.

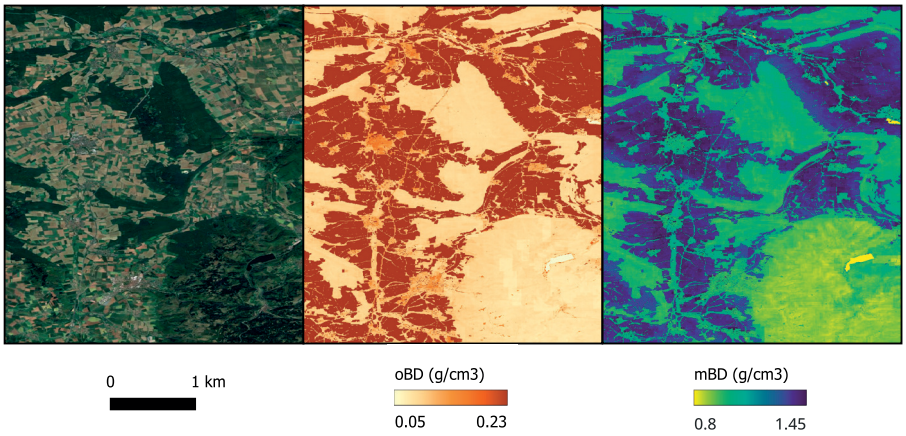


Figure 5.8: Maps of the predicted latent parameters oBD (middle) and mBD (right), shown alongside the corresponding satellite imagery (left). Predictions are generated at 30 m spatial resolution for the year 2018 over a region in central Germany (centre location: longitude 10.19320, latitude 51.98643). Satellite imagery: map data ©2025 GeoBasis-DE/BKG (©2009), Google.

5.4 Discussion

5.4.1 Model performance under sparse data conditions

Across all target variables with sparse data, including SOC density, BD, and CF, SiNN generally exhibits comparable or slightly better performance than MultiNN, followed by UniNN, with the clearest improvements observed for SOC density (see accuracy plots for other properties in Appendix Fig. 5.9). These improvements, however, are not consistent across all land-cover classes.

An exception is SOC content, for which prediction accuracy decreases from UniNN ($R^2 = 0.59$) to MultiNN ($R^2 = 0.49$) and further to SiNN ($R^2 = 0.48$). UniNN achieves the highest accuracy for SOC content, which is also the most available variable, while slightly underperforming for properties with sparser observations. From a purely univariate accuracy perspective, this suggests that different model structures may be preferable depending on data availability: the soil science-informed model is advantageous for SOC density prediction under sparse auxiliary data, whereas MultiNN achieves comparable accuracy for BD and CF with lower computational cost, and the univariate approach remains competitive for well-represented variables such as SOC content.

5.4.2 Plausibility of predictions from the perspective of soil science

The higher SOC content accuracy achieved by UniNN appears to be associated with its ability to predict very high SOC values, whereas MultiNN and SiNN tend to underestimate the upper tail of the SOC distribution (Fig. 5.9-a). However, when the joint SOC–BD relationship is considered, these high SOC predictions by UniNN are often implausible. In the observations, very high SOC content typically co-occurs with very low BD values (below approximately 0.2 $g\,cm^{-3}$), whereas UniNN frequently predicts such high SOC values at moderate BD levels (approximately 0.5–1.0 $g\,cm^{-3}$), thereby placing probability mass in sparsely supported regions of the SOC–BD space.

Temporal transferability experiments further show that UniNN exhibits the largest performance degradation when models trained on one survey year are transferred to other years. This pronounced drop indicates a higher sensitivity of UniNN to temporal domain shifts and a less stable generalisation across time, compared with the more constrained multivariate and soil science-informed model structures. Consistent with this finding, temporal plausibility analyses reveal that UniNN produces noisier SOC density trajectories, with larger extreme changes exceeding 60 $g\,cm^{-3}$. Such magnitudes are highly implausible over the nine-year LUCAS interval, even under strong anthropogenic intervention (Gubler et al., 2019; Poeplau et al., 2011). In contrast, MultiNN shows intermediate behaviour, while SiNN exhibits the most stable temporal dynamics. Overall, the soil science-informed model provides the most plausible and temporally coherent SOC density time series under temporal domain shifts.

5.4.3 Latent soil physical parameters

Beyond predicting the target soil properties, SiNN additionally yields latent estimates of oBD and mBD. Although direct ground truth data for these latent parameters are not available for validation, their inferred ranges are broadly consistent with values reported in the literature, and their plausibility is supported by systematic patterns observed across land cover classes and soil texture space. It should be noted that the learning of oBD and mBD is constrained by prior information specified in this study. In our case, these priors are knowledge-based and lead to physically reasonable and numerically stable solutions. However, fixed priors implicitly assume their validity beyond the calibration domain, and their suitability may vary. Therefore, the use of the SiNN requires selecting or adapting prior ranges based on domain knowledge and the specific environmental context of the application.

The latent parameter estimates also provide insights into the spatial variability of oBD and mBD. As illustrated in Fig. 5.8, clear land cover and land use associations are evident for both parameters. Such patterns are expected for oBD, as land cover and land use influence organic matter inputs and accumulation (Smith, 2008). For mBD, however, the interpretation is less straightforward. The observed spatial variation may reflect differences

in soil compaction associated with land use, differences in soil texture, or a combination of both. In the absence of independent ground truth data, these hypotheses cannot be explicitly tested within this study.

Estimates of oBD and mBD can be further linked to soil porosity and compaction characteristics (Robinson et al., 2022). Soil porosity, defined as the fraction of soil volume occupied by pore space, plays a key role in regulating aeration, water movement, and nutrient transport, and is therefore closely related to soil structure and ecosystem functioning (Assouline and Or, 2013; Hao et al., 2008; Pagliai and De Nobili, 1993). Using the latent parameters inferred by SiNN, soil porosity can be derived following the analytical model developed by Robinson et al. (2022). This illustrates that the SiNN enables large-scale porosity estimation by combining legacy soil observations with EO covariates, without requiring direct porosity measurements. As shown in Appendix 5.6.2, the derived porosity patterns are consistent with those reported by Robinson et al. (2022) based on soil data from Great Britain and Wales: porosity generally increases with SOC content and levels off at SOC contents of approximately 200 $\mathrm{g\,kg^{-1}}$ (Fig. 5.11).

Soil compaction is a major form of soil degradation (Nawaz et al., 2013; Shah et al., 2017). BD and its derivatives, such as packing density, are commonly used as proxies for soil compaction because compaction directly leads to increased BD (Panagos et al., 2024b; Stolf et al., 2011). However, BD is influenced not only by compaction but also by SOC content, which complicates its interpretation. By explicitly estimating mBD, SiNN rules out the influence of SOC on BD. When further normalised by the mineral particle density (mPD), assumed to be constant, the ratio mBD/mPD can be used as a proxy indicator of soil compaction. Using this compaction indicator, Fig. 5.11 compares its distribution across land-cover classes. The highest mean values are observed in *Cropland* and *Woodland*, with *Woodland* exhibiting the largest spread. These patterns may reflect the effects of mechanised operations associated with intensive cropping and forest silviculture (Nawaz et al., 2013). This is consistent with the findings of Panagos et al. (2024b), who reported a higher susceptibility of arable land to soil compaction using packing density as an indicator.

The case of porosity and compaction illustrate how the soil science-informed model structure enables interpretation of soil physical characteristics beyond the directly predicted variables, providing added soil science value and practical relevance for soil assessment. Furthermore, these latent parameter estimates also enhance model interpretability by explicitly linking SOC content and BD within the model structure. This improves transparency and supports a more process-consistent understanding of model behaviour.

5.4.4 Limitations

Despite the added value introduced by the soil science-informed structure, SiNN does not fully resolve the challenge of predicting highly organic soils. Both MultiNN and SiNN fail

to accurately reproduce the upper tail of SOC content associated with very low BD. This limitation can be viewed from two related perspectives: an underestimation of extreme SOC and BD values, and an incomplete reproduction of their joint relationship.

As illustrated in Fig. 5.10, highly organic soils—characterised by SOC content exceeding 200 $g\,kg^{-1}$ and BD below approximately 0.2 $g\,cm^{-3}$—remain at the lower edge of the predicted BD range for all models, yet their predicted BD values are generally higher than observed. While both MultiNN and SiNN largely fail to predict SOC content above 200 $g\,kg^{-1}$, they preserve the qualitative association between high SOC and low BD. In contrast, UniNN is able to generate high SOC predictions, but these are weakly constrained, often occurring at moderate BD values and failing to reproduce both the low-BD regime and the joint SOC–BD structure.

Two factors may contribute to this limitation. First, the input data may be insufficient to distinguish highly organic soils. They are intrinsically under-represented in the training set, and the available covariates may not capture the factors that differentiate these soils. Although most highly organic samples originate from *Woodland* and *Grassland*, many other samples from the same land-cover classes occupy similar covariate space while exhibiting lower SOC content and higher BD, thereby limiting the discriminative power of the predictors. Second, all model experiments are conducted in a transformed target space, with logarithmic scaling applied. This transformation compresses differences at the upper end of the distribution (e.g., between 200 and 600 $g\,kg^{-1}$), reducing the influence of extreme values during loss optimisation and contributing to systematic underestimation of the highest SOC contents.

A further limitation is the lack of repeated SOC density measurements across survey years, which prevents direct validation of predicted temporal changes. This constraint is intrinsic to the sparse availability of BD and CF data and necessitates indirect assessments of temporal transferability and plausibility using SOC content and relative change metrics. As future LUCAS survey rounds become available and provide more complete measurements, this limitation is expected to be alleviated.

In addition, this study focuses primarily on accuracy-based evaluation and does not include uncertainty quantification, which has become increasingly common in DSM to support map interpretation and use. Although the gains in predictive accuracy introduced by structural constraints are modest, such constraints may contribute to reduced predictive uncertainty. Given the availability of established uncertainty quantification methods for neural networks (Huang et al., 2025), incorporating such approaches is identified as a priority for future work.

5.5 Conclusions

This study evaluated three NN architectures for reconstructing SOC density time series while preserving consistent relationships between soil properties, particularly between SOC content and BD. The results show that the soil relation informed model generally achieved the most favourable balance between prediction accuracy, temporal stability, and joint SOC–BD plausibility. While improvements in SOC density accuracy are modest, SiNN consistently produced less noisy SOC density trajectories and better preserved the physically meaningful association between SOC and BD, followed by MultiNN and UniNN. In addition, the latent parameters estimated by SiNN enable the derivation, interpretation and assessment of soil physical properties, such as porosity and compaction-related indicators.

In contrast, the univariate model performed well for abundantly sampled variables such as SOC content, but failed to reproduce plausible joint SOC–BD behaviour while exhibiting greater sensitivity to temporal domain shifts. These findings highlight that, in soil property modelling and mapping, evaluating models solely based on univariate accuracy can be misleading. Incorporating multivariate structure and soil constraints provides added value by improving robustness, plausibility, temporal coherence, and interpretability from a soil science perspective.

5.6 Appendix

5.6.1 Complementary accuracy assessment and joint distribution analysis

Figure 5.9 presents complementary cross-validation accuracy plots for the three models for the remaining target variables: SOC content (a), BD (b), and CF (c). UniNN achieves the highest prediction accuracy for SOC content ($R^2 = 0.59$), substantially outperforming MultiNN ($R^2 = 0.49$) and SiNN ($R^2 = 0.48$). In contrast, for BD and CF, UniNN shows slightly lower performance compared to the multivariate and soil science-informed models.

Figure 5.10 shows the joint distribution of SOC content and BD for highly organic soils, defined by SOC content exceeding 200 $g\,kg^{-1}$ and BD below approximately 0.2 $g\,cm^{-3}$. Across all models, these samples remain at the lower edge of the predicted BD range; however, predicted BD values are generally higher than those observed. Both MultiNN and SiNN largely underpredict SOC content above 200 $g\,kg^{-1}$, although their predictions remain associated with low BD values. In contrast, UniNN is able to generate high SOC predictions, but these frequently occur at moderate BD levels.

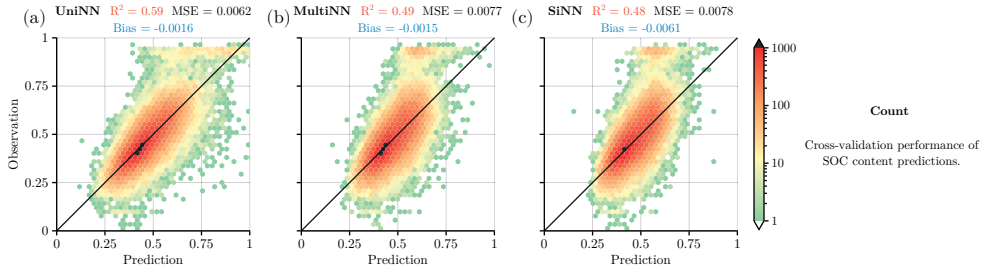
5.6.2 Porosity and compaction

Once oBD and mBD are obtained from SiNN, soil porosity can be derived. Following Robinson et al. (2022), particle densities of the organic and mineral soil constituents (oPD and mPD) are assumed to be constant, with oPD = 1.4 $\mathrm{g\,cm^{-3}}$ and mPD = 2.7 $\mathrm{g\,cm^{-3}}$. Soil porosity is then calculated as:

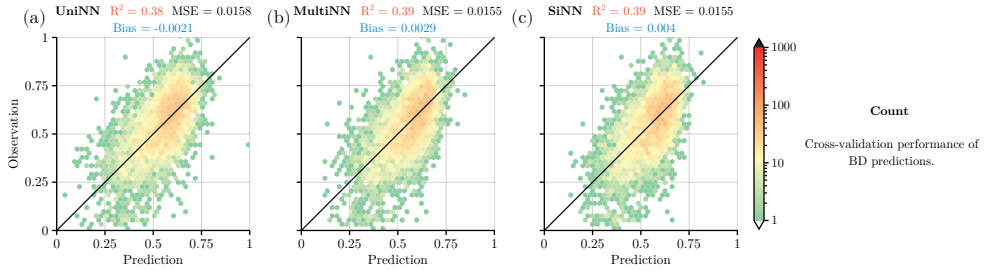
$$\varphi = 1 - \left[\left(\frac{SOM}{oPD} + \frac{1 - SOM}{mPD} \right) \div \left(\frac{SOM}{oBD} + \frac{1 - SOM}{mBD} \right) \right], \tag{5.3}$$

Figure 5.11$(a) - (g)$ shows the derived soil porosity against SOC content across land-cover classes. Porosity is lowest in *Cropland* and *Bareland*, and highest in *Wetland* and *Woodland*. Consistent with estimates reported by Robinson et al. (2022) using soil data from Great Britain and Wales, porosity generally increases with SOC content. This increase, however, levels off at SOC contents of approximately 200 $\mathrm{g\,kg^{-1}}$.

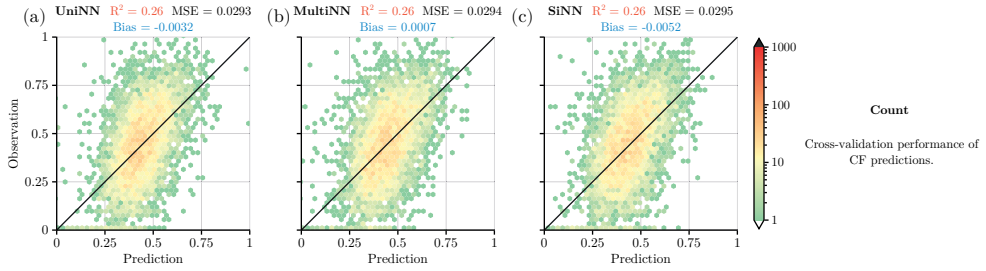
The inverse of mineral soil porosity, expressed as the ratio of mineral BD to mineral particle density (mBD/mPD), could be interpreted as a proxy indicator related to soil compaction. This mBD-based indicator is a modelling construct intended to reduce the influence of SOC-driven density effects. Figure 5.11 (e) compares the distribution of this compaction-related indicator across land-cover classes. The highest mean values are observed in *Cropland*, whereas the lowest mean values occur in *Shrubland*. The widest distributions are found in *Woodland* and *Wetland*.



a SOC content prediction accuracy



b BD prediction accuracy



c CF prediction accuracy

Figure 5.9: Accuracy of cross-validated predictions for the three models, UniNN, MultiNN, and SiNN, across different soil properties. All results are shown in the transformed target space used during model training.

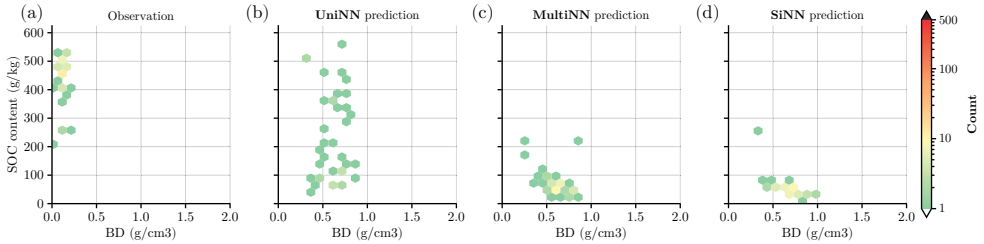


Figure 5.10: Joint distribution of SOC content ($g\,kg^{-1}$) and BD ($g\,cm^{-3}$) for observations and model predictions, restricted to samples with SOC content exceeding 200 $g\,kg^{-1}$ and BD below 0.2 $g\,cm^{-3}$.

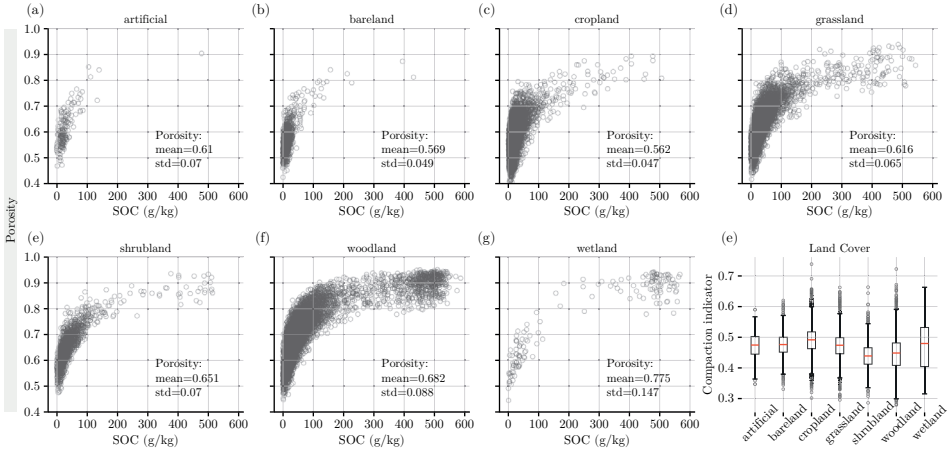


Figure 5.11: $(a) - (g)$. Porosity derived from predicted oBD and mBD, plotted against SOC content for each land cover class. (e) Compaction indicator calculated for each land cover.

Chapter 6

Synthesis

6.1 Revisiting the research objectives

The main aim of this thesis is to explore the operational feasibility, plausibility, and potential methodological improvements of high-resolution, dynamic Digital Soil Mapping (DSM) of Soil Organic Carbon (SOC) across continental Europe. Rather than merely presenting a high-resolution pan-European SOC mapping pipeline, this thesis critically evaluates what such a pipeline can, and cannot, deliver for the analysis of SOC dynamics.

In this final synthesis chapter, the main findings of each chapter are brought together and discussed in relation to the four research objectives outlined in Chapter 1:

1. Transforming EO time series into covariates for large-scale dynamic SOC modelling across Europe.
2. Modelling SOC using EO data and ML–based DSM across space, depth, and time at the European scale.
3. Evaluating the reliability of SOC change detection derived from DSM-based predictions, explicitly accounting for model uncertainty.
4. Integrating soil-scientific knowledge into Machine Learning (ML) models to support SOC dynamics predictions.

The methodological and practical implications of the work are then reflected upon, and the contributions of the thesis are considered in a broader scientific, societal, and personal context. The chapter concludes with an outlook on future research directions for large-scale, spatio-temporal SOC modelling and monitoring.

6.2 Main findings

6.2.1 Transforming EO time series into covariates for large-scale dynamic SOC modelling across Europe

Chapter 2 presents an operational framework for transforming long-term, high–temporal-resolution Earth Observation (EO) time series into spatially complete, temporally coherent, and analysis-ready covariates for large-scale, pan-European SOC modelling. This framework proceeds stepwise by (i) filtering low-quality observations, (ii) reconstructing residual gaps, (iii) deriving thematically relevant spectral indices, (iv) applying feature engineering to obtain multi-temporal-resolution covariates, and (v) conducting systematic quality control.

Because of its long temporal record, high spatial resolution, and open accessibility, Landsat imagery was selected as the backbone of the EO covariates data cube. Although Landsat ARD products already provide extensive preprocessing (Potapov et al., 2020), additional

steps were required to ensure spatial completeness, temporal consistency, and suitability for continental-scale modelling. To achieve spatial completeness, low-quality observations were removed and remaining gaps were addressed through temporal aggregation and reconstruction. Specifically, 16-day Landsat ARD V2 observations were aggregated into bimonthly composites, and residual gaps—most pronounced during winter periods and in high-latitude or high-elevation regions—were reconstructed using the Time Series Iteration-free Reconstruction Framework (TSIRF) algorithm, which exploits temporal neighbourhood information within each pixel time series.

Validation using artificially introduced gaps indicated that the reconstructed bimonthly time series achieve coefficients of determination exceeding 0.6 across the spectral bands. Despite this decent overall accuracy, reconstruction artifacts were found in winter at high latitudes and elevations, because of the very few valid observations in the time series owing to persistent snow and cloud cover. These artifacts reflect inherent limitations of optical EO observations rather than deficiencies of the reconstruction algorithm. Overall, the procedure proves effective in generating spatially complete and temporally coherent EO predictors.

Building on the reconstructed reflectance bands, a suite of spectral indices was derived to represent land-surface processes related to soil, vegetation, crops, and water. Plausibility checks, including visual inspection and comparison with available ground-based observations, indicated that most derived indices are physically consistent and interpretable. Model experiments further showed that expanding the predictor space by incorporating multiple covariate groups generally improves predictive performance. Reflectance bands and well-established indices such as NDVI and NDWI performed well in both plausibility assessments and SOC prediction. In contrast, more complex or highly engineered indices require careful validation and domain-informed interpretation. In this study, indices such as Bare Soil Fraction (BSF) contributed meaningful information to SOC prediction, whereas crop-related indices, e.g. Number of Seasons (NOS) and Crop Duration Ratio (CDR), tended to introduce noise, underscoring the importance of feature selection when working with comprehensive EO-derived predictor sets.

Covariates of multiple temporal resolutions were then developed to capture dynamics at different temporal aggregation scales. Bimonthly reflectance bands and indices were aggregated to annual and multi-year summaries using statistical descriptors such as means or quantiles. Regression experiments showed that models relying exclusively on long-term EO predictors outperformed all other feature combinations for SOC estimation. In contrast, land-cover classification benefited more from predictors with finer temporal resolution. This divergence reflected fundamental differences in process dynamics: SOC responds slowly to environmental and management drivers, typically over multi-year to decadal timescales. Given the limited temporal span of repeated SOC observations in the Land Use/Cover Area Frame Statistical Survey (LUCAS) dataset, it remains difficult to disentangle the

relative contributions of EO data length, high-frequency signal quality, and measurement uncertainty to the limited added value of high-temporal-resolution predictors.

The resulting EO data cube integrates fine spatial resolution, extensive temporal coverage, broad thematic representation, and an analysis-ready structure, thereby providing a robust data foundation for dynamic SOC mapping. At the same time, the consistent superiority of long-term aggregated predictors over higher-temporal-resolution covariates revealed an unresolved tension between EO temporal resolution and the temporal sensitivity of SOC processes, motivating further investigation.

6.2.2 SOC modelling using EO data and ML–based DSM across space, depth, and time at the European scale

In Chapter 3 a pan-European 3D+T SOC density map was created using a reproducible and updatable automated soil mapping framework. The reliability of the resulting maps was systematically examined through qualitative visual assessment, quantitative accuracy evaluation, explicit quantification of the prediction uncertainty, and assessment of the extrapolation risk.

At the continental scale, visual inspection indicated that the predicted SOC density patterns are physically plausible, with coherent spatial variation across major land-cover types. Overall model accuracy assessed through cross-validation and independent test data is satisfactory. However, model performance varied across land-cover classes. In particular, accuracy declined for land covers with sparse SOC density observations, such as shrubland, wetlands, and artificial surfaces. In these classes, uncertainty estimates became less reliable due to limited data support. Although high SOC density regions, such as peatlands and forested areas, were generally identified in spatial terms, the model tended to underestimate high SOC density values. This behaviour was accompanied by increasing prediction uncertainty at higher SOC levels. Whether this reflects the smoothing effect typical of ensemble learning algorithms, limitations in the feature space, SOC data quality, or a combination of these factors remains unresolved.

In addition to accuracy and uncertainty, spatial reliability was further assessed through explicit quantification of extrapolation risk based on feature space coverage. At the continental scale, the extrapolation risk map highlighted regions characterized by environmental conditions that are weakly represented in the training data. In particular, high-altitude regions exhibited elevated extrapolation risk, including parts of Turkey, Norway, Iceland, and Switzerland. In contrast, some regions with limited SOC observations, such as Ukraine, displayed relatively low extrapolation risk. This reflects similarities in feature space to better-represented regions, such as Poland or Germany, indicating that sparse sampling alone does not necessarily imply extrapolation if environmental conditions are well covered by the training data.

The predictor set used in this study lacks explicit information describing depth-specific below-ground processes, due to the scarcity of consistent depth-resolved datasets at continental scale. Therefore, the depth below the surface was included as an explicit covariate, and emerges as the most influential predictor of the SOC profiles over depth. Despite the substantially smaller number of SOC density observations below 20 cm compared to topsoil, model accuracy did not markedly decrease for subsoil layers between 20 and 50 cm. In contrast, performance deteriorated for deeper layers below 50 cm, where prediction uncertainty increases substantially and overall reliability is reduced.

Regarding the temporal dimension, the spatio-temporal mapping framework provides temporally explicit SOC predictions; however, these predictions do not necessarily reflect true SOC density dynamics. As shown in Chapter 3, temporal uncertainty is generally large, which limits the detectability of robust temporal trends. A Shapley value analysis indicated that static predictors dominated both in number and average importance relative to dynamic covariates, suggesting that the model primarily captures spatial variability rather than temporal change. Whether this dominance arises from limitations in EO-derived temporal covariates, the quality and temporal coverage of SOC measurements, or inherently weak SOC change signals over the 2000–2022 period could not be conclusively determined. Despite these limitations, the availability of temporally explicit predictions together with quantified uncertainty enables the cautious exploration of temporal trends, provided that such analyses are conducted in conjunction with uncertainty information, local context, and soil-specific knowledge.

The proposed framework enables the production of spatially and temporally explicit SOC density maps at the European scale with overall satisfactory quality. Given the available SOC observations, the resulting maps are most reliable for topsoil applications in cropland, grassland, and woodland systems. However, due to the limited temporal length of SOC measurements, the predominance of spatial patterns over temporal signals, and substantial predictive uncertainty, drawing robust conclusions about SOC density change over time remains challenging. The accompanying uncertainty and extrapolation risk layers are therefore essential for interpretation, as they allow users to identify areas where predictions are less reliable or where estimates rely on extrapolation beyond the feature space represented in the training data. Map use should be guided by local knowledge and application-specific requirements. In summary, the maps provide a consistent pan-European spatio-temporal SOC resource over two decades and a foundation for future methodological improvement.

6.2.3 Evaluating the reliability of SOC change detection derived from DSM-based predictions, explicitly accounting for model uncertainty.

Chapter 4 introduces a Signal-to-Noise Ratio (SNR) framework to explicitly assess the detectability of SOC change from DSM-based predictions. The framework leverages model-

derived prediction uncertainty, now routinely reported but rarely incorporated into SOC change analysis, to evaluate whether inferred changes exceed the model's own uncertainty, particularly in the absence of repeated SOC measurements for direct validation.

Using the LUCAS survey and national legacy datasets, the SNR framework was applied in an exploratory analysis of SOC change detectability. With three repeated SOC measurements available in LUCAS, two types of change estimates were considered: paired changes derived by simple differencing between two observations within the time series, and temporal trends fitted using all three observations. Across paired change estimates separated by three, six, and nice years, SNR values were consistently low and showed no clear temporal pattern. When comparing paired nine-year changes with nine-year trend estimates, trend-based changes systematically exhibited higher SNR values. This difference may reflect the additional information provided by the third observation, the noise-dampening effect of trend fitting, or artefacts arising from fitting trends to very short time series. Disentangling these effects requires longer SOC time series and cannot be resolved within the current data constraints.

Two modelling strategies for SOC change estimation were further examined. The first, a *state-first* approach, predicts SOC states at specific locations and times and subsequently derives change through differencing or trend fitting. The second, a *change-first* approach, directly models SOC change computed from repeated measurements, requiring sufficient temporal replication. For paired change prediction, the *change-first* approach predominantly produced near-zero change estimates with high baseline uncertainty, resulting in very low SNR values and no clear spatial or thematic patterns, except for marginally higher SNR in woodland systems. In contrast, the *state-first* approach produced coherent state predictions, but the SNR of derived paired changes remained low and similarly lacked clear structure. Trend-based change estimates yielded higher SNR for both approaches; however, SNR values generally remained below one and showed no consistent spatial patterns, except for systematically higher SNR observed in woodland land-cover areas.

Overall, the SNR analysis indicates that, given a nine-year observation window and only three repeated SOC measurements, DSM-based SOC change predictions at the pixel level are dominated by uncertainty, irrespective of modelling strategy or change metric. Consistently low SNR values (<1) suggest that inferred changes are not reliably distinguishable from noise under these conditions. Although trend-based estimates show modestly higher SNR than paired changes, additional repeated observations are required to robustly assess whether this reflects genuine signal detection or methodological artefacts.

An illustrative aggregation analysis further demonstrates that spatial averaging substantially improves SNR by reducing random noise, thereby supporting SOC change assessments at broader spatial scales. However, the choice of aggregation scale and spatial unit requires careful consideration, as incorporating local knowledge into aggregation design is essential for both the interpretability and policy relevance of SOC change assessments.

Finally, the SNR should not be viewed as a replacement for predictive accuracy metrics. In this study, only weak correlations were observed between SNR and relative absolute error. Whereas accuracy metrics evaluate model performance against observations, SNR functions as an internal diagnostic indicating whether predicted changes exceed the model's own uncertainty, thereby offering an additional and conceptually distinct perspective on the reliability of DSM-based SOC change estimates.

6.2.4 Integrating soil-scientific knowledge into ML models to support SOC dynamics predictions.

Chapter 5 evaluated a soil-science-informed neural network for SOC density modelling that explicitly embeds physically constrained SOC–BD relationships, and by comparing its performance with univariate and multivariate neural network architectures under sparse and heterogeneous data conditions.

Across all three models, overall prediction accuracy for SOC density is comparable, with only modest improvements observed for the soil-science-informed model. However, when examining the joint distributions of predicted SOC content and bulk density, the soil-science-informed model consistently preserves physically plausible SOC–BD relationships, whereas the univariate model produces implausible SOC–BD combinations.

Temporal analysis further demonstrates the added value of soil-science-informed modelling for SOC density dynamics. In terms of plausibility, the SiNN yields smoother SOC density time series with substantially fewer extreme temporal fluctuations. In terms of transferability, it exhibits greater robustness to temporal domain shifts, as indicated by smaller performance degradation when models trained on one survey year are applied to other years. Together, these results show that embedding soil-scientific constraints suppresses unrealistic temporal variability and enhances temporal robustness, even when explicit SOC change signals remain weak.

Beyond the prediction of target variables, the soil-science-informed modelling framework provides additional value by yielding physically interpretable latent parameters. By explicitly decomposing Bulk Density (BD) into oBD and mBD components, the model produces latent estimates whose inferred ranges and spatial patterns are broadly consistent with values reported in the literature and exhibit systematic associations with land cover and soil texture. These latent parameters enable SOC density predictions to be linked to soil physical characteristics such as porosity and compaction through established analytical relationships, thereby enhancing model interpretability and transparency while also providing insights into soil physical processes and practical relevance.

Taken together, these results demonstrate that soil-science-informed modelling frameworks improve SOC dynamics prediction primarily by enhancing physical plausibility, temporal stability, and interpretability, rather than by maximising pointwise accuracy alone. This

finding highlights the limitations of evaluating SOC models solely using univariate accuracy metrics and shows that embedding soil-scientific structure within ML models provides meaningful added value for dynamic SOC mapping, particularly when key target variables are sparsely observed.

6.3 Reflection and outlook

6.3.1 The role of EO in DSM

6.3.1.1 An enabler for large-scale SOC mapping

The use of EO data in soil survey and SOC prediction has a long history (Minasny et al., 2013; Mulder et al., 2011). The persistent absence of pan-European 3D+T SOC mapping is therefore not primarily a matter of scientific infeasibility, but rather of data preparation, harmonization, and operational complexity. Chapter 2 addressed this gap by constructing a consistent EO covariate data cube based on long-term Landsat time series, explicitly designed for large-scale SOC modelling.

Advances in EO technologies and products have substantially expanded the possibilities for large-scale modelling and mapping of SOC (Ladoni et al., 2010). The development of a broader range of spectral indices enables more detailed representations of land-surface processes through diverse combinations of satellite bands (Montero et al., 2023). Concurrent improvements in spatial and temporal resolution facilitate the mapping of finer spatial variability and provide opportunities to explore temporal signals. In addition, the increasing length of satellite archives supports long-term monitoring and retrospective analyses, thereby strengthening the role of EO data in SOC assessment and natural resource management.

6.3.1.2 Inherent limitations

However, fully realising the potential of EO-based SOC modelling requires careful consideration of the assumptions, uncertainties, and constraints. EO products are subject to intrinsic quality issues (Croft et al., 2012; Stevens et al., 2008). Even within a single satellite programme such as Landsat, the archive comprises multiple sensors with differing geometric accuracy, radiometric performance, and spectral configurations (Wulder et al., 2019). Although this thesis uses the ARD products, which are consistently pre-processed to provide radiometrically normalised reflectance and improved geometric consistency, residual errors and inconsistencies remain.

Producing gapfree ARD EO products for SOC mapping over continental extents and multi-decadal periods also requires extensive preprocessing, which inevitably introduces additional uncertainty that propagates into downstream SOC predictions. The Landsat-based EO data cube developed in Chapter 2 illustrates the trade-offs between data quality,

temporal resolution, and spatial completeness. Cloud and snow cover substantially degrade optical EO observations, leading to extensive gaps in the original time series (Potapov et al., 2020). These gaps are commonly addressed through temporal aggregation and gap reconstruction using information from neighbouring observations. As a result, higher temporal resolution generally requires more intensive gap-filling, creating an inherent tension between temporal detail and spatial completeness.

Although the reconstructed bimonthly time series presented in Chapter 2 exhibit high accuracy under controlled validation with artificially introduced gaps, uncertainty remains unavoidable. In particular, persistent cloud and snow cover at high latitudes and elevations lead to fewer valid observations, increasing the likelihood of artefacts in reconstructed time series (Consoli et al., 2024; Potapov et al., 2020). While such artefacts do not compromise large-scale analyses, they impose clear constraints on the interpretation of local temporal signals and must be explicitly considered when analysing EO-derived SOC dynamics.

Scale mismatch constitutes an additional challenge. In the spatio-temporal modelling framework adopted in this thesis, point-based SOC measurements are linked to EO predictors at pixel support. While SOC samples are typically collected at point scale or as composites of nearby subsamples, this support remains substantially smaller than the 30 m EO pixel used in this study. Consequently, sub-pixel variability in SOC is not resolved, leading to mixed-pixel effects and increased prediction uncertainty.

6.3.1.3 Feature richness versus soil-scientific relevance

Apart from the Landsat-based EO data cube developed in this thesis, in Chapter 3, additional EO-based predictors were incorporated, including direct signals from satellite (backscatter of synthetic aperture radar Shimada and Ohtaki, 2010; Wagner et al., 2021), variables directly derived from satellite observations (e.g. land surface temperature Wan, 2006, atmospheric water vapour Lyapustin and Wang, 2018), as well as products informed by model simulations and ground observations but supported by EO data, such as climatologies Karger et al., 2017; Karger et al., 2021, lithology Isik et al., 2024, and plant functional types (Harper et al., 2023). These predictors originate from diverse research domains, including vegetation science, crop science, climatology, and remote sensing, and collectively demonstrate their utility for SOC prediction. This cross-disciplinary integration highlights the value of leveraging EO products developed for different purposes to inform SOC modelling. At the same time, careful attention is required to avoid data leakage, whereby variables derived from related processes are used both as predictors and as intermediates in subsequent modelling steps.

The use of a broad and thematically diverse set of EO-derived predictors generally improves SOC model performance, as demonstrated in Chapter 2. Incorporating multiple covariate groups can enhance predictive accuracy and reduce uncertainty (Zeraatpisheh et al., 2022). However, expanding the predictor space also increases the risks of multicollinearity,

overfitting, and computational burden (Wadoux et al., 2020a). With the rapidly growing availability of high-resolution EO products, careful feature selection becomes increasingly critical to maintain parsimony and relevance in DSM. Such selection should be guided not only by statistical performance, but also by an understanding of predictor characteristics, modelling objectives, and task-specific constraints (Chen et al., 2022b; Nazari et al., 2020; Shi et al., 2018). In this context, the soil-scientific meaning of predictors should be considered alongside parsimony and redundancy principle during feature engineering (Richer-de-Forges et al., 2023); otherwise, soil may be treated merely as a material with poorly informed labels. This approach can result in information that lacks practical relevance and does not contribute to a deeper understanding of soil science or soil (Minasny et al., 2024).

6.3.1.4 Integrating other EO resources

One relatively under-exploited class of EO data for SOC modelling is radar remote sensing, which relies on active microwave signals to observe land-surface properties. Compared with optical EO data, radar observations are largely independent of illumination and weather conditions, are sensitive to surface roughness and soil moisture, and can partially penetrate vegetation canopies as well as soil surface (Baghdadi et al., 2008; Richer-de-Forges et al., 2023). These characteristics make radar data particularly valuable for mapping soil surface properties, wetlands, and SOC-related processes (Amani et al., 2021; Bartsch et al., 2016; Luo et al., 2024). Despite this potential, the use of radar products in DSM and SOC modelling remains limited. Although several studies, including this thesis, have incorporated radar-derived covariates and demonstrated their relevance for SOC prediction, such applications are still relatively sparse and often confined to local or regional scales (Chen et al., 2022b). Moreover, radar observations are predominantly exploited in the form of backscatter intensity alone, whereas other physically meaningful signal components-such as phase, interferometric coherence, polarimetric scattering parameters—are rarely utilised. The limited uptake of radar data in DSM can be attributed in part to the technical complexity of radar data processing, which has constrained the availability of analysis-ready and user-friendly radar products for soil applications (Flores-Anderson et al., 2019). As EO infrastructures and preprocessing pipelines continue to mature, further development in this direction holds considerable promise. In particular, radar EO may help improve the representation of high-SOC environments associated with soil moisture and hydrological conditions, which remain challenging to capture accurately using optical EO data alone (Feeney et al., 2022).

In this study, EO-based covariates act as proxies for environmental processes influencing SOC, rather than as direct measurements of soil properties. Some EO approaches aim to capture soil information more directly, for example through bare-soil reflectance composites that select the least vegetated observations from long time series (Demattê et al., 2020; Rogge et al., 2018). Such products can provide valuable information on soil surface

characteristics and serve as covariates for DSM. However, their spatial completeness is often limited by permanent vegetation cover, and their generation relies on masking procedures that may introduce bias by excluding certain soil conditions (Broeg et al., 2026).

While this thesis focuses on large-scale, space-borne EO products, EO more broadly includes information acquired from airborne, drone-based, and proximal sensing platforms. High-resolution hyperspectral imagery, for instance, is sometimes treated as a surrogate for in situ measurements in field-scale soil mapping (Lagacherie and Gomez, 2018). In addition to radar sensing, gamma-ray spectrometry provides an integrated response over a greater soil depth. However, such data are mainly collected by low-altitude airborne sensors. Similarly, proximal sensing techniques such as electromagnetic induction, ground-penetrating radar, and X-ray-based methods are widely used at local scales but remain largely absent from continental-scale SOC mapping efforts (Chen et al., 2022a). These complementary EO modalities highlight both the current limitations of large-scale EO-based SOC modelling and the potential directions for future integration across scales.

6.3.2 Constraints owing to SOC sample data

6.3.2.1 Sampling design, support and density

The full DSM pipeline ideally starts with the design of a soil survey scheme, in which sampling design plays a crucial role (Brus, 2019; Gruijter et al., 2006). Although many DSM studies—including this thesis—rely predominantly on legacy soil data rather than newly designed sampling campaigns, sampling design remains critically important for model evaluation and performance interpretation of resulting maps (Piikki and Söderström, 2019). In this thesis, sampling design has not been explicitly addressed in earlier chapters. This omission partly reflects the reliance on multiple legacy datasets.

The primary data foundation of this thesis is the LUCAS soil survey, which adopts a multi-stage stratified random sampling design. Sampling locations are selected from a regular 2 km × 2 km grid covering the European Union, stratified by country and land-cover class, resulting in approximately 20,000 georeferenced soil samples. However, due to the availability of forest soil data from the BIOSOIL survey (Hiederer, DURRANT, et al., 2010), one-third of the original forest sampling points were reassigned to arable land and grassland. As noted by Tóth et al. (2013), there is a bias towards cropland, which ultimately accounts for approximately 43% of all samples, despite representing only about 34% of the EU-24 land area. This may partly explain the higher model performance in croplands and the lower performance in woodlands in our study.

An additional but often overlooked aspect concerns sampling support. LUCAS employs composite sampling, combining one central subsample with four additional subsamples collected within a 2 m radius. Consequently, the reported "point" SOC measurements

in this study represent actually an integrated support over a small area rather than a true point observation. When such measurements are linked to EO-based covariates and resulted maps, a clear spatial support mismatch arises. As discussed in earlier chapters, this mismatch contributes to increased prediction uncertainty and complicates the interpretation of validation metrics. Explicit reporting and consideration of sampling support are therefore essential for a meaningful assessment of model performance (Piikki et al., 2021).

Sampling density represents a further structural limitation, since this study, like many other large-scale SOC mapping efforts, relies on a relatively low density of soil observations (Chen et al., 2022b). With approximately 31,000 sites distributed over the mapped continental Europe (approximately 7 million km^2) in Chapter 3, the effective sampling density in this study is below 0.005 sample sites km^{-2}. Previous work on particle size distribution has shown that prediction accuracy depends on sampling density, with performance gains saturating only at densities approaching one sample site per km^2 (Loiseau et al., 2021). Although comparable analyses are lacking for SOC, it is reasonable to expect that insufficient sampling density similarly constrains SOC prediction accuracy. From this perspective, the Soil Health Data Cube produced in Chapter 3 can reasonably be characterised as "doing too much with too little" (Minasny et al., 2024), in the sense that it attempts to generate spatially detailed SOC predictions over large areas despite limited and unevenly distributed ground observations. The resulting maps should therefore be interpreted as an imperfect baseline that can be progressively improved as additional data becomes available.

Sampling design and optimisation have been a central focus in geostatistical DSM (Heuvelink and Webster, 2022). In contrast, many recent ML-based DSM studies either underreport or overlook sampling design (Chen et al., 2022b; Wadoux et al., 2020b). While EO data and ML methods enable large-scale SOC mapping at relatively low cost, they cannot replace field soil measurements, which remain the foundation of DSM (Richer-de-Forges et al., 2023). An overreliance on modelling at the expense of new observations risks weakening this foundation (Minasny et al., 2024). In this sense, remotely sensed data may be most effectively used not solely for prediction, but also to guide and optimise future soil sampling efforts (Ladoni et al., 2010).

6.3.2.2 Analytical uncertainty in SOC measurements

Beyond sampling limitations, inconsistencies in SOC measurements further constrain large-scale SOC dynamics modelling. Variation in analytical measurements arises from differences in laboratory conditions, protocols, operators, instruments, and random error (Leeuwen et al., 2024; Pleijsier, 1989). Grover et al. (2025) proposed using inter-laboratory analytical uncertainty, derived from sample exchanges and blind sample analyses across laboratories, as a proxy for acceptable prediction uncertainty bounds. Such measurement

uncertainty occurs not only between laboratories, but also within individual laboratories. Similarly, Breure et al. (2026) quantified intra-laboratory variation and used it as a benchmark to evaluate visible–near-infrared spectroscopy applied to the LUCAS survey. While these metrics provide valuable reference points for interpreting model uncertainty, their applicability depends on laboratories conducting dedicated uncertainty assessments and making the results publicly available. For studies relying on legacy soil datasets, such detailed information on analytical uncertainty is often unavailable.

When merging legacy soil datasets measured using different analytical methods is required, pedotransfer functions are often used to mitigate methodological inconsistencies. However, their effectiveness depends on the availability of calibration datasets in which multiple methods have been applied or on the existence of robust, well-established transfer relationships (Ciampalini et al., 2013). Moreover, pedotransfer functions themselves are affected by measurement error (Leeuwen et al., 2024), and harmonisation procedures cannot fully eliminate this source of uncertainty (Goidts et al., 2009; Hu et al., 2021; Libohova et al., 2019; Louis et al., 2014).

Importantly, these error sources originate even earlier than laboratory analysis, beginning with sampling strategies and their implementation. For example, although LUCAS applies a harmonised sampling protocol across Member States, differences in execution and field recording are unavoidable (Hiederer, 2020).

While such uncertainties may be of limited consequence for mapping spatial SOC patterns at a single point in time, they become critical when interpreting weak temporal signals associated with SOC change.

6.3.2.3 Limited representation in time

A fundamental constraint in modelling and mapping SOC dynamics is the inherently slow rate at which SOC changes over time (Gubler et al., 2019), even under agricultural management or land-use change (Ciais et al., 2010; Poeplau and Don, 2013). Detecting SOC change generally becomes more feasible over longer time spans, with the required duration typically exceeding a decade (Smith, 2004). In addition to sufficient temporal extent, sampling frequency needs to be enough to ensure that underlying change signals are not obscured by short-term fluctuations in SOC (Harbo et al., 2025). Weak temporal signals are further obscured by strong spatial heterogeneity, which can exceed temporal variability by an order of magnitude (Broeg et al., 2024b; Goidts et al., 2009; Heuvelink et al., 2020).

These intrinsic characteristics of SOC dynamics are compounded by the limited availability of repeated soil observations. Many soil surveys often require several years to complete regional coverage (e.g. Parcelas COS in Spain Serrano et al., 2022, BZE-LW in Germany Poeplau et al., 2020b), and repeated measurements are rare, often limited to a small region,

or the repeated measurements are yet to be conducted. A common workaround is to combine soil observations from multiple surveys conducted in different periods, often with differing sampling designs, laboratory methods, and implementation protocols (Chen et al., 2022b). While this strategy increases data availability, it introduces an additional source of uncertainty owing to survey heterogeneity. Sampling frequency further influences SOC change assessment: estimating change from only two time steps is particularly sensitive to noise and may lead to misleading conclusions. In contrast, longer time series help stabilise trend estimates (Harbo et al., 2025). At the time of writing, LUCAS was the only harmonised SOC monitoring programme with repeated coverage across the EU, spanning approximately nine years and offering at most three observations per location, which still spans a relatively short period for robust SOC change detection. The forthcoming LUCAS 2022 release is expected to improve this situation by extending both the temporal coverage and the range of measured properties.

6.3.3 Representing soil systems in models

6.3.3.1 Incorporating spatial context

As noted by Heuvelink and Webster (2022), most conventional ML algorithms are inherently non-spatial. Although tabular training data are constructed by linking point-based soil observations to covariate values extracted at their geographic locations, the learning algorithm itself has no spatial awareness. One strategy to approach this limitation is to encode spatial context into covariates, such as geographic covariates (Hengl et al., 2018), and feature-engineered multi-scale terrain attributes (Behrens et al., 2010). Also Random Forest spatial interpolation has been proposed, which incorporates observations from nearby locations and their distances to the prediction point (Sekulić et al., 2020). However, with these strategies, spatial relationships are only implicitly captured through input features, rather than being explicitly represented within the model formulation.

More recently, deep learning approaches have been explored to incorporate spatial context more directly, by processing neighbouring patches as structured inputs to the learning algorithm (Kamilaris and Prenafeta-Boldú, 2018). In particular, convolutional neural networks have been applied to DSM to capture local spatial patterns and landscape structure (Padarian et al., 2019; Wadoux, 2019a; Yang et al., 2021; Zhang et al., 2022), with recent extensions to transformer-based architectures that aim to model longer-range spatial dependencies (Tresson et al., 2024). By explicitly account for spatial context, such methods offer the potential to capture landscape structure, spatial gradients, and local heterogeneity that are difficult to represent using tabular predictors alone. At the same time, deep learning models are typically data-intensive, computationally demanding, and more challenging to interpret, which can limit their practical applicability in large-scale DSM.

Despite these advances, deep learning approaches do not explicitly model spatial autocorrelation in the geostatistical sense (Heuvelink and Webster, 2022). As noted by Wadoux et al. (2020b), applying a non-spatial model in DSM is not problematic in itself; however, modelling choices should not ignore or violate the spatial structure present in soil data. Examination of spatial autocorrelation in model residuals can help identify inadequately represented processes in the model (Wadoux et al., 2020b). In some cases, environmental covariates have limited explanatory power for the spatial variation of the target soil property because they are noisy, incomplete, or only weakly linked to the underlying processes. Under such conditions, the spatial autocorrelation of the target variable itself may become a dominant source of information. Moreover, spatial autocorrelation can provide meaningful insights into soil variability and underlying processes when interpreted within an appropriate domain context. From this perspective, how spatial dependence can be appropriately accounted for within ML-based DSM frameworks remains an important and valuable direction for future methodological development.

6.3.3.2 Depth as an underrepresented dimension

As a three-dimensional body, soil depth constitutes a crucial yet persistently underrepresented dimension in DSM. A review of large-scale DSM studies published between 2003 and 2021 by Chen et al. (2022b) shows that model performance generally declines with increasing depth. This pattern can be attributed both to limited availability of depth-resolved soil observations and to the difficulty of constructing meaningful depth-specific covariates.

The primary data source used in this thesis, the LUCAS soil survey, is predominantly focused on topsoil (0–20 cm). Only from 2018 onwards were deeper layers up to 30 cm included (Orgiazzi et al., 2018). Consequently, among the four chapters of this thesis, only Chapter 3 attempts to model SOC below 20 cm. Observations for deeper layers were mainly supplemented from national datasets in Germany (Poeplau et al., 2020c) and Spain (Serrano et al., 2022). In Germany, full soil profiles are sampled and recorded at each location, whereas in Spain, a separate dataset provides measurements down to 30 cm. This heterogeneity in depth coverage and sampling design hinders depth-consistent modelling at continental scale.

Depth-related challenges are not limited to soil observations but also extend to the covariates. Unlike vegetation or land-cover dynamics, SOC is a subsurface property, whereas most EO products suitable for large-scale applications are space-borne and primarily sensitive to land-surface conditions. Vegetation cover and built-up areas further obscure soil signals, limiting the ability of EO data to directly capture SOC information. Depth-resolved covariates are therefore difficult to prepare, both because of the substantial effort required and the limited availability of suitable datasets. When such covariates do exist, they are typically region-specific, as illustrated by the peat occurrence maps

developed by Helfenstein et al. (2024b), which provide depth-explicit information and have been shown to reduce prediction uncertainty.

With respect to modelling strategies, the suitability of different approaches depends on the spatial variation of the target soil property and the availability of depth-resolved observations (Chen et al., 2022b). In this thesis, a 3D modelling strategy was adopted in Chapter 3, in which depth was included as an explicit predictor within a Random Forest (RF) framework. While this approach enables joint modelling across depth intervals, it is known to produce artificial "stepped" depth patterns (Ma et al., 2021). Moreover, depth often emerges as the dominant predictor, exhibiting substantially higher feature importance than all other covariates. As noted by Ma et al. (2021), this dominance may cause tree-based models to prioritise depth-based splits, potentially limiting the model's ability to learn robust relationships between soil properties and environmental drivers. Alternative approaches, including multivariate ML models that jointly predict soil properties across multiple depths, have been proposed to address these limitations (Ließ and Sakhaee, 2024; Wadoux, 2019a). However, such approach also requires the support of consistent SOC measurements in depth.

An additional consideration concerns the effective soil depth represented by the predicted maps. In Chapter 3, SOC density is modelled down to the 100–200 cm depth interval across continental Europe. However, in many regions the depth to bedrock is shallower than this, rendering such predictions physically implausible at specific locations. While land management and agricultural practices primarily concern topsoil layers, reliable Monitoring, Reporting, and Verification (MRV) of SOC stocks requires this information. Global maps of depth to bedrock are available (Shangguan et al., 2017), but their accuracy and suitability for application at the European scale remain uncertain and require further evaluation. Explicitly accounting for spatial variability in soil depth therefore represents an important prerequisite for improving the physical realism and interpretability of depth-resolved SOC predictions.

6.3.3.3 Integrating soil knowledge

While ML has undoubtedly transformed DSM of SOC, with an emphasis on improving predictive accuracy, it also exhibits important limitations. As highlighted by Wadoux et al. (2020b), three elements require greater emphasis in DSM: *plausibility*, meaning that predictions should be consistent with established soil-scientific understanding; *interpretability* and *explainability*, which enable insight into model behaviour, validation of covariate–target relationships, and the generation of new hypotheses that may challenge or advance existing soil knowledge. None of these can be achieved without explicitly engaging with soil science.

The explicit quantification and integration of soil knowledge derived from field surveyors and soil scientists has long been pursued, yet remains challenging (Hudson, 1992; Ma et al.,

2019; Walter et al., 2006). In the context of soil science, Minasny et al. (2024) discussed three integration pathways proposed by Karniadakis et al. (2021): observational priors, model structure, and learning guidance. To date, most integration efforts in DSM have focused on the use of observational priors, primarily through the selection and construction of input data, including both covariates and soil observations.

Early foundations of DSM were rooted in soil science, conceptualising soil as a dynamic system linked to soil-forming factors (Jenny, 1941; McBratney et al., 2003). While data-driven covariate selection methods are widely used to refine predictor sets, the initial pool of candidate predictors is constructed based on these soil-forming factors (Hamzehpour et al., 2019; Zhang et al., 2023a). The Homosoil concept proposed by Mallavan et al. (2010) assumes similarity in soil-forming factors between reference areas and target regions. Hengl et al. (2017) introduced pseudo-observations based on expert knowledge to augment training data in poorly represented regions, an approach also adopted in Chapter 3. More recently, Zhang et al. (2024a) employed synthetic data generated from process-based SOC models to improve temporal representativeness.

Through model structure, soil-scientific relationships can be directly embedded. In this context, the multivariate ML approaches discussed earlier and the soil-science-informed neural network evaluated in Chapter 5 represent examples of structure-based integration, where part of the model is constructed by known soil relationships rather than purely statistical optimisation. The learning guidance can be realized through tailored loss functions. Both requires a flexible ML framework that allows customization of model structure and loss funciton.

Process-based models of SOC dynamics, such as RothC, MIMICS, ECOSYS, and Century, explicitly simulate carbon cycling within plant–soil systems and sometimes include crop growth dynamics (Ding et al., 2025). However, these models require substantial expertise to parameterise and depend on detailed input data, making them difficult to apply at large scales. Current efforts to integrate process-based models with ML have therefore primarily focused on the use of synthetic data to inform data-driven models (Liu et al., 2024; Zhang et al., 2024a).

Integration of soil knowledge should also include the evaluation of models. DSM products must also be evaluated in terms of plausibility, not solely statistical accuracy. While most studies focus on quantitative performance metrics, combining statistical validation with pedological interpretation can strengthen both DSM and soil science. For example, soil experts can assess the coherence of soil–landscape patterns in digital maps or apply digital soil products within land suitability frameworks (Holmes et al., 2021), contributing to what has been termed “conscious” DSM (Wadoux et al., 2020b).

Importantly, the flow of integration is bidirectional. While soil knowledge can guide modelling choices, DSM models developed for mapping purposes can also contribute to advancing soil understanding, a process referred to as knowledge discovery from DSM

(Ma et al., 2019; Wadoux et al., 2020b). Such models can be used to explore multivariate interactions among soil-forming factors, reveal spatially variable hierarchies of environmental controls, and quantify patterns of predictor influence at large scales, either through expert interpretation (Bui et al., 2006) or explainable ML techniques (Wadoux and Molnar, 2022). Chapter 5 represents a modest contribution in this direction, by exploring the integration of a simple and physically interpretable SOC–BD relationship into a neural network framework. Further progress is expected as open data availability increases, modelling tools mature, and collaborative efforts continue to bridge soil science and data-driven approaches.

6.3.4 Usability of SOC maps

6.3.4.1 Communication of map quality

The usability of digital soil maps fundamentally depends on their quality, which should be evaluated in relation to their intended use (Rossiter and Poggio, 2025). Nevertheless, evaluation practices in DSM remain dominated by point-based statistical metrics such as RMSE and R^2, typically reported as global summaries. While these metrics are informative for assessing predictive accuracy at observation locations, they are insufficient for evaluating whether maps meaningfully represent soil bodies and spatial patterns across landscapes, which is important for applications such as ecological zoning or land-use planning (Rossiter, 2021). Maps with similar point-based accuracy can exhibit substantially different spatial structures and degrees of spatial coherence. To address this limitation, Rossiter and Poggio (2025) proposed complementing point-based evaluation with pattern-based assessment, for example by analysing aggregated or segmented maps to identify spatially homogeneous units.

Map usability is inherently scale-dependent. Piikki and Söderström (2019) argued that accuracy should be evaluated across multiple spatial supports to reflect different application contexts. Soil maps may be used at field scale for site-specific management, or at municipal, regional, and national scales for decision support, policy development, and environmental modelling. Their study on Swedish arable land demonstrated that soil organic matter maps may exhibit inadequate accuracy for precision agriculture at the farm level. Similarly, a multiscale evaluation of DSM products by Lemercier et al. (2022) showed that maps produced for large spatial extents (e.g. global or continental) often fail to provide sufficiently accurate information at finer, local spatial supports that are typically relevant to end-users. These findings highlight the importance of explicitly linking accuracy assessment to intended use and spatial supports.

While considerable effort has been devoted to quantifying uncertainty in digital soil maps, communicating and using this information remains challenging. Surveys among end-users and decision-makers indicate widespread discomfort with uncertainty information (Padarian and McBratney, 2023; Richer-de-Forges et al., 2019). This discomfort contributes to a

persistent misconception that finer spatial resolution necessarily implies higher accuracy. Moreover, to avoid exposing decisions to contestation, polygon-based soil information is often preferred in legal and administrative contexts. Recent studies have questioned whether conventional DSM representations featured with both high spatial resolution and high uncertainty are appropriate for users (Arrouays et al., 2020a; Courteille et al., 2024; Vaysse et al., 2017). Although spatial aggregation reduces spatial detail, it can also decrease uncertainty and improve interpretability. In this context, Padarian and McBratney (2023) proposed encoding uncertainty through variable pixel sizes, ensuring that all map units meet a predefined uncertainty threshold. Similarly, Courteille et al. (2025) showed that users tend to favour intermediate levels of spatial aggregation that reduce uncertainty. Together, these findings suggest that the demand for ever finer spatial resolution is often driven by shortcomings in how uncertainty is communicated. Future DSM efforts should therefore place greater emphasis on designing map products that balance spatial resolution, aggregation, and uncertainty communication in ways that align with end-user needs and decision contexts (Arrouays et al., 2020a; Heuvelink and Webster, 2022).

6.3.4.2 Involving stakeholders

DSM has been closely linked to practical applications since its inception, particularly in agricultural contexts. Although DSM has since evolved into a research field in its own right, it should not be treated merely as a modelling exercise within data science. DSM products are not only scientific outputs, but also information products intended for use by practitioners, decision-makers, and stakeholders. As argued by Carré et al. (2007), DSM should not be regarded as an end in itself, but rather as a technique for providing data and information within a broader framework for soil assessment that integrates political, social, economic, and environmental considerations. The persistent difficulty faced by end-users and decision-makers in applying digital maps of soil classes and properties to their final objectives highlights the need for mappers and researchers to remain attentive to usability throughout the mapping process.

In this context, increasing model complexity and more sophisticated representations of map quality place growing demands on user competence. User communities thus require training and experience to correctly interpret map outputs, particularly with respect to uncertainty information and its use in risk assessment and policy development (Arrouays et al., 2020a; Courteille et al., 2025). At the same time, further advances in DSM critically depend on the availability of high-quality ground truth measurements, which remain largely insufficient. As noted by Schmidinger et al. (2025), open soil datasets at field and farm scales suitable for high-resolution precision agriculture are almost entirely lacking. Effective progress in DSM therefore requires sustained stakeholder involvement beyond model development alone, encompassing data provision, map design, interpretation, and feedback.

6.4 Conclusions

At the beginning of the twentieth century, Legros (2006) commented on the first world soil map produced by K. D. Glinka: "Its accuracy is very limited and the map must not be considered in detail." Looking back at the SOC maps produced and updated during my PhD project, I am left with a similar impression. Although these maps provide a consistent baseline for further improvement and reveal visually appealing large-scale spatial variation, their accuracy remains limited and they should not be interpreted in detail.

This observation is not intended to diminish the efforts invested in this research, but rather to place them in context and to look forward. By providing an open, analysis-ready EO data cube for Europe (RO1) and developing an operational pipeline for large-scale SOC modelling and mapping (RO2), this thesis contributes to advancing DSM practice. More importantly, it is situated at a moment when DSM is transitioning from demonstrating the ability to map towards addressing questions of plausibility, credibility, and usability of DSM products. Chapter 4 and Chapter 5 of this thesis therefore focus on quantifying the limitations of SOC change detection using ML and EO (RO3), and on exploring the potential of soil-science-informed modelling approaches to improve interpretability and robustness (RO4).

This PhD project is embedded in two European research initiatives, one focusing on soil health assessment and the other on carbon credit MRV. For soil health management, the findings of this study indicate that ever-increasing spatial resolution is not necessarily the primary objective for supporting land managers. Instead, soil information aggregated to predefined spatial units with reduced uncertainty may be more suitable for decision-making. In parallel, MRV applications for carbon markets require reliable information on SOC dynamics. However, robust detection of temporal SOC change remains challenging with current EO-based DSM and ML approaches. These challenges point to key directions for future development, with the aim of ensuring that DSM better serves societal needs and environmental objectives. Achieving this will require not only continued technological development, but also closer and more sustained engagement with stakeholders.

In summary, this thesis underscores that the future of large-scale DSM lies not solely in methodological development, but in a balanced integration of data quality, soil-scientific knowledge, appropriate modelling choices, and meaningful stakeholder involvement. Such an approach is essential for ensuring that DSM evolves from technically impressive products into robust and trustworthy tools for understanding, managing, and monitoring soils under changing environmental and societal conditions.

Acronyms

BD Bulk Density. 6, 10, 11, 13, 73, 74, 75, 102, 154, 155, 156, 157, 159, 160, 162, 163, 164, 166, 167, 168, 169, 170, 171, 173, 181, 192

BSF Bare Soil Fraction. 16, 19, 22, 24, 25, 28, 30, 31, 32, 38, 39, 40, 52, 53, 54, 55, 56, 59, 177

CDR Crop Duration Ratio. 16, 19, 23, 28, 30, 31, 32, 39, 43, 44, 177

CF Coarse Fragment. 10, 73, 155, 156, 157, 159, 166, 169, 170

DSM Digital Soil Mapping. 4, 5, 7, 8, 9, 10, 11, 12, 13, 38, 41, 69, 100, 118, 119, 126, 130, 135, 142, 143, 154, 155, 156, 169, 176, 178, 184, 185, 186, 188, 189, 190, 191, 192, 193, 194

EO Earth Observation. 5, 8, 9, 11, 13, 17, 18, 21, 37, 38, 39, 40, 43, 44, 68, 69, 70, 104, 118, 119, 136, 137, 140, 141, 142, 144, 155, 168, 176, 177, 178, 182, 183, 184, 185, 186, 189, 194

FAPAR Fraction of Absorbed Photosynthetically Active Radiation. 16, 19, 22

GHG Green House Gases. 6

LUCAS Land Use/Cover Area Frame Statistical Survey. vii, 7, 8, 9, 13, 16, 24, 25, 27, 28, 29, 41, 53, 54, 55, 57, 58, 69, 70, 72, 73, 74, 75, 76, 86, 96, 97, 101, 102, 104, 120, 121, 122, 123, 131, 134, 136, 137, 138, 140, 141, 142, 143, 155, 156, 157, 163, 167, 169, 177, 180, 185, 187, 188, 189

minNDTI Minimum Normalized Difference Tillage Index. 16, 26, 30, 31, 32, 33, 39, 43, 52, 60, 61

ML Machine Learning. 5, 9, 10, 11, 13, 14, 69, 70, 72, 73, 83, 118, 119, 140, 142, 143, 144, 154, 155, 156, 176, 182, 186, 188, 189, 190, 191, 192, 194

MRV Monitoring, Reporting, and Verification. 7, 190, 194

NDSI Normalised Difference Snow Index. 16, 19, 22

NDTI Normalised Difference Tillage Index. 16, 19, 22

NDVI Normalised Difference Vegetation Index. 16, 19, 22, 23, 24, 28, 30, 31, 32, 38, 39, 40, 50, 51, 57, 58

NDWI Normalised Difference Water Index. 16, 19, 22, 23, 24, 28, 29, 30, 31, 39, 40, 50, 51, 53, 54, 55

NOS Number of Seasons. 16, 19, 24, 25, 28, 30, 31, 32, 36, 38, 39, 40, 41, 44, 177

QRF Quantile Regression Forests. 9, 68, 72, 81, 118, 119, 126, 127, 128, 129, 143

RF Random Forest. vi, vii, 68, 70, 78, 80, 81, 82, 84, 86, 88, 96, 104, 126, 127, 128, 143, 190

RS Remote Sensing. 5

SAVI Soil Adjusted Vegetation Index. 16, 19, 22

SNR Signal-to-Noise Ratio. 13, 118, 120, 121, 123, 127, 128, 129, 131, 133, 134, 135, 136, 137, 138, 139, 142, 143, 144, 145, 146, 147, 148, 150, 151, 179, 180, 181

SOC Soil Organic Carbon. v, vi, vii, viii, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 20, 27, 28, 33, 34, 35, 36, 37, 40, 41, 42, 44, 49, 62, 63, 69, 73, 84, 94, 98, 101, 102, 118, 119, 120, 121, 122, 123, 124, 126, 127, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 140, 141, 142, 143, 144, 145, 146, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 166, 167, 168, 169, 170, 171, 173, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 194

SOCD Soil Organic Carbon Density. 68, 69, 70, 72, 73, 74, 75, 76, 77, 79, 80, 81, 82, 84, 85, 86, 87, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 106, 107, 110, 112, 113, 115

TSIRF Time Series Iteration-free Reconstruction Framework. 21, 24, 29, 42, 177

References

A. Wentz, E. and M. Shimizu (2018). "Measuring spatial data fitness-for-use through multiple criteria decision making". *Annals of the American Association of Geographers* 108.4, 1150–1167. DOI: 10.1080/24694452.2017.1411246.

Adams, W. (1973). "The effect of organic matter on the bulk and true densities of some uncultivated podzolic soils". *Journal of Soil Science* 24.1, 10–17. DOI: 10.1111/j.1365-2389.1973.tb00737.x.

Adekanmbi, A. A., X. Shu, Y. Zou, and T. Sizmur (2022). "Legacy effect of constant and diurnally oscillating temperatures on soil respiration and microbial community structure". *European Journal of Soil Science* 73.6, e13319. DOI: 10.1111/ejss.13319.

Adhikari, K., A. Hartemink, B. Minasny, R. Kheir, M. Greve, and M. Greve (2014). "Digital mapping of soil organic carbon contents and stocks in Denmark". *PloS ONE* 9, 105519. DOI: 10.1371/journal.pone.0105519.

Alonso, L., B. Ahrens, M. Reichstein, X. Tian, and Q.-F. Bi (2025). *Earthy-Science/EasyHybrid.jl: v0.1.7*. Version v0.1.7. DOI: 10.5281/zenodo.17794983.

Amani, M., V. Poncos, B. Brisco, F. Foroughnia, E. R. DeLancey, and S. Ranjbar (2021). "InSAR coherence analysis for wetlands in Alberta, Canada using time-series Sentinel-1 data". *Remote Sensing* 13.16, 3315. DOI: 10.3390/rs13163315.

Andrews, S. S., D. L. Karlen, and C. A. Cambardella (2004). "The soil management assessment framework: a quantitative soil quality evaluation method". *Soil Science Society of America Journal* 68.6, 1945–1962. DOI: 10.2136/sssaj2004.1945.

Angelini, M. E., G. Heuvelink, and B. Kempen (2017). "Multivariate mapping of soil with structural equation modelling". *European Journal of Soil Science* 68.5, 575–591. DOI: 10.1111/ejss.12446.

Angelopoulou, T., N. Tziolas, A. Balafoutis, G. Zalidis, and D. Bochtis (2019). "Remote sensing techniques for soil organic carbon estimation: A review". *Remote Sensing* 11.6, 676. DOI: 10.3390/rs11060676.

Araza, A., S. de Bruin, M. Herold, S. Quegan, N. Labriere, P. Rodriguez-Veiga, V. Avitabile, M. Santoro, E. T. Mitchard, C. M. Ryan, O. L. Phillips, S. Willcock, H. Verbeeck, J. Carreiras, L. Hein, M.-J. Schelhaas, A. M. Pacheco-Pascagaza, P. da

Conceição Bispo, G. V. Laurin, G. Vieilledent, F. Slik, A. Wijaya, S. L. Lewis, A. Morel, J. Liang, H. Sukhdeo, D. Schepaschenko, J. Cavlovic, H. Gilani, and R. Lucas (2022). "A comprehensive framework for assessing the accuracy and uncertainty of global above-ground biomass maps". *Remote Sensing of Environment* 272, 112917. DOI: 10.1016/j.rse.2022.112917.

Arrouays, D., M. G. Grundy, A. E. Hartemink, J. W. Hempel, G. B. Heuvelink, S. Y. Hong, P. Lagacherie, G. Lelyk, A. B. McBratney, N. J. McKenzie, et al. (2014a). "GlobalSoilMap: Toward a fine-resolution global grid of soil properties". *Advances in agronomy* 125, 93–134. DOI: 10.1016/B978-0-12-800137-0.00003-0.

Arrouays, D., P. Lagacherie, and A. E. Hartemink (2017). *Digital soil mapping across the globe*. DOI: 10.1016/j.geodrs.2017.03.002.

Arrouays, D., A. McBratney, J. Bouma, Z. Libohova, A. C. Richer-de-Forges, C. L. Morgan, P. Roudier, L. Poggio, and V. L. Mulder (2020a). "Impressions of digital soil maps: The good, the not so good, and making them ever better". *Geoderma Regional* 20, e00255. DOI: 10.1016/j.geodrs.2020.e00255.

Arrouays, D., A. B. Mcbratney, B. Minasny, J. W. Hempel, G. Heuvelink, R. MacMillan, A. Hartemink, P. Lagacherie, N. J. Mckenzie, et al. (2014b). "The GlobalSoilMap project specifications". *GlobalSoilMap: Basis of the global spatial soil information system*.

Arrouays, D., L. Poggio, O. A. S. Guerrero, and V. L. Mulder (2020b). "Digital soil mapping and GlobalSoilMap. Main advances and ways forward". *Geoderma Regional* 21, e00265. DOI: 10.1016/j.geodrs.2020.e00265.

Assouline, S. and D. Or (2013). "Conceptual and parametric representation of soil hydraulic properties: A review". *Vadose Zone Journal* 12.4, vzj2013–07. DOI: 10.2136/vzj2013.07.0121.

Baghdadi, N., O. Cerdan, M. Zribi, V. Auzet, F. Darboux, M. El Hajj, and R. B. Kheir (2008). "Operational performance of current synthetic aperture radar sensors in mapping soil surface characteristics in agricultural environments: application to hydrological and erosion modelling". *Hydrological Processes: An International Journal* 22.1, 9–20. DOI: 10.1002/hyp.6609.

Barbet-Massin, M., F. Jiguet, C. H. Albert, and W. Thuiller (2012). "Selecting pseudo-absences for species distribution models: How, where and how many?" *Methods in ecology and evolution* 3.2, 327–338. DOI: 10.1111/j.2041-210X.2011.00172.x.

Bartsch, A., B. Widhalm, P. Kuhry, G. Hugelius, J. Palmtag, and M. B. Siewert (2016). "Can C-band synthetic aperture radar be used to estimate soil organic carbon storage in tundra?" *Biogeosciences* 13.19, 5453–5470. DOI: 10.5194/bg-13-5453-2016.

Batjes, N. (2009). "Harmonized soil profile data for applications at global and continental scales: updates to the WISE database". *Soil Use and Management* 25.2, 124–127. DOI: 10.1111/j.1475-2743.2009.00202.x.

Batjes, N. H., E. Ribeiro, A. van Oostrum, J. Leenaars, T. Hengl, and J. M. de Jesus (2017). "WoSIS: providing standardised soil profile data for the world". *Earth System Science Data* 9.1, 1.

Batjes, N. H., E. Ribeiro, and A. Van Oostrum (2020). "Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019)". *Earth System Science Data* 12.1, 299–320. DOI: 10.5194/essd-12-299-2020.

Bauer-Marschallinger, B., S. Cao, C. Navacchi, V. Freeman, F. Reuß, D. Geudtner, B. Rommen, F. C. Vega, P. Snoeij, E. Attema, C. Reimer, and W. Wagner (2021). "The normalised Sentinel-1 Global Backscatter Model, mapping Earth's land surface with C-band microwaves". *Scientific Data* 8.1. DOI: 10.1038/s41597-021-01059-7.

Bauer-Marschallinger, B., V. Freeman, S. Cao, C. Paulik, S. Schaufler, T. Stachl, S. Modanesi, C. Massari, L. Ciabatta, L. Brocca, et al. (2018). "Toward global soil moisture monitoring with Sentinel-1: Harnessing assets and overcoming obstacles". *IEEE Transactions on Geoscience and Remote Sensing* 57.1, 520–539. DOI: 10.1109/TGRS.2018.2858004.

Baumann, P. (2024). "On the analysis-readiness of spatio-temporal earth data and suggestions for its enhancement". *Environmental Modelling & Software*, 106017. DOI: 10.1016/j.envsoft.2024.106017.

Beeson, P. C., C. S. Daughtry, and S. A. Wallander (2020). "Estimates of conservation tillage practices using landsat archive". *Remote Sensing* 12.16, 2665. DOI: 10.3390/rs12162665.

Behrens, T., K. Schmidt, R. A. MacMillan, and R. A. Viscarra Rossel (2018). "Multi-scale digital soil mapping with deep learning". *Scientific reports* 8.1, 15244. DOI: 10.1038/s41598-018-33516-6.

Behrens, T. and T. Scholten (2006). "Digital soil mapping in Germany—a review". *Journal of Plant Nutrition and Soil Science* 169.3, 434–443. DOI: 10.1002/jpln.200521962.

Behrens, T., A.-X. Zhu, K. Schmidt, and T. Scholten (2010). "Multi-scale digital terrain analysis and feature selection for digital soil mapping". *Geoderma* 155.3-4, 175–185. DOI: 10.1016/j.geoderma.2009.07.010.

Bellamy, P. H., P. J. Loveland, R. I. Bradley, R. M. Lark, and G. J. Kirk (2005). "Carbon losses from all soils across England and Wales 1978–2003". *Nature* 437.7056, 245–248. DOI: 10.1038/nature04038.

Blanco-Canqui, H., C. A. Shapiro, C. S. Wortmann, R. A. Drijber, M. Mamo, T. M. Shaver, and R. B. Ferguson (2013). "Soil organic carbon: The value to soil properties". *Journal of Soil and Water Conservation* 68.5, 129A–134A. DOI: 10.2489/jswc.68.5.129A.

Bolton, D. K., J. M. Gray, E. K. Melaas, M. Moon, L. Eklundh, and M. A. Friedl (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.

Breure, T. S., S. Haefele, J. A. Hannam, R. Corstanje, R. Webster, S. Moreno-Rojas, and A. Milne (2022). "A loss function to evaluate agricultural decision-making under uncertainty: a case study of soil spectroscopy". *Precision Agriculture* 23.4, 1333–1353. DOI: 10.1007/s11119-022-09887-2.

Breure, T., A. Jones, and P. Panagos (2026). "Evaluating visible near-infrared spectroscopy in context of a repeated sampling survey across the European Union". *Geoderma* 465, 117647. DOI: 10.1016/j.geoderma.2025.117647.

Broeg, T., A. Don, A. Gocht, T. Scholten, R. Taghizadeh-Mehrjardi, and S. Erasmi (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., A. Don, T. Scholten, and S. Erasmi (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., A. Don, M. Wiesmeier, T. Scholten, and S. Erasmi (2024b). "Spatiotemporal Monitoring of Cropland Soil Organic Carbon Changes From Space". *Global Change Biology* 30.12, e17608. DOI: 10.1111/gcb.17608.

Brown, C. F., S. P. Brumby, B. Guzder-Williams, T. Birch, S. B. Hyde, J. Mazzariello, W. Czerwinski, V. J. Pasquarella, R. Haertel, S. Ilyushchenko, et al. (2022). "Dynamic World, Near real-time global 10 m land use land cover mapping". *Scientific Data* 9.1, 251. DOI: 10.1038/s41597-022-01307-4.

Brown, M., K. De Beurs, and M. Marshall (2012). "Global phenological response to climate change in crop areas using satellite remote sensing of vegetation, humidity and temperature over 26 years". *Remote Sensing of Environment* 126, 174–183. DOI: 10.1016/j.rse.2012.08.009.

Brummitt, C. D., C. A. Mathers, R. A. Keating, K. O'Leary, M. Easter, M. A. Friedl, M. DuBuisson, E. E. Campbell, R. Pape, S. J. Peters, et al. (2024). "Solutions and insights for agricultural monitoring, reporting, and verification (MRV) from three consecutive issuances of soil carbon credits". *Journal of Environmental Management* 369, 122284. DOI: 10.1016/j.jenvman.2024.122284.

Brun, P., N. E. Zimmermann, C. Hari, L. Pellissier, and D. N. Karger (2022). "Global climate-related predictors at kilometer resolution for the past and future". *Earth System Science Data* 14.12, 5573–5603. DOI: 10.5194/essd-14-5573-2022.

Brus, D. (2019). "Sampling for digital soil mapping: A tutorial supported by R scripts". *Geoderma* 338, 464–480. DOI: 10.1016/j.geoderma.2018.07.036.

Bui, E. N., B. L. Henderson, and K. Viergever (2006). "Knowledge discovery from models of soil properties developed through data mining". *Ecological Modelling* 191.3-4, 431–446. DOI: 10.1016/j.ecolmodel.2005.05.021.

Bundesanstalt für Geowissenschaften und Rohstoffe (2005). *Soil Regions Map of the European Union and Adjacent Countries 1:5,000,000 (Version 2.0)*. Special Publication. EU catalogue number S.P.I.05.134. Ispra.

Bünemann, E. K., G. Bongiorno, Z. Bai, R. E. Creamer, G. De Deyn, R. De Goede, L. Fleskens, V. Geissen, T. W. Kuyper, P. Mäder, et al. (2018). "Soil quality–A critical review". *Soil biology and biochemistry* 120, 105–125. DOI: 10.1016/j.soilbio.2018.01.030.

Caparros-Santiago, J. A., V. Rodriguez-Galiano, and J. Dash (2021). "Land surface phenology as indicator of global terrestrial ecosystem dynamics: A systematic review". *ISPRS Journal of Photogrammetry and Remote Sensing* 171, 330–347. DOI: 10.1016/j.isprsjprs.2020.11.019.

Carré, F., A. B. McBratney, T. Mayr, and L. Montanarella (2007). "Digital soil assessments: Beyond DSM". *Geoderma* 142.1-2, 69–79. DOI: 10.1016/j.geoderma.2007.08.015.

Castaldi, F., S. Chabrillat, A. Don, and B. van Wesemael (2019). "Soil organic carbon mapping using LUCAS topsoil database and Sentinel-2 data: An approach to reduce soil moisture and crop residue effects". *Remote Sensing* 11.18, 2121. DOI: 10.3390/rs11182121.

Ceddia, M. B., A. S. Gomes, G. M. Vasques, and É. F. Pinheiro (2017). "Soil carbon stock and particle size fractions in the central Amazon predicted from remotely sensed relief, multispectral and radar data". *Remote Sensing* 9.2, 124. DOI: 10.3390/rs9020124.

Celik, M. F., M. S. Isik, O. Yuzugullu, N. Fajraoui, and E. Erten (2022). "Soil Moisture Prediction from Remote Sensing Images Coupled with Climate, Soil Texture and Topography via Deep Learning". *Remote Sensing* 14.21. DOI: 10.3390/rs14215584.

Chahal, I., D. Hooker, B. Deen, K. Janovicek, and L. Van Eerd (2021). "Long-term effects of crop rotation, tillage, and fertilizer nitrogen on soil health indicators and crop productivity in a temperate climate". *Soil and Tillage Research* 213, 105121. DOI: 10.1016/j.still.2021.105121.

Chatenoux, B., J.-P. Richard, D. Small, C. Roeoesli, V. Wingate, C. Poussin, D. Rodila, P. Peduzzi, C. Steinmeier, C. Ginzler, et al. (2021). "The Swiss data cube, analysis ready data archive using earth observations of Switzerland". *Scientific data* 8.1, 295. DOI: 10.1038/s41597-021-01076-6.

Chen, D., N. Chang, J. Xiao, Q. Zhou, and W. Wu (2019). "Mapping dynamics of soil organic matter in croplands with MODIS data and machine learning algorithms". *Science of the Total Environment* 669, 844–855. DOI: 10.1016/j.scitotenv.2019.03.151.

Chen, S., D. Arrouays, V. L. Mulder, L. Poggio, B. Minasny, P. Roudier, Z. Libohova, P. Lagacherie, Z. Shi, J. Hannam, et al. (2022a). "Digital mapping of GlobalSoilMap soil properties at a broad scale: A review". *Geoderma* 409, 115567. DOI: 10.1016/j.geoderma.2021.115567.

Chen, Y., L. Ma, D. Yu, H. Zhang, K. Feng, X. Wang, and J. Song (2022b). "Comparison of feature selection methods for mapping soil organic matter in subtropical restored forests". *Ecological Indicators* 135, 108545. DOI: 10.1016/j.ecolind.2022.108545.

Chen, Y., X. Song, S. Wang, J. Huang, and L. R. Mansaray (2016). "Impacts of spatial heterogeneity on crop area mapping in Canada using MODIS data". *ISPRS Journal of Photogrammetry and Remote Sensing* 119, 451–461. DOI: 10.1016/j.isprsjprs.2016.07.007.

Chen, Z., Q. Shuai, Z. Shi, D. Arrouays, A. C. Richer-de-Forges, and S. Chen (2023). "National-scale mapping of soil organic carbon stock in France: New insights and lessons learned by direct and indirect approaches". *Soil & Environmental Health* 1.4, 100049. DOI: 10.1016/j.seh.2023.100049.

Ciais, P., M. Wattenbach, N. Vuichard, P. Smith, S. L. Piao, A. Don, S. Luyssaert, I. A. Janssens, A. Bondeau, R. Dechow, et al. (2010). "The European carbon balance. Part 2: croplands". *Global Change Biology* 16.5, 1409–1428. DOI: 10.1111/j.1365-2486.2009.02055.x.

Ciampalini, R., P. Lagacherie, C. Gomez, O. Grünberger, M. H. Hamrouni, I. Mekki, and A. Richard (2013). "Detecting, correcting and interpreting the biases of measured soil profile data: A case study in the Cap Bon Region (Tunisia)". *Geoderma* 192, 68–76. DOI: 10.1016/j.geoderma.2012.07.022.

Conrad, O., B. Bechtel, M. Bock, H. Dietrich, E. Fischer, L. Gerlitz, J. Wehberg, V. Wichmann, and J. Böhner (2015). "System for Automated Geoscientific Analyses (SAGA) v. 2.1.4". *Geoscientific Model Development* 8.7, 1991–2007. DOI: 10.5194/gmd-8-1991-2015.

Consoli, D., L. Parente, R. Simoes, M. Şahin, X. Tian, M. Witjes, L. Sloat, and T. Hengl (2024). "A computational framework for processing time-series of earth observation data based on discrete convolution: global-scale historical Landsat cloud-free aggregates at 30 m spatial resolution". *PeerJ* 12, e18585. DOI: 10.7717/peerj.18585.

Courteille, L., P. Lagacherie, N. Boukhelifa, E. Lutton, and L. Tardieu (2024). "Using spatial aggregation of soil multifunctionality maps to support uncertainty-aware planning decisions". *European Journal of Soil Science* 75.4, e13523. DOI: 10.1111/ejss.13523.

Courteille, L., L. Tardieu, N. Boukhelifa, E. Lutton, and P. Lagacherie (2025). "What is the best way to communicate the uncertainty of a digital soil mapping product? Some lessons from an end-users survey". *Geoderma* 459, 117302. DOI: 10.1016/j.geoderma.2025.117302.

Croft, H., N. Kuhn, and K. Anderson (2012). "On the use of remote sensing techniques for monitoring spatio-temporal soil organic carbon dynamics in agricultural systems". *Catena* 94, 64–74. DOI: 10.1016/j.catena.2012.01.001.

Crowther, T. W., K. E. O. Todd-Brown, C. W. Rowe, W. R. Wieder, J. C. Carey, M. B. Machmuller, B. L. Snoek, S. Fang, G. Zhou, S. D. Allison, J. M. Blair, S. D. Bridgham,

A. J. Burton, Y. Carrillo, P. B. Reich, J. S. Clark, A. T. Classen, F. A. Dijkstra, B. Elberling, B. A. Emmett, M. Estiarte, S. D. Frey, J. Guo, J. Harte, L. Jiang, B. R. Johnson, G. Kröel-Dulay, K. S. Larsen, H. Laudon, J. M. Lavallee, Y. Luo, M. Lupascu, L. N. Ma, S. Marhan, A. Michelsen, J. Mohan, S. Niu, E. Pendall, J. Peñuelas, L. Pfeifer-Meister, C. Poll, S. Reinsch, L. L. Reynolds, I. K. Schmidt, S. Sistla, N. W. Sokol, P. H. Templer, K. K. Treseder, J. M. Welker, and M. A. Bradford (2016). "Quantifying global soil carbon losses in response to warming". *Nature* 540.7631, 104–108. DOI: 10.1038/nature20150.

d'Andrimont, R., A. Verhegghen, G. Lemoine, P. Kempeneers, M. Meroni, and M. Van der Velde (2021). "From parcel to continental scale–A first European crop type map based on Sentinel-1 and LUCAS Copernicus in-situ observations". *Remote sensing of environment* 266, 112708. DOI: 10.1016/j.rse.2021.112708.

d'Andrimont, R., M. Yordanov, L. Martinez-Sanchez, B. Eiselt, A. Palmieri, P. Dominici, J. Gallego, H. I. Reuter, C. Joebges, G. Lemoine, et al. (2020). "Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union". *Scientific data* 7.1, 352. DOI: 10.6084/m9.figshare.12841202.

Dark, S. J. and D. Bram (2007). "The modifiable areal unit problem (MAUP) in physical geography". *Progress in physical geography* 31.5, 471–479. DOI: 10.1177/0309133307083294.

Daughtry, C. S., G. Serbin, J. B. Reeves III, P. C. Doraiswamy, and E. R. Hunt Jr (2010). "Spectral reflectance of wheat residue during decomposition and remotely sensed estimates of residue cover". *Remote Sensing* 2.2, 416–431. DOI: 10.3390/rs2020416.

De Beurs, K. M. and G. M. Henebry (2005). "Land surface phenology and temperature variation in the International Geosphere–Biosphere Program high-latitude transects". *Global Change Biology* 11.5, 779–790. DOI: 10.1111/j.1365-2486.2005.00949.x.

– (2010). "Spatio-temporal statistical methods for modelling land surface phenology". *Phenological research: methods for environmental and climate change analysis*, 177–208. DOI: 10.1007/978-90-481-3335-2_9.

de Brogniez, D., C. Ballabio, A. Stevens, R. Jones, L. Montanarella, and B. van Wesemael (2015). "A map of the topsoil organic carbon content of Europe generated by a generalized additive model". *European Journal of Soil Science* 66.1, 121–134. DOI: 10.1111/ejss.12193.

De Bruin, S., D. J. Brus, G. B. Heuvelink, T. van Ebbenhorst Tengbergen, and A. M.-C. Wadoux (2022). "Dealing with clustered samples for assessing map accuracy by cross-validation". *Ecological Informatics* 69, 101665. DOI: 10.1016/j.ecoinf.2022.101665.

De Rosa, D., C. Ballabio, E. Lugato, M. Fasiolo, A. Jones, and P. Panagos (2024). "Soil organic carbon stocks in European croplands and grasslands: How much have we lost in the past decade?" *Global Change Biology* 30.1, e16992. DOI: 10.1111/gcb.16992.

Dee, D., M. Balmaseda, G. Balsamo, R. Engelen, A. Simmons, and J.-N. Thépaut (2014). "Toward a consistent reanalysis of the climate system". *Bulletin of the American Meteorological Society* 95.8, 1235–1248. DOI: 10.1175/BAMS-D-13-00043.1.

Demattê, J. A., J. L. Safanelli, R. R. Poppiel, R. Rizzo, N. E. Q. Silvero, W. d. S. Mendes, B. R. Bonfatti, A. C. Dotto, D. F. U. Salazar, F. A. d. O. Mello, et al. (2020). "Bare earth's surface spectra as a proxy for soil resource monitoring". *Scientific reports* 10.1, 4461. DOI: 10.1038/s41598-020-61408-1.

Deodoro, S. C., R. A. Moral, R. Fealy, T. McCarthy, and R. Fealy (2023). "An assessment of Sentinel-1 synthetic aperture radar, geophysical and topographical covariates for estimating topsoil particle-size fractions". *European Journal of Soil Science* 74.5, e13414. DOI: 10.1111/ejss.13414.

Diek, S., F. Fornallaz, M. E. Schaepman, and R. De Jong (2017). "Barest pixel composite for agricultural areas using landsat time series". *Remote Sensing* 9.12, 1245. DOI: 10.3390/rs9121245.

Ding, Z., K. Liu, S. Grunwald, P. Smith, P. Ciais, B. Wang, A. M.-C. Wadoux, C. Ferreira, S. Karunaratne, N. Shurpali, et al. (2025). "Advancing Soil Organic Carbon Prediction: A Comprehensive Review of Technologies, AI, Process-Based and Hybrid Modelling Approaches". *Advanced Science*, e04152. DOI: 10.1002/advs.202504152.

DLR Geoservice of the Earth Observation Center (2019). *World Settlement Footprint 2019 (WSF2019)*. Accessed: 2024-08-30. German Aerospace Center (DLR).

Dokuchaev, V. (1883). "Russian Chernozem (1883)". *Israel Program for Scientific Translations Ltd.(for USDA-NSF), S. Monson, Jerusalem.*

Don, A., F. Seidel, J. Leifeld, T. Kätterer, M. Martin, S. Pellerin, D. Emde, D. Seitz, and C. Chenu (2024). "Carbon sequestration in soils and climate change mitigation—Definitions and pitfalls". *Global Change Biology* 30.1, e16983. DOI: 10.1111/gcb.16983.

Durante, P., J. M. Requena-Mullor, R. Vargas, M. Guevara, D. Alcaraz-Segura, and C. Oyonarte (2024). "Soil organic carbon maps and associated uncertainty at 90 m for peninsular Spain". *Earth System Science Data Discussions* 2024, 1–29. DOI: 10.5194/essd-2024-431.

Dvorakova, K., U. Heiden, K. Pepers, G. Staats, G. van Os, and B. van Wesemael (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.

Edlinger, A., G. Garland, S. Banerjee, F. Degrune, P. García-Palacios, S. Hallin, C. Herzog, J. Jansa, E. Kost, F. T. Maestre, D. Sánchez-Pescador, L. Philippot, M. Rillig, S. Romdhane, A. Saghaı, A. Spor, E. Frossard, M. van der Heijden, K. Hartman, and A. Held (2022). "Edlinger_etal_database.xlsx". DOI: 10.6084/m9.figshare.15134328.v4.

Edlinger, A., G. Garland, S. Banerjee, F. Degrune, P. García-Palacios, C. Herzog, D. Sánchez Pescador, S. Romdhane, M. Ryo, A. Saghaı, S. Hallin, F. T. Maestre, L. Philippot, M. Rillig, and M. van der Heijden (2023). "The impact of agricultural management on soil aggregation and carbon storage is regulated by climate. Digging deeper Dataset." DOI: 10.6084/m9.figshare.19762114.v5.

Ellili, Y., C. Walter, D. Michot, P. Pichelin, and B. Lemercier (2019). "Mapping soil organic carbon stock change by soil monitoring and digital soil mapping at the landscape scale". *Geoderma* 351, 1–8. DOI: 10.1016/j.geoderma.2019.03.005.

Elvidge, C. D., M. Zhizhin, T. Ghosh, F.-C. Hsu, and J. Taneja (2021). "Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019". *Remote Sensing* 13.5, 922. DOI: 10.3390/rs13050922.

Emick, E., C. Babcock, G. W. White, A. T. Hudak, G. M. Domke, and A. O. Finley (2023). "An approach to estimating forest biomass while quantifying estimate uncertainty and correcting bias in machine learning maps". *Remote Sensing of Environment* 295, 113678. DOI: 10.1016/j.rse.2023.113678.

Emmett, B. A., B. Reynolds, P. M. Chamberlain, E. Rowe, D. Spurgeon, S. A. Brittain, Z. Frogbrook, S. Hughes, A. J. Lawlor, J. Poskitt, E. Potter, D. A. Robinson, A. Scott, C. M. Wood, and C. Woods (2016). *Soil physico-chemical properties 2007 [Countryside Survey]*. NERC Environmental Information Data Centre. DOI: 10.5285/79669141-cde5-49f0-b24d-f3c6a1a52db8.

Eskandari, I., H. Navid, and K. Rangzan (2016). "Evaluating spectral indices for determining conservation and conventional tillage systems in a vetch-wheat rotation". *International Soil and Water Conservation Research* 4.2, 93–98. DOI: 10.1016/j.iswcr.2016.04.002.

Estel, S., T. Kuemmerle, C. Alcántara, C. Levers, A. Prishchepov, and P. Hostert (2015). "Mapping farmland abandonment and recultivation across Europe using MODIS NDVI time series". *Remote Sensing of Environment* 163, 312–325. DOI: 10.1016/j.rse.2015.03.028.

Estel, S., T. Kuemmerle, C. Levers, M. Baumann, and P. Hostert (2016). "Mapping cropland-use intensity across Europe using MODIS NDVI time series". *Environmental Research Letters* 11.2, 024015. DOI: 10.1088/1748-9326/11/2/024015.

Ettehadi Osgouei, P., S. Kaya, E. Sertel, and U. Alganci (2019). "Separating built-up areas from bare land in mediterranean cities using Sentinel-2A imagery". *Remote Sensing* 11.3, 345. DOI: 10.3390/rs11030345.

European Space Agency (2024). *Copernicus Global Digital Elevation Model*. OpenTopography. DOI: 10.5069/G9028PQB.

Even, R. J., M. B. Machmuller, J. M. Lavallee, T. J. Zelikova, and M. F. Cotrufo (2025). "Large errors in soil carbon measurements attributed to inconsistent sample processing". *Soil* 11.1, 17–34. DOI: 10.5194/soil-11-17-2025.

Fan, X., A. Vrieling, B. Muller, and A. Nelson (2020). "Winter cover crops in Dutch maize fields: Variability in quality and its drivers assessed from multi-temporal Sentinel-2 imagery". *International Journal of Applied Earth Observation and Geoinformation* 91, 102139. DOI: 10.1016/j.jag.2020.102139.

Federer, C., D. Turcotte, and C. Smith (1993). "The organic fraction–bulk density relationship and the expression of nutrient content in forest soils". *Canadian Journal of Forest Research* 23.6, 1026–1032. DOI: 10.1139/x93-131.

Feeney, C., B. Cosby, D. Robinson, A. Thomas, B. Emmett, and P. Henrys (2022). "Multiple soil map comparison highlights challenges for predicting topsoil organic carbon concentration at national scale". *Scientific Reports* 12.1, 1379. DOI: 10.1038/s41598-022-05476-5.

Fernández-Ugalde, O., A. Jones, and R. G. Meuli (2020). "Comparison of sampling with a spade and gouge auger for topsoil monitoring at the continental scale". *European Journal of Soil Science* 71.2, 137–150. DOI: 10.1111/ejss.12862.

Flores-Anderson, A. I., K. E. Herndon, R. B. Thapa, and E. Cherrington (2019). *The SAR handbook: Comprehensive methodologies for forest monitoring and biomass estimation*. NASA.

Friedlingstein, P., M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti (2014). "Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks". *Journal of Climate* 27.2, 511–526. DOI: 10.1175/JCLI-D-12-00579.1.

Fritsch, F. N. and J. Butland (1984). "A method for constructing local monotone piecewise cubic interpolants". *SIAM journal on scientific and statistical computing* 5.2, 300–304. DOI: 10.1137/0905021.

Froger, C., E. Tondini, D. Arrouays, K. Oorts, C. Poeplau, J. Wetterlind, E. Putku, N. P. Saby, M. Fantappiè, Q. Styc, et al. (2024). "Comparing LUCAS Soil and national systems: Towards a harmonized European Soil monitoring network". *Geoderma* 449, 117027. DOI: 10.1016/j.geoderma.2024.117027.

Ganguly, S., M. A. Friedl, B. Tan, X. Zhang, and M. Verma (2010). "Land surface phenology from MODIS: Characterization of the Collection 5 global land cover dynamics product". *Remote sensing of environment* 114.8, 1805–1816. DOI: 10.1016/j.rse.2010.04.005.

Gao, B.-C. (1996). "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space". *Remote sensing of environment* 58.3, 257–266. DOI: 10.1016/S0034-4257(96)00067-3.

Gasch, C. K., T. Hengl, B. Gräler, H. Meyer, T. S. Magney, and D. J. Brown (2015). "Spatio-temporal interpolation of soil water, temperature, and electrical conductivity in 3D+ T: The Cook Agronomy Farm data set". *Spatial Statistics* 14, 70–90. DOI: 10.1016/j.spasta.2015.04.001.

Giuliani, G., H. Cazeaux, P.-Y. Burgi, C. Poussin, J.-P. Richard, and B. Chatenoux (2021). "SwissEnveo: A FAIR national environmental data repository for earth observation open science". *Data Science Journal* 20, 22–22. DOI: 10.5334/dsj-2021-022.

Giuliani, G., B. Chatenoux, A. De Bono, D. Rodila, J.-P. Richard, K. Allenbach, H. Dao, and P. Peduzzi (2017). "Building an earth observations data cube: lessons learned from the swiss data cube (sdc) on generating analysis ready data (ard)". *Big Earth Data* 1.1-2, 100–117. DOI: 10.1080/20964471.2017.1398903.

Gocke, M. I., J. Guigue, S. L. Bauke, D. Barkusky, M. Baumecker, A. E. Berns, E. Hobley, B. Honermeier, I. Kögel-Knabner, S. Koszinski, et al. (2023). "Interactive effects of agricultural management on soil organic carbon accrual: A synthesis of long-term field experiments in Germany". *Geoderma* 438, 116616. DOI: 10.1016/j.geoderma.2023.116616.

Goidts, E., B. Van Wesemael, and M. Crucifix (2009). "Magnitude and sources of uncertainties in soil organic carbon (SOC) stock assessments at various scales". *European Journal of Soil Science* 60.5, 723–739. DOI: 10.1111/j.1365-2389.2009.01157.x.

Gomes, L. C., R. M. Faria, E. de Souza, G. V. Veloso, C. E. G. Schaefer, and E. I. Fernandes Filho (2019). "Modelling and mapping soil organic carbon stocks in Brazil". *Geoderma* 340, 337–350. DOI: 10.1016/j.geoderma.2019.01.007.

Gomes, V. C., G. R. Queiroz, and K. R. Ferreira (2020). "An overview of platforms for big earth observation data management and analysis". *Remote Sensing* 12.8, 1253. DOI: 10.3390/rs12081253.

Gómez, C., J. C. White, and M. A. Wulder (2016). "Optical remotely sensed time series data for land cover classification: A review". *ISPRS Journal of photogrammetry and Remote Sensing* 116, 55–72. DOI: 10.1016/j.isprsjprs.2016.03.008.

Gong, Z., X. Zhang, J. Chen, and G. Zhang (2003). "Origin and development of soil science in ancient China". *Geoderma* 115.1-2, 3–13. DOI: 10.1016/S0016-7061(03)00071-5.

Goovaerts, P. (2001). "Geostatistical modelling of uncertainty in soil science". *Geoderma* 103.1-2, 3–26. DOI: 10.1016/S0016-7061(01)00067-2.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore (2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone". *Remote sensing of Environment* 202, 18–27. DOI: 10.1016/j.rse.2017.06.031.

GRASS Development Team (2023). *Geographic Resources Analysis Support System (GRASS GIS) Software, Version 8.3*. Open Source Geospatial Foundation. USA. DOI: 10.5281/zenodo.5176030. URL: https://grass.osgeo.org.

Gross, C. D. and R. B. Harrison (2019). "The case for digging deeper: soil organic carbon storage, dynamics, and controls in our changing world". *Soil Systems* 3.2, 28. DOI: 10.3390/soilsystems3020028.

Grover, K., J. L. Safanelli, J. Sanderman, B. G. Hopkins, and C. Brungard (2025). "The Utility of Laboratory Measurement Uncertainty: A Case-Study Using the Open Soil Spectral Library Service". *European Journal of Soil Science* 76.4, e70166. DOI: 10.1111/ejss.70166.

Gruijter, J. J. de, M. F. Bierkens, D. J. Brus, and M. Knotters (2006). *Sampling for natural resource monitoring*. Springer.

Gubler, A., D. Wächter, P. Schwab, M. Müller, and A. Keller (2019). "Twenty-five years of observations of soil organic carbon in Swiss croplands showing stability overall but with some divergent trends". *Environmental Monitoring and Assessment* 191, 1–17. DOI: 10.1007/s10661-019-7435-y.

Guevara, M., C. Arroyo, N. Brunsell, C. O. Cruz, G. Domke, J. Equihua, J. Etchevers, D. Hayes, T. Hengl, A. Ibelles, K. Johnson, B. de Jong, Z. Libohova, R. Llamas, L. Nave, J. L. Ornelas, F. Paz, R. Ressl, A. Schwartz, A. Victoria, S. Wills, and R. Vargas (2020). "Soil Organic Carbon Across Mexico and the Conterminous United States (1991–2010)". *Global Biogeochemical Cycles* 34.3, no–no.

Hackländer, J., L. Parente, Y.-F. Ho, T. Hengl, R. Simoes, D. Consoli, M. Şahin, X. Tian, M. Jung, M. Herold, G. Duveiller, M. Weynants, and I. Wheeler (2024). "Land potential assessment and trend-analysis using 2000–2021 FAPAR monthly time-series at 250 m spatial resolution". *PeerJ* 12, e16972. DOI: 10.7717/peerj.16972.

Hamzehpour, N., H. Shafizadeh-Moghadam, and R. Valavi (2019). "Exploring the driving forces and digital mapping of soil organic carbon using remote sensing and soil texture". *Catena* 182, 104141. DOI: 10.1016/j.catena.2019.104141.

Hao, X., B. Ball, J. Culley, M. Carter, and G. Parkin (2008). "Soil density and porosity". *Soil sampling and methods of analysis* 2.1, 743–759.

Harbo, L. S., R. Lama, C. Lemming, and L. Elsgaard (2025). "Importance of sampling frequency for the observed dynamics of SOC content in the Danish long-term monitoring network". *Geoderma Regional*, e00931. DOI: 10.1016/j.geodrs.2025.e00931.

Harbo, L. S., J. E. Olesen, C. Lemming, B. T. Christensen, and L. Elsgaard (2023). "Limitations of farm management data in analyses of decadal changes in SOC stocks in the Danish soil-monitoring network". *European Journal of Soil Science* 74.3, e13379. DOI: 10.1111/ejss.13379.

Harper, K. L., C. Lamarche, A. Hartley, P. Peylin, C. Ottlé, V. Bastrikov, R. San Martín, S. I. Bohnenstengel, G. Kirches, M. Boettcher, et al. (2023). "A 29-year time series of annual 300 m resolution plant-functional-type maps for climate models". *Earth System Science Data* 15.3, 1465–1499. DOI: 10.5194/essd-15-1465-2023.

Hartemink, A. E., P. Krasilnikov, and J. Bockheim (2013). "Soil maps of the world". *Geoderma* 207, 256–267. DOI: 10.1016/j.geoderma.2013.05.003.

Hateffard, F., L. Steinbuch, and G. B. Heuvelink (2024). "Evaluating the extrapolation potential of random forest digital soil mapping". *Geoderma* 441, 116740. DOI: 10.1016/j.geoderma.2023.116740.

He, X., L. Yang, A. Li, L. Zhang, F. Shen, Y. Cai, and C. Zhou (2021). "Soil organic carbon prediction using phenological parameters and remote sensing variables generated from Sentinel-2 images". *Catena* 205, 105442. DOI: 10.1016/j.catena.2021.105442.

Heiden, U., P. d'Angelo, P. Schwind, P. Karlshöfer, R. Müller, S. Zepp, M. Wiesmeier, and P. Reinartz (2022). "Soil reflectance composites—improved thresholding and performance evaluation". *Remote Sensing* 14.18, 4526. DOI: 10.3390/rs14184526.

Helfenstein, A., V. L. Mulder, M. J. Hack-ten Broeke, M. van Doorn, K. Teuling, D. J. Walvoort, and G. Heuvelink (2024a). "BIS-4D: mapping soil properties and their uncertainties at 25 m resolution in the Netherlands". *Earth System Science Data* 16.6, 2941–2970. DOI: 10.5194/essd-16-2941-2024.

Helfenstein, A., V. L. Mulder, G. B. Heuvelink, and M. J. Hack-ten Broeke (2024b). "Three-dimensional space and time mapping reveals soil organic matter decreases across anthropogenic landscapes in the Netherlands". *Communications Earth & Environment* 5.1, 130. DOI: 10.1038/s43247-024-01293-y.

Henebry, G. M. and K. M. De Beurs (2013). "Remote sensing of land surface phenology: A prospectus". In: *Phenology: An integrative environmental science*. Springer, 385–411. DOI: 10.1007/978-94-007-6925-0_21.

Hengl, T. (2023). *Annual time series of global VIIRS nighttime lights for 2000-2021 at 500-m spatial resolution extrapolated using logistic regression*. Version v0.2. [Dataset]. Wageningen, NL. DOI: 10.5281/zenodo.8277198.

Hengl, T., J. M. de Jesus, R. A. MacMillan, N. H. Batjes, G. B. Heuvelink, E. Ribeiro, A. Samuel-Rosa, B. Kempen, J. G. Leenaars, M. G. Walsh, et al. (2014). "SoilGrids1km—global soil information based on automated mapping". *PloS one* 9.8, e105992.

Hengl, T. and R. A. MacMillan (2019). "Predictive soil mapping with R". *Wageningen: OpenGeoHub foundation*.

Hengl, T., J. Mendes de Jesus, G. B. Heuvelink, M. Ruiperez Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, et al. (2017). "SoilGrids250m: Global gridded soil information based on machine learning". *PLoS one* 12.2, e0169748. DOI: 10.1371/journal.pone.0169748.

Hengl, T., M. Nussbaum, M. N. Wright, G. B. Heuvelink, and B. Gräler (2018). "Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables". *PeerJ* 6, e5518. DOI: 10.7717/peerj.5518.

Hernández, T. D. B., B. K. Slater, J. M. Shaffer, and N. Basta (2023). "Comparison of methods for determining organic carbon content of urban soils in Central Ohio". *Geoderma Regional* 34, e00680. DOI: 10.1016/j.geodrs.2023.e00680.

Heuvelink, G., M. Angelini, L. Poggio, Z. Bai, N. Batjes, H. van den Bosch, D. Bossio, S. Estella, J. Lehmann, G. Olmedo, and J. Sanderman (2020). "Machine learning in space and time for modelling soil organic carbon change". *European Journal of Soil Science*, NA.

Heuvelink, G. B. M. (2018). "Uncertainty and Uncertainty Propagation in Soil Mapping and Modelling". In: *Pedometrics*. Ed. by A. B. McBratney, B. Minasny, and U. Stockmann. Cham: Springer International Publishing, 439–461. DOI: 10.1007/978-3-319-63439-5_14.

– (1998). *Error propagation in environmental modelling with GIS*. CRC press. DOI: 10.4324/9780203016114.

– (2014). "Uncertainty quantification of GlobalSoilMap products". *GlobalSoilMap: Basis of the global spatial soil information system*, 335–340. DOI: 10.1201/b16500-62.

Heuvelink, G. B., M. E. Angelini, L. Poggio, Z. Bai, N. H. Batjes, R. van den Bosch, D. Bossio, S. Estella, J. Lehmann, G. F. Olmedo, et al. (2021). "Machine learning in space and time for modelling soil organic carbon change". *European Journal of Soil Science* 72.4, 1607–1623. DOI: 10.1111/ejss.12998.

Heuvelink, G. B., J. Kros, G. J. Reinds, and W. De Vries (2016). "Geostatistical prediction and simulation of European soil property maps". *Geoderma Regional* 7.2, 201–215. DOI: 10.1016/j.geodrs.2016.04.002.

Heuvelink, G. B. and R. Webster (2022). "Spatial statistics and soil mapping: A blossoming partnership under pressure". *Spatial statistics* 50, 100639. DOI: 10.1016/j.spasta.2022.100639.

Hicks Pries, C. E., R. Ryals, B. Zhu, K. Min, A. Cooper, S. Goldsmith, J. Pett-Ridge, M. Torn, and A. A. Berhe (2023). "The deep soil organic carbon response to global change". *Annual Review of Ecology, Evolution, and Systematics* 54.1, 375–401. DOI: 10.1146/annurev-ecolsys-102320-085332.

Hiederer, R. (2020). "Data Evaluation of LUCAS Soil Survey Laboratory 2009 to 2015 Data". *EUR Report, Publications Office of the European Union* EUR 30092 EN. JRC119881. DOI: 10.2760/791714.

Hiederer, R., H. T. DURRANT, et al. (2010). "Evaluation of BioSoil Demonstration Project-Preliminary Data Analysis".

Hill, M. J. and J. P. Guerschman (2022). "Global trends in vegetation fractional cover: Hotspots for change in bare soil and non-photosynthetic vegetation". *Agriculture, Ecosystems & Environment* 324, 107719. DOI: 10.1016/j.agee.2021.107719.

Ho, Y.-F., C. H. Grohmann, J. Lindsay, H. I. Reuter, L. Parente, M. Witjes, and T. Hengl (2025). "Global Ensemble Digital Terrain modeling and parametrization at 30 m resolution (GEDTM30): a data fusion approach based on ICESat-2, GEDI and multisource data". *ResearchSquare*. DOI: 10.21203/rs.3.rs-6280607/v1.

Ho, Y. F., T. Hengl, and L. Parente (2023). *Ensemble Digital Terrain Model (EDTM) of the world*. Zenodo. DOI: 10.5281/zenodo.7634679.

Holmes, K. W., E. A. Griffin, and D. van Gool (2021). "Digital soil mapping of coarse fragments in southwest Australia: Targeting simple features yields detailed maps". *Geoderma* 404, 115282. DOI: 10.1016/j.geoderma.2021.115282.

Hu, B., H. Bourennane, D. Arrouays, P. Denoroy, B. Lemercier, and N. P. Saby (2021). "Developing pedotransfer functions to harmonize extractable soil phosphorus content measured with different methods: A case study across the mainland of France". *Geoderma* 381, 114645. DOI: 10.1016/j.geoderma.2020.114645.

Hu, P.-L., S.-J. Liu, Y.-Y. Ye, W. Zhang, K.-L. Wang, and Y.-R. Su (2018). "Effects of environmental factors on soil organic carbon under natural or managed vegetation restoration". *Land Degradation & Development* 29.3, 387–397. DOI: 10.1002/ldr.2876.

Huang, Q., G. B. Heuvelink, P. He, J. G. Leenaars, and A. G. Schut (2024). "Combining production ecology principles with random forest to model potato yield in China". *Field Crops Research* 319, 109619. DOI: 10.1016/j.fcr.2024.109619.

Huang, Y.-C., J. Padarian, B. Minasny, and A. B. McBratney (2025). "Using Monte Carlo conformal prediction to evaluate the uncertainty of deep-learning soil spectral models". *Soil* 11.2, 553–563. DOI: 10.5194/soil-11-553-2025.

Huat, B. B., S. Kazemian, A. Prasad, and M. Barghchi (2011). "State of an art review of peat: General perspective". *International Journal of the Physical Sciences* 6.8, 1988–1996. DOI: 10.5897/IJPS11.192.

Hudson, B. D. (1992). "The soil survey as paradigm-based science". *Soil Science Society of America Journal* 56.3, 836–841. DOI: 10.2136/sssaj1992.03615995005600030027x.

Huete, A. R. (1988). "A soil-adjusted vegetation index (SAVI)". *Remote sensing of environment* 25.3, 295–309. DOI: 10.1016/0034-4257(88)90106-X.

Iosifescu Enescu, I., L. de Espona, D. Haas-Artho, R. Kurup Buchholz, D. Hanimann, M. Rüetschi, D. N. Karger, G.-K. Plattner, M. Hägeli, C. Ginzler, et al. (2021). "Cloud optimized raster encoding (core): A web-native streamable format for large environmental time series". *Geomatics* 1.3, 369–382. DOI: 10.3390/geomatics1030021.

Isik, S., R. Minarik, and T. Hengl (2024). *Continental Europe surface lithology based on EGDI/OneGeology map at 1:1M scale*. Version V2. Zenodo. DOI: 10.5281/zenodo.12607973.

Jackson, R. B., K. Lajtha, S. E. Crow, G. Hugelius, M. G. Kramer, and G. Piñeiro (2017). "The Ecology of Soil Carbon: Pools, Vulnerabilities, and Biotic and Abiotic Controls". *Annual Review of Ecology, Evolution, and Systematics* 48.1.

Japan Aerospace Exploration Agency (2021). *ALOS World 3D 30 meter DEM. V3.2*. OpenTopography. DOI: 10.5069/G94M92HB.

Jenny, H. (1994). *Factors of soil formation: a system of quantitative pedology*. Dover books on Earth sciences. Dover Publications.

Jenny, H. (1941). *Factors of soil formation: a system of quantitative pedology*. Courier Corporation.

Jobbágy, E. G. and R. B. Jackson (2000). "The vertical distribution of soil organic carbon and its relation to climate and vegetation". *Ecological applications* 10.2, 423–436.

Jones, R. J., R. Hiederer, E. Rusco, and L. Montanarella (2005). "Estimating organic carbon in the soils of Europe for policy support". *European Journal of Soil Science* 56.5, 655–671. DOI: 10.1111/j.1365-2389.2005.00728.x.

Jong, R. de and S. de Bruin (2012). "Linear trends in seasonal vegetation time series and the modifiable temporal unit problem". *Biogeosciences* 9.1, 71–77. DOI: 10.5194/bg-9-71-2012.

Kamilaris, A. and F. X. Prenafeta-Boldú (2018). "Deep learning in agriculture: A survey". *Computers and electronics in agriculture* 147, 70–90. DOI: 10.1016/j.compag.2018.02.016.

Kansakar, P. and F. Hossain (2016). "A review of applications of satellite earth observation data for global societal benefit and stewardship of planet earth". *Space Policy* 36, 46–54. DOI: 10.1016/j.spacepol.2016.05.005.

Karger, D. N., O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, H. P. Linder, and M. Kessler (2017). "Climatologies at high resolution for the earth's land surface areas". *Scientific data* 4.1, 1–20. DOI: 10.1038/sdata.2017.122.

Karger, D. N., A. M. Wilson, C. Mahony, N. E. Zimmermann, and W. Jetz (2021). "Global daily 1 km land surface precipitation based on cloud cover-informed downscaling". *Scientific Data* 8.1, 307. DOI: https://www.nature.com/articles/s41597-021-01084-6.

Karniadakis, G. E., I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang (2021). "Physics-informed machine learning". *Nature Reviews Physics* 3.6, 422–440. DOI: 10.1038/s42254-021-00314-5.

Kasraei, B., B. Heung, D. D. Saurette, M. G. Schmidt, C. E. Bulmer, and W. Bethel (2021). "Quantile regression as a generic approach for estimating uncertainty of digital soil maps produced from machine-learning". *Environmental Modelling & Software* 144, 105139. DOI: 10.1016/j.envsoft.2021.105139.

Keel, S., A. Johannes, P. Boivin, S. Burgos, R. Charles, F. Hagedorn, B. Kulli, J. Leifeld, A. Saluz, and S. Zimmermann (2021). *Soil carbon sequestration in Switzerland: analysis of potentials and measures*. Bern, CH: Report by Agroscope, 129.

Kerr, D. D. and T. E. Ochsner (2020). "Soil organic carbon more strongly related to soil moisture than soil temperature in temperate grasslands". *Soil Science Society of America Journal* 84.2, 587–596. DOI: 10.1002/saj2.20018.

Kidd, D., R. Searle, M. Grundy, A. McBratney, N. Robinson, L. O'Brien, P. Zund, D. Arrouays, M. Thomas, J. Padarian, et al. (2020). "Operationalising digital soil mapping–Lessons from Australia". *Geoderma Regional* 23, e00335. DOI: 10.1016/j.geodrs.2020.e00335.

Killough, B. (2018). "Overview of the open data cube initiative". In: *IGARSS 2018-2018 IEEE international geoscience and remote sensing symposium*. IEEE, 8629–8632. DOI: 10.1109/IGARSS.2018.8517694.

– (2019). "The impact of analysis ready data in the Africa regional data cube". In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 5646–5649. DOI: 10.1109/IGARSS.2019.8898321.

Konen, M. E., P. M. Jacobs, C. L. Burras, B. J. Talaga, and J. A. Mason (2002). "Equations for predicting soil organic carbon using loss-on-ignition for north central US soils". *Soil science society of America journal* 66.6, 1878–1881. DOI: 10.2136/sssaj2002.1878.

Kooch, Y., M. Amani, and M. Abedi (2022). "Vegetation degradation threatens soil health in a mountainous semi-arid regiony". *Science of The Total Environment* 830, 154827. DOI: 10.1016/j.scitotenv.2022.154827.

Kumar, S., Y. S. Ghotekar, and V. K. Dadhwal (2019). "C-equivalent correction factor for soil organic carbon inventory by wet oxidation, dry combustion and loss on ignition methods in Himalayan region". *Journal of Earth System Science* 128, 1–10. DOI: 10.1007/s12040-019-1086-9.

Lacagnina, C., R. David, A. Nikiforova, M.-E. Kuusniemi, C. Cappiello, O. Biehlmaier, L. Wright, C. Schubert, A. Bertino, H. Thiemann, et al. (2022). "TOWARDS A DATA QUALITY FRAMEWORK FOR EOSC Authorship Community". PhD thesis. EOSC Association. DOI: 10.5281/zenodo.7515816.

Ladoni, M., H. A. Bahrami, S. K. Alavipanah, and A. A. Norouzi (2010). "Estimating soil organic carbon from soil reflectance: a review". *Precision Agriculture* 11.1, 82–99. DOI: 10.1007/s11119-009-9123-3.

Lagacherie, P. and A. McBratney (2006). "Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping". *Developments in soil science* 31, 3–22. DOI: 10.1016/S0166-2481(06)31001-X.

Lagacherie, P. and C. Gomez (2018). "Vis-NIR-SWIR remote sensing products as new soil data for digital soil mapping". In: *Pedometrics*. Springer, 415–437. DOI: 10.1007/978-3-319-63439-5_13.

Lal, R. (2004). "Soil carbon sequestration to mitigate climate change". *Geoderma* 123.1-2, 1–22. DOI: 10.1016/j.geoderma.2004.01.032.

Lal, R., J. Bouma, E. Brevik, L. Dawson, D. J. Field, B. Glaser, R. Hatano, A. E. Hartemink, T. Kosaki, B. Lascelles, et al. (2021). "Soils and sustainable development goals of the

United Nations: An International Union of Soil Sciences perspective". *Geoderma Regional* 25, e00398. DOI: 10.1016/j.geodrs.2021.e00398.

Lamb, A., R. Green, I. Bateman, M. Broadmeadow, T. Bruce, J. Burney, P. Carey, D. Chadwick, E. Crane, R. Field, K. Goulding, H. Griffiths, A. Hastings, T. Kasoar, D. Kindred, B. Phalan, J. Pickett, P. Smith, E. Wall, E. K. H. J. zu Ermgassen, and A. Balmford (2016). "The potential for land sparing to offset greenhouse gas emissions from agriculture". *Nature Climate Change* 6.5, 488–492.

Lamichhane, S., L. Kumar, and B. Wilson (2019). "Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review". *Geoderma* 352, 395–413. DOI: 10.1016/j.geoderma.2019.05.031.

Lampach, N., J. O. Skoien, H. Ramos, J. Gaffuri, R. Koeble, L. See, and M. Van der Velde (2025). "Statistical Atlas of European Agriculture: Gridded Data from the Agricultural Census 2020 and the Spatial Distribution of CAP Contextual Indicators". *Earth System Science Data Discussions* 2025, 1–35. DOI: 10.5194/essd-17-3893-2025.

Lawrence, I. and K. Lin (1989). "A concordance correlation coefficient to evaluate reproducibility". *Biometrics*, 255–268. DOI: 10.2307/2532051.

Leenaars, J., B. Kempen, A. van Oostrum, and N. Batjes (2014). "Africa Soil Profiles Database: A compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa". *Arrouays et al.(eds.), 2014b*, 51–57.

Leeuwen, C. C. van, V. L. Mulder, N. H. Batjes, and G. B. Heuvelink (2024). "Effect of measurement error in wet chemistry soil data on the calibration and model performance of pedotransfer functions". *Geoderma* 442, 116762. DOI: 10.1016/j.geoderma.2023.116762.

Legros, J.-P. (2006). *Mapping of the Soil*. Science Publishers.

Lehmann, J., D. A. Bossio, I. Kögel-Knabner, and M. C. Rillig (2020). "The concept and future prospects of soil health". *Nature Reviews Earth & Environment* 1.10, 544–553. DOI: 10.1038/s43017-020-0080-8.

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

Lettens, S., J. Van Orshoven, B. Van Wesemael, B. Muys, and D. Perrin (2005). "Soil organic carbon changes in landscape units of Belgium between 1960 and 2000 with reference to 1990". *Global Change Biology* 11.12, 2128–2140. DOI: Citations:110SuzannaLettens,tel.+3216329755,fax+3216329760,e-mail:suzanna.lettens@biw.kuleuven.beSectionsPDF.

Lewis, A., S. Oliver, L. Lymburner, B. Evans, L. Wyborn, N. Mueller, G. Raevksi, J. Hooke, R. Woodcock, J. Sixsmith, et al. (2017). "The Australian geoscience data

cube—foundations and lessons learned". *Remote Sensing of Environment* 202, 276–292. DOI: 10.1016/j.rse.2017.03.015.

Li, H., Y. Wu, S. Liu, J. Xiao, W. Zhao, J. Chen, G. Alexandrov, and Y. Cao (2022). "Decipher soil organic carbon dynamics and driving forces across China using machine learning". *Global Change Biology* 28.10, 3394–3410. DOI: 10.1111/gcb.16154.

Li, L., M. A. Friedl, Q. Xin, J. Gray, Y. Pan, and S. Frolking (2014). "Mapping crop cycles in China using MODIS-EVI time series". *Remote Sensing* 6.3, 2473–2493. DOI: 10.3390/rs6032473.

Libohova, Z., C. Seybold, K. Adhikari, S. Wills, D. Beaudette, S. Peaslee, D. Lindbo, and P. Owens (2019). "The anatomy of uncertainty for soil pH measurements and predictions: Implications for modellers and practitioners". *European journal of soil science* 70.1, 185–199. DOI: 10.1111/ejss.12770.

Ließ, M. and A. Sakhaee (2024). "Deep Learning with a Multi-Task Convolutional Neural Network to Generate a National-Scale 3D Soil Data Product: The Particle Size Distribution of the German Agricultural Soil Landscape". *Agriculture* 14.8, 1230. DOI: 10.3390/agriculture14081230.

Liu, C., Q. Zhang, S. Tao, J. Qi, M. Ding, Q. Guan, B. Wu, M. Zhang, M. Nabil, F. Tian, et al. (2020). "A new framework to map fine resolution cropping intensity across the globe: Algorithm, validation, and implication". *Remote Sensing of Environment* 251, 112095. DOI: 10.1016/j.rse.2020.112095.

Liu, F. T., K. M. Ting, and Z.-H. Zhou (2008). "Isolation forest". In: *2008 eighth IEEE International Conference on Data Mining*. IEEE, 413–422. DOI: 10.1109/ICDM.2008.17.

Liu, F., D. G. Rossiter, X.-D. Song, G.-L. Zhang, R.-M. Yang, Y.-G. Zhao, D.-C. Li, and B. Ju (2016). "A similarity-based method for three-dimensional prediction of soil organic matter concentration". *Geoderma* 263, 254–263. DOI: 10.1016/j.geoderma.2015.05.013.

Liu, L., W. Zhou, K. Guan, B. Peng, S. Xu, J. Tang, Q. Zhu, J. Till, X. Jia, C. Jiang, et al. (2024). "Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems". *Nature communications* 15.1, 357. DOI: 10.1038/s41467-023-43860-5.

Loiseau, T., D. Arrouays, A. C. Richer-de-Forges, P. Lagacherie, C. Ducommun, and B. Minasny (2021). "Density of soil observations in digital soil mapping: A study in the Mayenne region, France". *Geoderma Regional* 24, e00358. DOI: 10.1016/j.geodrs.2021.e00358.

Loiseau, T., S. Chen, V. L. Mulder, M. R. Dobarco, A. C. Richer-de-Forges, S. Lehmann, H. Bourennane, N. P. Saby, M. P. Martin, E. Vaudour, et al. (2019). "Satellite data integration for soil clay content modelling at a national scale". *International Journal of Applied Earth Observation and Geoinformation* 82, 101905. DOI: 10.1016/j.jag.2019.101905.

Lokers, R., R. Knapen, S. Janssen, Y. van Randen, and J. Jansen (2016). "Analysis of Big Data technologies for use in agro-environmental science". *Environmental modelling & software* 84, 494–504. DOI: 10.1016/j.envsoft.2016.07.017.

Lorenz, K. and R. Lal (2005). "The depth distribution of soil organic carbon in relation to land use and management and the potential of carbon sequestration in subsoil horizons". *Advances in agronomy* 88, 35–66. DOI: 10.1016/S0065-2113(05)88002-2.

Louis, B., N. Saby, T. Orton, E. Lacarce, L. Boulonne, C. Jolivet, C. Ratié, and D. Arrouays (2014). "Statistical sampling design impact on predictive quality of harmonization functions between soil monitoring networks". *Geoderma* 213, 133–143. DOI: 10.1016/j.geoderma.2013.07.018.

Luo, C., W. Zhang, X. Zhang, and H. Liu (2024). "Mapping the soil organic matter content in a typical black-soil area using optical data, radar data and environmental covariates". *Soil and Tillage research* 235, 105912. DOI: 10.1016/j.still.2023.105912.

Luo, Z., W. Feng, Y. Luo, J. Baldock, and E. Wang (2017). "Soil organic carbon dynamics jointly controlled by climate, carbon inputs, soil properties and soil carbon fractions". *Global change biology* 23.10, 4430–4439. DOI: 10.1111/gcb.13767.

Lyapustin, A. and Y. Wang (2018). *MCD19A2 MODIS/Terra+Aqua Land Aerosol Optical Depth Daily L2G Global 1km SIN Grid V006*. NASA EOSDIS Land Processes Distributed Active Archive Center. DOI: 10.5067/MODIS/MCD19A2.006.

Ma, Y., D. Woolf, M. Fan, L. Qiao, R. Li, and J. Lehmann (2023). "Global crop production increase by soil organic carbon". *Nature Geoscience* 16.12, 1159–1165. DOI: 10.1038/s41561-023-01302-3.

Ma, Y., B. Minasny, B. P. Malone, and A. B. Mcbratney (2019). "Pedology and digital soil mapping (DSM)". *European Journal of Soil Science* 70.2, 216–235. DOI: 10.1111/ejss.12790.

Ma, Y., B. Minasny, A. McBratney, L. Poggio, and M. Fajardo (2021). "Predicting soil properties in 3D: Should depth be a covariate?" *Geoderma* 383, 114794. DOI: 10.1016/j.geoderma.2020.114794.

Magdoff, F., H. Van Es, et al. (2000). *Building soils for better crops*. Vol. 4. Sustainable Agriculture Network Beltsville.

Mallavan, B., B. Minasny, and A. McBratney (2010). "Homosoil, a methodology for quantitative extrapolation of soil information across the globe". In: *Digital soil mapping: bridging research, environmental application, and operation*. Springer, 137–150. DOI: 10.1007/978-90-481-8863-5_12.

Malone, B. P., B. Minasny, A. B. McBratney, B. P. Malone, B. Minasny, and A. B. McBratney (2017). "Some methods for the quantification of prediction uncertainties for digital soil mapping". *Using R for digital soil mapping*, 169–219. DOI: 10.1007/978-3-319-44327-0_7.

MapBiomas (2023). *Annual mapping of soil organic carbon stock in Brazil 1985-2021 (beta collection) V1*. MapBiomas Data. DOI: 10.58053/MapBiomas/DHAYLZ.

Martin, M., T. Orton, E. Lacarce, J. Meersmans, N. Saby, J. Paroissien, C. Jolivet, L. Boulonne, and D. Arrouays (2014). "Evaluation of modelling approaches for predicting the spatial distribution of soil organic carbon stocks at the national scale". *Geoderma* 223, 97–107. DOI: 10.1016/j.geoderma.2014.01.005.

Mattina, D., H. Erdogan, I. Wheeler, N. Crossman, R. Cumani, and S. Minelli (2018). "Default data: methods and interpretation". In: *A guidance document for the 2018 UNCCD reporting*. Bonn, Germany: United Nations Convention to Combat Desertification (UNCCD).

Maynard, J. J. and M. R. Levi (2017). "Hyper-temporal remote sensing for digital soil mapping: Characterizing soil-vegetation response to climatic variability". *Geoderma* 285, 94–109. DOI: 10.1016/j.geoderma.2016.09.024.

McBratney, A., D. J. Field, and A. Koch (2014a). "The dimensions of soil security". *Geoderma* 213, 203–213. DOI: 10.1016/j.geoderma.2013.08.013.

McBratney, A. B., M. M. Santos, and B. Minasny (2003). "On digital soil mapping". *Geoderma* 117.1-2, 3–52. DOI: 10.1016/S0016-7061(03)00223-4.

McBratney, A. B., U. Stockmann, D. A. Angers, B. Minasny, and D. J. Field (2014b). "Challenges for soil organic carbon research". In: *Soil carbon*. Springer, 3–16.

McKenzie, N. J. and P. J. Ryan (1999). "Spatial prediction of soil properties using environmental correlation". *Geoderma* 89.1-2, 67–94. DOI: 10.1016/S0016-7061(98)00137-2.

Meersmans, J., B. van Wesemael, E. Goidts, M. Van Molle, S. De Baets, and F. De Ridder (2011). "Spatial analysis of soil organic carbon evolution in Belgian croplands and grasslands, 1960–2006". *Global Change Biology* 17.1, 466–479. DOI: 10.1111/j.1365-2486.2010.02183.x.

Meinshausen, N. (2006). "Quantile regression forests." *Journal of machine learning research* 7.6.

Meng, X., Y. Bao, C. Luo, X. Zhang, and H. Liu (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.

Meroni, M., R. d'Andrimont, A. Vrieling, D. Fasbender, G. Lemoine, F. Rembold, L. Seguini, and A. Verhegghen (2021). "Comparing land surface phenology of major European crops as derived from SAR and multispectral data of Sentinel-1 and-2". *Remote sensing of environment* 253, 112232. DOI: 10.1016/j.rse.2020.112232.

Meyer, H. and E. Pebesma (2021). "Predicting into unknown space? Estimating the area of applicability of spatial prediction models". *Methods in Ecology and Evolution* 12.9, 1620–1633. DOI: 10.1111/2041-210X.13650.

Minasny, B., T. Bandai, T. A. Ghezzehei, Y.-C. Huang, Y. Ma, A. B. McBratney, W. Ng, S. Norouzi, J. Padarian, A. Sharififar, et al. (2024). "Soil science-informed machine learning". *Geoderma* 452, 117094. DOI: 10.1016/j.geoderma.2024.117094.

Minasny, B., B. P. Malone, A. B. McBratney, D. A. Angers, D. Arrouays, A. Chambers, V. Chaplot, Z.-S. Chen, K. Cheng, B. S. Das, et al. (2017). "Soil carbon 4 per mille". *Geoderma* 292, 59–86. DOI: 10.1016/j.geoderma.2017.01.002.

Minasny, B. and A. B. McBratney (2016). "Digital soil mapping: A brief history and some lessons". *Geoderma* 264, 301–311. DOI: 10.1016/j.geoderma.2015.07.017.

Minasny, B., A. B. McBratney, B. P. Malone, and I. Wheeler (2013). "Digital mapping of soil carbon". *Advances in agronomy* 118, 1–47. DOI: 10.1016/B978-0-12-405942-9.00001-3.

Moinet, G. Y., R. Hijbeek, D. P. van Vuuren, and K. E. Giller (2023). "Carbon for soils, not soils for carbon". *Global Change Biology* 29.9, 2384–2398. DOI: 10.1111/gcb.16570.

Montero, D., C. Aybar, M. D. Mahecha, F. Martinuzzi, M. Söchting, and S. Wieneke (2023). "A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research". *Scientific Data* 10.1, 197. DOI: 10.1038/s41597-023-02096-0.

Mu, H., X. Li, Y. Wen, J. Huang, P. Du, W. Su, S. Miao, and M. Geng (2022). "A global record of annual terrestrial Human Footprint dataset from 2000 to 2018". *Scientific Data* 9.1, 176. DOI: 10.1038/s41597-022-01284-8.

Mulder, V., S. De Bruin, M. E. Schaepman, and T. Mayr (2011). "The use of remote sensing in soil and terrain mapping—A review". *Geoderma* 162.1-2, 1–19. DOI: 10.1016/j.geoderma.2010.12.018.

Myneni, R. and D. Williams (1994). "On the relationship between FAPAR and NDVI". *Remote Sensing of Environment* 49.3, 200–211. DOI: 10.1016/0034-4257(94)90016-7.

Mzid, N., S. Pignatti, W. Huang, and R. Casa (2021). "An analysis of bare soil occurrence in arable croplands for remote sensing topsoil applications". *Remote Sensing* 13.3, 474. DOI: 10.3390/rs13030474.

National Academies of Sciences, Engineering, and Medicine (2021). *Exploring a dynamic soil information system: proceedings of a workshop*. Washington, DC: The National Academies Press, 93. DOI: 10.17226/26170.

Nauman, T. W. and M. C. Duniway (2019). "Relative prediction intervals reveal larger uncertainty in 3D approaches to predictive digital soil mapping of soil properties with legacy data". *Geoderma* 347, 170–184. DOI: 10.1016/j.geoderma.2019.03.037.

Nawaz, M. F., G. Bourrie, and F. Trolard (2013). "Soil compaction impact and modelling. A review". *Agronomy for sustainable development* 33.2, 291–309. DOI: 10.1007/s13593-011-0071-8.

Nazari, S., M. Rostaminia, S. Ayoubi, A. Rahmani, and S. R. Mousavi (2020). "Efficiency of Different Feature Selection Methods in Digital Mapping of Subgroup and Soil Family Classes with Data Mining Algorithms". *Water and Soil* 34.4, 973–987. DOI: 10.22067/jsw.v34i4.85748.

Nenkam, A. M., A. M.-C. Wadoux, B. Minasny, A. B. McBratney, P. C. Traore, G. N. Falconnier, and A. M. Whitbread (2022). "Using homosoils for quantitative extrapolation of soil mapping models". *European Journal of Soil Science* 73.5, e13285. DOI: 10.1111/ejss.13285.

Ng, W., B. Minasny, M. Montazerolghaem, J. Padarian, R. Ferguson, S. Bailey, and A. B. McBratney (2019). "Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra". *Geoderma* 352, 251–267. DOI: 10.1016/j.geoderma.2019.06.016.

Nguyen, L. H., D. R. Joshi, D. E. Clay, and G. M. Henebry (2020). "Characterizing land cover/land use from multiple years of Landsat and MODIS time series: A novel approach using land surface phenology modeling and random forest classifier". *Remote sensing of environment* 238, 111017. DOI: 10.1016/j.rse.2018.12.016.

Nguyen, T. T., T. D. Pham, C. T. Nguyen, J. Delfos, R. Archibald, K. B. Dang, N. B. Hoang, W. Guo, and H. H. Ngo (2022). "A novel intelligence approach based active and ensemble learning for agricultural soil organic carbon prediction using multispectral and SAR data fusion". *Science of The Total Environment* 804, 150187. DOI: 10.1016/j.scitotenv.2021.150187.

Nussbaum, M., S. Zimmermann, L. Walthert, and A. Baltensweiler (2023). "Benefits of hierarchical predictions for digital soil mapping—An approach to map bimodal soil pH". *Geoderma* 437, 116579. DOI: 10.1016/j.geoderma.2023.116579.

Obuchowicz, C., C. Poussin, and G. Giuliani (2023). "Change in observed long-term greening across Switzerland–evidence from a three decades NDVI time-series and its relationship with climate and land cover factors". *Big Earth Data*, 1–32. DOI: 10.1080/20964471.2023.2268322.

Odgers, N. P., A. B. McBratney, and B. Minasny (2015). "Digital soil property mapping and uncertainty estimation using soil class probability rasters". *Geoderma* 237, 190–198. DOI: 10.1016/j.geoderma.2014.09.009.

Openshaw, S. and P. Taylor (1979). "A million or so correlated coefficients: three experiment on the modifiable areal unit problem". *Statistical applications in the spatial sciences*.

Orgiazzi, A., C. Ballabio, P. Panagos, A. Jones, and O. Fernández-Ugalde (2018). "LUCAS Soil, the largest expandable soil dataset for Europe: a review". *European Journal of Soil Science* 69.1, 140–153. DOI: 10.1111/ejss.12499.

Osborne, J. (2010). "Improving your data transformations: Applying the Box-Cox transformation". *Practical Assessment, Research, and Evaluation* 15.1. DOI: 10.7275/qbpc-gk17.

Oshiro, T. M., P. S. Perez, and J. A. Baranauskas (2012). "How many trees in a random forest?" In: *Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings 8*. Springer, 154–168. DOI: 10.1007/978-3-642-31537-4_13.

Pacini, L., F. Yunta, A. Jones, L. Montanarella, P. Barrè, S. Saia, S. Chen, and C. Schillaci (2023). "Fine earth soil bulk density at 0.2 m depth from Land Use and Coverage Area Frame Survey (LUCAS) soil 2018". *European Journal of Soil Science* 74.4, e13391. DOI: 10.1111/ejss.13391.

Padarian, J. and A. B. McBratney (2023). "QuadMap: Variable resolution maps to better represent spatial uncertainty". *Computers & Geosciences* 181, 105480. DOI: 10.1016/j.cageo.2023.105480.

Padarian, J., A. B. McBratney, and B. Minasny (2020). "Game theory interpretation of digital soil mapping convolutional neural networks". *SOIL Discussions* 2020, 1–12. DOI: 10.5194/soil-6-389-2020.

Padarian, J., B. Minasny, and A. B. McBratney (2019). "Using deep learning for digital soil mapping". *Soil* 5.1, 79–89. DOI: 10.5194/soil-5-79-2019.

Pagliai, M. and M. De Nobili (1993). "Relationships between soil porosity, root development and soil enzyme activity in cultivated soils". *Geoderma* 56.1-4, 243–256. DOI: 10.1016/B978-0-444-81490-6.50024-8.

Päivänen, J. (1969). "The bulk density of peat and its determination". *Silva Fennica* 3.1. DOI: 10.14214/sf.a14569.

Panagos, P., P. Borrelli, A. Jones, and D. A. Robinson (2024a). "A 1 billion euro mission: A Soil Deal for Europe". *European Journal of Soil Science* 75.1, e13466. DOI: 10.1111/ejss.13466.

Panagos, P., D. De Rosa, L. Liakos, M. Labouyrie, P. Borrelli, and C. Ballabio (2024b). "Soil bulk density assessment in Europe". *Agriculture, Ecosystems & Environment* 364, 108907. DOI: 10.1016/j.agee.2024.108907.

Panagos, P., R. Hiederer, M. Van Liedekerke, and F. Bampa (2013). "Estimating soil organic carbon in Europe based on data collected through an European network". *Ecological Indicators* 24, 439–450. DOI: 10.1016/j.ecolind.2012.07.020.

Panagos, P., M. Van Liedekerke, P. Borrelli, J. Köninger, C. Ballabio, A. Orgiazzi, E. Lugato, L. Liakos, J. Hervas, A. Jones, et al. (2022). "European Soil Data Centre 2.0: Soil data and knowledge in support of the EU policies". *European Journal of Soil Science* 73.6, e13315. DOI: 10.1111/ejss.13315.

Parente, L., R. Simoes, and T. Hengl (2023). *Monthly aggregated Water Vapor MODIS MCD19A2 (1 km): Monthly time-series (2018-2020)*. Zenodo. DOI: 10.5281/zenodo.8193736.

Patel, J. H. and M. P. Oza (2014). "Deriving crop calendar using NDVI time-series". *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 40, 869–873. DOI: 10.5194/isprsarchives-XL-8-869-2014.

Paul, C., B. Bartkowski, C. Dönmez, A. Don, S. Mayer, M. Steffens, S. Weigl, M. Wiesmeier, A. Wolf, and K. Helming (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.

Paustian, K., S. Collier, J. Baldock, R. Burgess, J. Creque, M. DeLonge, J. Dungait, B. Ellert, S. Frank, T. Goddard, et al. (2019). "Quantifying carbon for agricultural soil management: from the current status toward a global soil information system". *Carbon Management* 10.6, 567–587. DOI: 10.1080/17583004.2019.1633231.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011). "Scikit-learn: Machine Learning in Python". *Journal of Machine Learning Research* 12, 2825–2830.

Perie, C. and R. Ouimet (2008). "Organic carbon, organic matter and bulk density relationships in boreal forest soils". *Canadian journal of soil science* 88.3, 315–325. DOI: 10.4141/CJSS06008.

Pietrzykowski, M. (2008). "Soil and plant communities development and ecological effectiveness of reclamation on a sand mine cast". *Journal of forest science* 54.12, 554–565.

Pietrzykowski, M. and W. Krzaklewski (2007). "An assessment of energy efficiency in reclamation to forest". *Ecological engineering* 30.4, 341–348. DOI: 10.1016/j.ecoleng.2007.04.003.

Piikki, K. and M. Söderström (2019). "Digital soil mapping of arable land in Sweden–Validation of performance at multiple scales". *Geoderma* 352, 342–350. DOI: 10.1016/j.geoderma.2017.10.049.

Piikki, K., J. Wetterlind, M. Söderström, and B. Stenberg (2021). "Perspectives on validation in digital soil mapping of continuous attributes—A review". *Soil Use and Management* 37.1, 7–21. DOI: 10.1111/sum.12694.

Pleijsier, L. (1989). "Variability in soil data."

Pôças, I., J. Gonçalves, B. Marcos, J. Alonso, P. Castro, and J. P. Honrado (2014). "Evaluating the fitness for use of spatial data sets to promote quality in ecological assessment and monitoring". *International Journal of Geographical Information Science* 28.11, 2356–2371. DOI: 10.1080/13658816.2014.924627.

Poeplau, C. and A. Don (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., A. Don, H. Flessa, A. Heidkamp, A. Jacobs, and R. Prietz (2020a). *Erste Bodenzustandserhebung Landwirtschaft – Kerndatensatz*. de. LZS FDM-BTi 02-04-2024 Atlas-Link-Aktualisierung 23-10-2024 FDM-BTi. Göttingen. DOI: 10.3220/DATA20200203151139.

– (2020b). *First German Agricultural Soil Inventory – Core dataset*. Göttingen: OpenAgrar. DOI: 10.3220/DATA20200203151139.

Poeplau, C., A. Don, L. Vesterdal, J. Leifeld, B. Van Wesemael, J. Schumacher, and A. Gensior (2011). "Temporal dynamics of soil organic carbon after land-use change in the temperate zone–carbon response functions as a model approach". *Global change biology* 17.7, 2415–2427. DOI: 10.1111/j.1365-2486.2011.02408.x.

Poeplau, C., A. Jacobs, A. Don, C. Vos, F. Schneider, M. Wittnebel, B. Tiemeyer, A. Heidkamp, R. Prietz, and H. Flessa (2020c). "Stocks of organic carbon in German agricultural soils—Key results of the first comprehensive inventory". *Journal of Plant Nutrition and Soil Science* 183.6, 665–681. DOI: 10.1002/jpln.202000113.

Poeplau, C., R. Prietz, and A. Don (2022). "Plot-scale variability of organic carbon in temperate agricultural soils—Implications for soil monitoring#". *Journal of Plant Nutrition and Soil Science* 185.3, 403–416. DOI: 10.1002/jpln.202100393.

Poeplau, C., C. Vos, and A. Don (2017). "Soil organic carbon stocks are systematically overestimated by misuse of the parameters bulk density and rock fragment content". *Soil* 3.1, 61–66. DOI: 10.5194/soil-3-61-2017.

Poggio, L., L. M. De Sousa, N. H. Batjes, G. B. Heuvelink, B. Kempen, E. Ribeiro, and D. Rossiter (2021). "SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty". *Soil* 7.1, 217–240. DOI: 10.5194/soil-7-217-2021.

Poggio, L. and A. Gimona (2014). "National scale 3D modelling of soil organic carbon stocks with uncertainty propagation—an example from Scotland". *Geoderma* 232, 284–299. DOI: 10.1016/j.geoderma.2014.05.004.

Potapov, P., M. C. Hansen, I. Kommareddy, A. Kommareddy, S. Turubanova, A. Pickens, B. Adusei, A. Tyukavina, and Q. Ying (2020). "Landsat analysis ready data for global land cover and land cover change mapping". *Remote Sensing* 12.3, 426. DOI: 10.3390/rs12030426.

Potapov, P., S. Turubanova, M. C. Hansen, A. Tyukavina, V. Zalles, A. Khan, X.-P. Song, A. Pickens, Q. Shen, and J. Cortez (2022). "Global maps of cropland extent and change show accelerated cropland expansion in the twenty-first century". *Nature Food* 3.1, 19–28. DOI: 10.1038/s43016-021-00429-z.

Pouladi, N., A. Gholizadeh, V. Khosravi, and L. Borůvka (2023). "Digital mapping of soil organic carbon using remote sensing data: A systematic review". *Catena* 232, 107409. DOI: 10.1016/j.catena.2023.107409.

Poussin, C., A. Massot, C. Ginzler, D. Weber, B. Chatenoux, P. Lacroix, T. Piller, L. Nguyen, and G. Giuliani (2021). "Drying conditions in Switzerland–indication from a 35-year Landsat time-series analysis of vegetation water content estimates to support SDGs". *Big Earth Data* 5.4, 445–475. DOI: 10.1080/20964471.2021.1974681.

Poussin, C., P. Timoner, B. Chatenoux, G. Giuliani, and P. Peduzzi (2023). "Improved Landsat-based snow cover mapping accuracy using a spatiotemporal NDSI and generalized linear mixed model". *Science of Remote Sensing* 7, 100078. DOI: 10.1016/j.srs.2023.100078.

Radeloff, V. C., D. P. Roy, M. A. Wulder, M. Anderson, B. Cook, C. J. Crawford, M. Friedl, F. Gao, N. Gorelick, M. Hansen, S. Healey, P. Hostert, G. Hulley, J. L. Huntington, D. M. Johnson, C. Neigh, A. Lyapustin, L. Lymburner, N. Pahlevan, J.-F. Pekel, T. A. Scambos, C. Schaaf, P. Strobl, C. E. Woodcock, H. K. Zhang, and Z. Zhu (2024). "Need and vision for global medium-resolution Landsat and Sentinel-2 data products". *Remote Sensing of Environment* 300, 113918. DOI: 10.1016/j.rse.2023.113918.

Ramesh, T., N. S. Bolan, M. B. Kirkham, H. Wijesekara, M. Kanchikerimath, C. S. Rao, S. Sandeep, J. Rinklebe, Y. S. Ok, B. U. Choudhury, H. Wang, C. Tang, X. Wang, Z. Song, and O. W. I. Freeman (2019). "Soil organic carbon dynamics: Impact of land use changes and management practices: A review". *Advances in agronomy* 156, 1–107. DOI: 10.1016/bs.agron.2019.02.001.

Ramos, T. B., A. Horta, M. C. Gonçalves, F. P. Pires, D. Duffy, and J. C. Martins (2017). "The INFOSOLO database as a first step towards the development of a soil information system in Portugal". *Catena* 158, 390–412. DOI: 10.1016/j.catena.2017.07.020.

Ranghetti, L., L. Busetto, A. Crema, M. Fasola, E. Cardarelli, and M. Boschetti (2016). "Testing estimation of water surface in Italian rice district from MODIS satellite data". *International journal of applied earth observation and geoinformation* 52, 284–295. DOI: 10.1016/j.jag.2016.06.018.

Reddy, N. M., I. A. Peerzada, M. Moonis, and O. Singh (2022). "Application of biophysical, soil, and vegetation indices to better understand forest dynamics and develop strategies for forest conservation". In: *Forest Dynamics and Conservation: Science, Innovations and Policies*. Springer, 399–415. DOI: 10.1007/978-981-19-0071-6_19.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, et al. (2019). "Deep learning and process understanding for data-driven Earth system science". *Nature* 566.7743, 195–204.

Rhyma, P., K. Norizah, O. Hamdan, I. Faridah-Hanum, and A. Zulfa (2020). "Integration of normalised different vegetation index and Soil-Adjusted Vegetation Index for mangrove

vegetation delineation". *Remote Sensing Applications: Society and Environment* 17, 100280. DOI: 10.1016/j.rsase.2019.100280.

Richer-de-Forges, A. C., D. Arrouays, M. Bardy, A. Bispo, P. Lagacherie, B. Laroche, B. Lemercier, J. Sauter, and M. Voltz (2019). "Mapping of Soils and Land-Related Environmental attributes in France: analysis of end-users' needs". *Sustainability* 11.10, 2940. DOI: 10.3390/su11102940.

Richer-de-Forges, A. C., Q. Chen, N. Baghdadi, S. Chen, C. Gomez, S. Jacquemoud, G. Martelet, V. L. Mulder, D. Urbina-Salazar, E. Vaudour, et al. (2023). "Remote sensing data for digital soil mapping in French research—a review". *Remote Sensing* 15.12, 3070. DOI: 10.3390/rs15123070.

Rizzo, R., A. M.-C. Wadoux, J. A. Demattê, B. Minasny, V. Barrón, E. Ben-Dor, N. Francos, I. Savin, R. Poppiel, N. E. Silvero, et al. (2023). "Remote sensing of the Earth's soil color in space and time". *Remote Sensing of Environment* 299, 113845. DOI: 10.1016/j.rse.2023.113845.

Robinson, D., A. Thomas, S. Reinsch, I. Lebron, C. Feeney, L. Maskell, C. Wood, F. Seaton, B. Emmett, and B. Cosby (2022). "Analytical modelling of soil porosity and bulk density across the soil organic matter and land-use continuum". *Scientific reports* 12.1, 7085. DOI: 10.1038/s41598-022-11099-7.

Robinson, N. P., B. W. Allred, W. K. Smith, M. O. Jones, A. Moreno, T. A. Erickson, D. E. Naugle, and S. W. Running (2018). "Terrestrial primary production for the conterminous United States derived from Landsat 30 m and MODIS 250 m". *Remote Sensing in Ecology and Conservation* 4.3, 264–280. DOI: 10.1002/rse2.74.

Rogge, D., A. Bauer, J. Zeidler, A. Mueller, T. Esch, and U. Heiden (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.

Rossiter, D. G. (2021). "Soil mapping today: Computer-generated predictive soil maps–their role in soil survey and land evaluation". *Agric. Dev* 44.5.

Rossiter, D. G. and L. Poggio (2025). "Representing soil landscapes from digital soil mapping products–helping the map to speak for itself". *Soil* 11.2, 849–881. DOI: 10.5194/soil-11-849-2025.

Roudier, P., O. R. Burge, S. J. Richardson, J. K. McCarthy, G. J. Grealish, and A.-G. Ausseil (2020). "National scale 3D mapping of soil pH using a data augmentation approach". *Remote Sensing* 12.18, 2872. DOI: 10.3390/rs12182872.

Rusakova, E., E. Sukhacheva, and A. E. Hartemink (2022). "Vasiliy Dokuchaev–A biographical sketch on the occasion of his 175th birthday". *Geoderma* 412, 115718. DOI: 10.1016/j.geoderma.2022.115718.

Ryan, P., N. McKenzie, D. O'Connell, A. Loughhead, P. Leppert, D. Jacquier, and L. Ashton (2000). "Integrating forest soils information across scales: spatial prediction of soil properties under Australian forests". *Forest Ecology and Management* 138.1-3, 139–157. DOI: 10.1016/S0378-1127(00)00393-5.

Safanelli, J. L., S. Chabrillat, E. Ben-Dor, and J. A. Demattê (2020). "Multispectral models from bare soil composites for mapping topsoil properties over Europe". *Remote Sensing* 12.9, 1369. DOI: 10.3390/rs12091369.

Salcedo-Sanz, S., P. Ghamisi, M. Piles, M. Werner, L. Cuadra, A. Moreno-Martínez, E. Izquierdo-Verdiguier, J. Muñoz-Marí, A. Mosavi, and G. Camps-Valls (2020). "Machine learning information fusion in Earth observation: A comprehensive review of methods, applications and data sources". *Information Fusion* 63, 256–272. DOI: 10.1016/j.inffus.2020.07.004.

Salomonson, V. V. and I. Appel (2006). "Development of the Aqua MODIS NDSI fractional snow cover algorithm and validation results". *IEEE Transactions on geoscience and remote sensing* 44.7, 1747–1756. DOI: 10.1109/TGRS.2006.876029.

Sanchez, P. A., S. Ahamed, F. Carré, A. E. Hartemink, J. Hempel, J. Huising, P. Lagacherie, A. B. McBratney, N. J. McKenzie, M. D. L. Mendonça-Santos, et al. (2009). "Digital soil map of the world". *Science* 325.5941, 680–681. DOI: 10.1126/science.1175084.

Sanderman, J., T. Hengl, G. Fiske, K. Solvik, M. F. Adame, L. Benson, J. J. Bukoski, P. Carnell, M. Cifuentes-Jara, D. Donato, C. Duncan, E. M. Eid, P. zu Ermgassen, C. J. E. Lewis, P. I. Macreadie, L. Glass, S. Gress, S. L. Jardine, T. G. Jones, E. N. Nsombo, M. M. Rahman, C. J. Sanders, M. Spalding, and E. Landis (2018). "A global map of mangrove forest soil carbon at 30 m spatial resolution". *Environmental Research Letters* 13.5, 055002. DOI: 10.1088/1748-9326/aabe1c.

Sanderman, J., T. Hengl, and G. J. Fiske (2017). "Soil carbon debt of 12,000 years of human land use". *Proceedings of the National Academy of Sciences* 114.36, 9575–9580. DOI: 10.1073/pnas.1706103114.

Sandström, E., A. Namasivayam, S. Oostdijk, N. Scherpenhuijzen, N. Debonne, and P. Verburg (2023). *Land system map for Europe*. Version V6. DOI: 10.34894/THARMK.

Schmidinger, J. and G. B. Heuvelink (2023). "Validation of uncertainty predictions in digital soil mapping". *Geoderma* 437, 116585. DOI: 10.1016/j.geoderma.2023.116585.

Schmidinger, J., S. Vogel, V. Barkov, A.-D. Pham, R. Gebbers, H. Tavakoli, J. Correa, T. R. Tavares, P. Filippi, E. J. Jones, et al. (2025). "LimeSoDa: A dataset collection for benchmarking of machine learning regressors in digital soil mapping". *Geoderma* 459, 117337. DOI: 10.1016/j.geoderma.2025.117337.

Schneider, F., C. Poeplau, and A. Don (2021). "Predicting ecosystem responses by data-driven reciprocal modelling". *Global Change Biology* 27.21, 5670–5679. DOI: 10.1111/gcb.15817.

Seifert, C. A. and D. B. Lobell (2015). "Response of double cropping suitability to climate change in the United States". *Environmental Research Letters* 10.2, 024002. DOI: 10.1088/1748-9326/10/2/024002.

Sekulić, A., M. Kilibarda, G. B. Heuvelink, M. Nikolić, and B. Bajat (2020). "Random forest spatial interpolation". *Remote Sensing* 12.10, 1687. DOI: 10.3390/rs12101687.

Sen, P. K. (1968). "Estimates of the regression coefficient based on Kendall's tau". *Journal of the American statistical association* 63.324, 1379–1389. DOI: 10.1080/01621459.1968.10480934.

Sena, N. C., G. V. Veloso, E. I. Fernandes-Filho, M. R. Francelino, and C. E. G. Schaefer (2020). "Analysis of terrain attributes in different spatial resolutions for digital soil mapping application in southeastern Brazil". *Geoderma Regional* 21, e00268. DOI: 10.1016/j.geodrs.2020.e00268.

Serrano, L. R., A. M. Ruiz, R. B. Llorente, J. J. S. Donoso, and S. M. Pérez (2022). *Mapa del carbono orgánico del suelo en España: Estimación a partir de los datos del Inventario Nacional de Erosión de Suelos*. Los contenidos de esta publicación podrán ser reutilizados citando la fuente y la fecha, en su caso, de la última actualización. Plaza de San Juan de la Cruz 10, 28003 Madrid, España: Ministerio para la Transición Ecológica y el Reto Demográfico (MITECO).

Shah, A. N., M. Tanveer, B. Shahzad, G. Yang, S. Fahad, S. Ali, M. A. Bukhari, S. A. Tung, A. Hafeez, and B. Souliyanonh (2017). "Soil compaction effects on soil health and cropproductivity: an overview". *Environmental Science and Pollution Research* 24.11, 10056–10067. DOI: 10.1007/s11356-017-8421-y.

Shamrikova, E., B. Kondratenok, E. Tumanova, E. Vanchikova, E. Lapteva, T. Zonova, E. Lu-Lyan-Min, A. Davydova, Z. Libohova, and N. Suvannang (2022). "Transferability between soil organic matter measurement methods for database harmonization". *Geoderma* 412, 115547. DOI: 10.1016/j.geoderma.2021.115547.

Shangguan, W., T. Hengl, J. Mendes de Jesus, H. Yuan, and Y. Dai (2017). "Mapping the global depth to bedrock for land surface modeling". *Journal of Advances in Modeling Earth Systems* 9.1, 65–88. DOI: 10.1002/2016MS000686.

Shapley, L. (1997). "7. A Value for n-Person Games. Contributions to the Theory of Games II (1953) 307-317." In: *Classics in Game Theory*. Ed. by H. W. Kuhn. Princeton: Princeton University Press, 69–79. DOI: 10.1515/9781400829156-012.

Sharma, P., A. Singh, C. S. Kahlon, A. S. Brar, K. K. Grover, M. Dia, and R. L. Steiner (2018). "The role of cover crops towards sustainable soil health and agriculture—A review paper". *American Journal of Plant Sciences* 9.9, 1935–1951. DOI: 10.4236/ajps.2018.99140.

Shi, J., L. Yang, A.-X. Zhu, C. Qin, P. Liang, C. Zeng, and T. Pei (2018). "Machine-Learning Variables at Different Scales vs. Knowledge-based Variables for Mapping

Multiple Soil Properties". *Soil Science Society of America Journal* 82.3, 645–656. DOI: https://doi-org.ezproxy.library.wur.nl/10.2136/sssaj2017.11.0392.

Shimada, M. and T. Ohtaki (2010). "Generating large-scale high-quality SAR mosaic datasets: Application to PALSAR data for global monitoring". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 3.4, 637–656. DOI: JSTARS.2010.2077619.

Siebert, S., F. T. Portmann, and P. Döll (2010). "Global patterns of cropland use intensity". *Remote Sensing* 2.7, 1625–1643. DOI: 10.3390/rs2071625.

Skadell, L. E., F. Schneider, M. I. Gocke, J. Guigue, W. Amelung, S. L. Bauke, E. U. Hobley, D. Barkusky, B. Honermeier, I. Kögel-Knabner, et al. (2023). "Twenty percent of agricultural management effects on organic carbon stocks occur in subsoils–Results of ten long-term experiments". *Agriculture, Ecosystems & Environment* 356, 108619. DOI: 10.1016/j.agee.2023.108619.

Smart, S., D. Allen, J. Murphy, P. Carey, B. Emmett, B. Reynolds, I. Simpson, R. Evans, J. Skates, W. Scott, et al. (2009). "Countryside survey: wales results from 2007".

Smith, P. (2004). "How long before a change in soil organic carbon can be detected?" *Global Change Biology* 10.11, 1878–1883. DOI: 10.1111/j.1365-2486.2004.00854.x.

– (2008). "Land use change and soil organic carbon dynamics". *Nutrient Cycling in Agroecosystems* 81.2, 169–178. DOI: 10.1007/s10705-007-9138-y.

– (2012). "Soils and climate change". *Current opinion in environmental sustainability* 4.5, 539–544. DOI: 10.1016/j.cosust.2012.06.005.

Smith, P., J.-F. Soussana, D. Angers, L. Schipper, C. Chenu, D. P. Rasse, N. H. Batjes, F. Van Egmond, S. McNeill, M. Kuhnert, et al. (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* 26.1, 219–241. DOI: 10.1111/gcb.14815.

Soils Revealed (2025). *Soils Revealed: Global Soil Organic Carbon Mapping Platform*. https://soilsrevealed.org/. Accessed: 15 July 2025.

Song, L. and L. Estes (2023). "itsdm: Isolation forest-based presence-only species distribution modelling and explanation in R". *Methods in Ecology and Evolution* 14.3, 831–840. DOI: 10.1111/2041-210X.14067.

Sonmez, N. K. and B. Slater (2016). "Measuring intensity of tillage and plant residue cover using remote sensing". *European Journal of Remote Sensing* 49.1, 121–135. DOI: 10.5721/EuJRS20164907.

Soon, Y. and S. Abboud (1991). "A comparison of some methods for soil organic carbon determination". *Communications in Soil Science and Plant Analysis* 22.9-10, 943–954. DOI: 10.1080/00103629109368465.

Sothe, C., A. Gonsamo, J. Arabian, and J. Snider (2022). "Large scale mapping of soil organic carbon concentration with 3D machine learning and satellite observations". *Geoderma* 405, 115402. DOI: 10.1016/j.geoderma.2021.115402.

Sousa, G. P. B. de, H. Bellinaso, J. T. F. Rosas, D. C. de Mello, N. A. Rosin, M. T. A. Amorim, B. dos Anjos Bartsch, M. C. Cardoso, S. Mallah, M. R. Francelino, R. Falcioni, A. M. Rodrigo, and J. A. M. Demattê (2024). "Assessing soil degradation in Brazilian agriculture by a remote sensing approach to monitor bare soil frequency: impact on soil carbon". *Soil Advances* 2, 100011. DOI: 10.1016/j.soilad.2024.100011.

Stanimirova, R., K. Tarrio, K. Turlej, K. McAvoy, S. Stonebrook, K.-T. Hu, P. Arévalo, E. L. Bullock, Y. Zhang, C. E. Woodcock, P. Olofsson, Z. Zhu, C. P. Barber, C. M. Souza, S. Chen, J. A. Wang, F. Mensah, M. Calderón-Loor, M. Hadjikakou, B. A. Bryan, J. Graesser, D. L. Beyene, B. Mutasha, S. Siame, A. Siampale, and M. A. Friedl (2023). "A global land cover training dataset from 1984 to 2020". *Scientific Data* 10.1, 879.

Stern, C., R. Abernathey, J. Hamman, R. Wegener, C. Lepore, S. Harkins, and A. Merose (2022). "Pangeo Forge: Crowdsourcing Analysis-Ready, Cloud Optimized Data Production". *Frontiers in Climate* 3, 782909. DOI: 10.3389/fclim.2021.782909.

Stevens, A., B. van Wesemael, H. Bartholomeus, D. Rosillon, B. Tychon, and E. Ben-Dor (2008). "Laboratory, field and airborne spectroscopy for monitoring organic carbon content in agricultural soils". *Geoderma* 144.1-2, 395–404. DOI: 10.1016/j.geoderma.2007.12.009.

Stewart, V., W. Adams, and H. Abdulla (1970). "Quantitative pedological studies on soils derived from Silurian mudstones: II. The relationship between stone content and the apparent density of the fine earth". *Journal of Soil Science* 21.2, 248–255. DOI: 10.1111/j.1365-2389.1970.tb01174.x.

Stockmann, U., M. A. Adams, J. W. Crawford, D. J. Field, N. Henakaarchchi, M. Jenkins, B. Minasny, A. B. McBratney, V. d. R. De Courcelles, K. Singh, et al. (2013). "The knowns, known unknowns and unknowns of sequestration of soil organic carbon". *Agriculture, Ecosystems & Environment* 164, 80–99. DOI: 10.1016/j.agee.2012.10.001.

Stolf, R., Á. d. M. Thurler, O. O. S. Bacchi, and K. Reichardt (2011). "Method to estimate soil macroporosity and microporosity based on sand content and bulk density". *Revista Brasileira de Ciência do Solo* 35, 447–459. DOI: 10.1590/S0100-06832011000200014.

Storey, E. A., D. A. Stow, and J. F. O'Leary (2016). "Assessing postfire recovery of chamise chaparral using multi-temporal spectral vegetation index trajectories derived from Landsat imagery". *Remote Sensing of Environment* 183, 53–64. DOI: 10.1016/j.rse.2016.05.018.

Sun, Q., P. Zhang, X. Jiao, X. Lin, W. Duan, S. Ma, Q. Pan, L. Chen, Y. Zhang, S. You, S. Liu, J. Hao, H. Li, and D. Sun (2023). "A global estimate of monthly vegetation and soil fractions from spatio-temporally adaptive spectral mixture analysis during 2001–2022". *Earth System Science Data Discussions* 2023, 1–30. DOI: 10.5194/essd-16-1333-2024.

Szabo, S., Z. Gácsi, and B. Balazs (2016). "Specific features of NDVI, NDWI and MNDWI as reflected in land cover categories". *Acta Geographica Debrecina. Landscape & Environment Series* 10.3/4, 194. DOI: 10.21120/LE/10/3-4/13.

Szatmári, G., A. Laborczi, J. Mészáros, K. Takács, A. Benő, S. Koós, Z. Bakacsi, and L. Pásztor (2024a). "Gridded, temporally referenced spatial information on soil organic carbon for Hungary". *Scientific Data* 11.1, 1–13. DOI: 10.1038/s41597-024-04158-3.

Szatmári, G. and L. Pásztor (2019). "Comparison of various uncertainty modelling approaches based on geostatistics and machine learning algorithms". *Geoderma* 337, 1329–1340. DOI: 10.1016/j.geoderma.2018.09.008.

Szatmári, G., L. Pásztor, and G. B. Heuvelink (2021). "Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics". *Geoderma* 403, 115356. DOI: 10.1016/j.geoderma.2021.115356.

Szatmári, G., L. Pásztor, K. Takács, J. Mészáros, A. Benő, and A. Laborczi (2024b). "Space-time modelling of soil organic carbon stock change at multiple scales: Case study from Hungary". *Geoderma* 451, 117067. DOI: 10.1016/j.geoderma.2024.117067.

Szatmári, G., B. Pirkó, S. Koós, A. Laborczi, Z. Bakacsi, J. Szabó, and L. Pásztor (2019). "Spatio-temporal assessment of topsoil organic carbon stock change in Hungary". *Soil and Tillage Research* 195, 104410.

Taghizadeh-Mehrjardi, R., M. Mahdianpari, F. Mohammadimanesh, T. Behrens, N. Toomanian, T. Scholten, and K. Schmidt (2020). "Multi-task convolutional neural networks outperformed random forest for mapping soil particle size fractions in central Iran". *Geoderma* 376, 114552. DOI: 10.1016/j.geoderma.2020.114552.

Taha, A. A. and A. Hanbury (2015). "Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool". *BMC medical imaging* 15, 1–28. DOI: 10.1186/s12880-015-0068-x.

Tayebi, M., J. T. Fim Rosas, W. d. S. Mendes, R. R. Poppiel, Y. Ostovari, L. F. C. Ruiz, N. V. dos Santos, C. E. P. Cerri, S. H. G. Silva, N. Curi, N. E. Q. Silvero, and J. A. M. Demattê (2021). "Drivers of organic carbon stocks in different LULC history and along soil depth for a 30 years image time series". *Remote Sensing* 13.11, 2223. DOI: 10.3390/rs13112223.

Theil, H. (1950). "A rank-invariant method of linear and polynomial regression analysis". *Indagationes mathematicae* 12.85, 173.

Tian, X., B. Ahrens, L. Rossdeutscher, L. Alonso, and L. Parente (2026a). "Soil science-informed neural networks for soil organic carbon density modelling under scarce bulk density data". *EGUsphere* 2026, 1–22. DOI: 10.5194/egusphere-2026-229.

Tian, X., S. de Bruin, F. Schneider, M. Herold, and K. de Beurs (2026b). "Modeling Soil Organic Carbon Changes Using Signal-To-Noise Analysis: A Case Study Using European Soil Survey Datasets". *Global Change Biology* 32.4, e70813. DOI: 10.1111/gcb.70813.

Tian, X., S. de Bruin, R. Simoes, M. S. Isik, R. Minarik, Y.-F. Ho, M. Şahin, M. Herold, D. Consoli, and T. Hengl (2025a). "Spatiotemporal prediction of soil organic carbon density in Europe (2000–2022) using earth observation and machine learning". *PeerJ* 13, e19605. DOI: 10.7717/peerj.19605.

Tian, X., D. Consoli, L. Parente, H. Yufeng, and T. Hengl (2024). *Landsat-based spectral indices for pan-EU 2000-2022*. Version v20240319. Zenodo. DOI: 10.5281/zenodo.10776891.

Tian, X., D. Consoli, M. Witjes, F. Schneider, L. Parente, M. Şahin, Y.-F. Ho, R. Minařík, and T. Hengl (2025b). "Time series of Landsat-based bimonthly and annual spectral indices for continental Europe for 2000–2022". *Earth System Science Data* 17.2, 741–772. DOI: 10.5194/essd-17-741-2025.

Tian, X., Y.-F. Ho, M. Witjes, L. Parente, T. Hengl, and R. Minarik (2023). *Pan-EU Landmask: 10m Resolution Geospatial Land Coverage with Administrative Boundary details on country and regional level*. Zenodo. DOI: 10.5281/zenodo.8171861.

Tóth, G., A. Jones, L. Montanarella, C. Alewell, C. Ballabio, F. Carre, B. D. DE, R. A. Guicharnaud, C. Gardi, T. Hermann, et al. (2013). *LUCAS Topoil Survey-methodology, data and results*. Publications Office of the European Union. DOI: 10.2788/97922.

Tresson, P., M. Dumont, M. Jaeger, F. Borne, S. Boivin, L. Marie-Louise, J. François, H. Boukcim, and H. Goëau (2024). "Self-supervised learning of Vision Transformers for digital soil mapping using visual data". *Geoderma* 450, 117056. DOI: 10.1016/j.geoderma.2024.117056.

Tucker, C. J. (1979). "Red and photographic infrared linear combinations for monitoring vegetation". *Remote sensing of Environment* 8.2, 127–150. DOI: 10.1016/0034-4257(79)90013-0.

Turmel, M.-S., A. Speratti, F. Baudron, N. Verhulst, and B. Govaerts (2015). "Crop residue management and soil health: A systems analysis". *Agricultural Systems* 134, 6–16. DOI: 10.1016/j.agsy.2014.05.009.

Tziolas, N., N. Tsakiridis, U. Heiden, and B. van Wesemael (2024). "Soil organic carbon mapping utilizing convolutional neural networks and Earth observation data, a case study in Bavaria state Germany". *Geoderma* 444, 116867. DOI: 10.1016/j.geoderma.2024.116867.

Ugbaje, S. U., S. Karunaratne, T. Bishop, L. Gregory, R. Searle, K. Coelli, and M. Farrell (2024). "Space-time mapping of soil organic carbon stock and its local drivers: Potential for use in carbon accounting". *Geoderma* 441, 116771. DOI: 10.1016/j.geoderma.2023.116771.

Van Der Westhuizen, S., G. B. Heuvelink, and D. P. Hofmeyr (2023). "Multivariate random forest for digital soil mapping". *Geoderma* 431, 116365. DOI: 10.1016/j.geoderma.2023.116365.

Van Deventer, A., A. Ward, P. Gowda, and J. Lyon (1997). "Using thematic mapper data to identify contrasting soil plains and tillage practices". *Photogrammetric engineering and remote sensing* 63, 87–93.

Van Ranst, E., L. Vanmechelen, A. Thomasson, L. ResourcePartnership, A. H. Pembroke, U. Dyfed, J. Hollis, R. Jones, M. Jarnagne, and D. King (1995). "11.3. ELABORATION OF AN EXTENDED KNOWLEDGE DATABASE TO INTERPRET THE 1: 1,000,000 EU SOIL MAP FOR ENVIRONMENTAL PURPOSES". *INSTITUTE FOR REMOTE SENSING APPLICATIONS*.

Vaysse, K., G. B. Heuvelink, and P. Lagacherie (2017). "Spatial aggregation of soil property predictions in support of local land management". *Soil Use and Management* 33.2, 299–310. DOI: 10.1111/sum.12350.

Vaysse, K. and P. Lagacherie (2017). "Using quantile regression forest to estimate uncertainty of digital soil mapping products". *Geoderma* 291, 55–64. DOI: 10.1016/j.geoderma.2016.12.017.

Venter, Z. S., H.-J. Hawkins, M. D. Cramer, and A. J. Mills (2021a). "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.

– (2021b). "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.

Veronesi, F. and C. Schillaci (2019). "Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation". *Ecological Indicators* 101, 1032–1044. DOI: 10.1016/j.ecolind.2019.02.026.

Wadoux, A.-C. (2019a). "Using deep learning for multivariate mapping of soil with quantified uncertainty". *Geoderma* 351, 59–70. DOI: 10.1016/j.geoderma.2019.05.012.

Wadoux, A. M.-C., B. Minasny, and A. B. McBratney (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.-C. (2019b). "Using deep learning for multivariate mapping of soil with quantified uncertainty". *Geoderma* 351, 59–70. DOI: 10.1016/j.geoderma.2019.05.012.

Wadoux, A. M.-C. and G. B. Heuvelink (2023). "Uncertainty of spatial averages and totals of natural resource maps". *Methods in Ecology and Evolution* 14.5, 1320–1332. DOI: 10.1111/2041-210X.14106.

Wadoux, A. M.-C. and G. B. Heuvelink (2025). "Scientists yet to consider spatial correlation in assessing uncertainty of spatial averages and totals". *International Journal of Applied Earth Observation and Geoinformation* 139, 104472. DOI: 10.1016/j.jag.2025.104472.

Wadoux, A. M.-C., B. Minasny, and A. B. McBratney (2020b). "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.-C. and C. Molnar (2022). "Beyond prediction: methods for interpreting complex models of soil variation". *Geoderma* 422, 115953. DOI: 10.1016/j.geoderma.2022.115953.

Wadoux, A. M.-C., N. P. Saby, and M. P. Martin (2023). "Shapley values reveal the drivers of soil organic carbon stock prediction". *Soil* 9.1, 21–38. DOI: 10.5194/soil-9-21-2023.

Wagemann, J., S. Siemen, B. Seeger, and J. Bendix (2021). "Users of open Big Earth data–An analysis of the current state". *Computers & Geosciences* 157, 104916. DOI: 10.1016/j.cageo.2021.104916.

Wagner, W., B. Bauer-Marschallinger, C. Navacchi, F. Reuß, S. Cao, C. Reimer, M. Schramm, and C. Briese (2021). "A Sentinel-1 backscatter datacube for global land monitoring applications". *Remote Sensing* 13.22, 4622. DOI: 10.3390/rs13224622.

Walkley, A. and I. A. Black (1934). "An examination of the Degtjareff method for determining soil organic matter, and a proposed modification of the chromic acid titration method". *Soil science* 37.1, 29–38. DOI: 10.1097/00010694-193401000-00003.

Walter, C., P. Lagacherie, and S. Follain (2006). "Integrating pedological knowledge into digital soil mapping". *Developments in soil science* 31, 281–615. DOI: 10.1016/S0166-2481(06)31022-7.

Wan, Z. (2006). "MODIS land surface temperature products users' guide". *Institute for Computational Earth System Science, University of California: Santa Barbara, CA, USA* 805, 26.

Wang, B., J. M. Gray, C. M. Waters, M. R. Anwar, S. E. Orgill, A. L. Cowie, P. Feng, and D. Li Liu (2022). "Modelling and mapping soil organic carbon stocks under future climate change in south-eastern Australia". *Geoderma* 405, 115442. DOI: 10.1016/j.geoderma.2021.115442.

Webster, R. and M. A. Oliver (2007). *Geostatistics for environmental scientists*. John Wiley & Sons.

Weindorf, D. C. and S. Chakraborty (2024). "Balancing machine learning and artificial intelligence in soil science with human perspective and experience". *Pedosphere* 34.1, 9–12. DOI: 10.1016/j.pedsph.2023.09.010.

Wesemael, B. van, A. Abdelbaki, E. Ben-Dor, S. Chabrillat, P. d'Angelo, J. A. Demattê, G. Genova, A. Gholizadeh, U. Heiden, P. Karlshoefer, et al. (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.

White, M. A., P. E. Thornton, and S. W. Running (1997). "A continental phenology model for monitoring vegetation responses to interannual climatic variability". *Global biogeochemical cycles* 11.2, 217–234. DOI: 10.1029/97GB00330.

Widyastuti, M. T., B. Minasny, J. Padarian, and F. Maggi (2024). *PEATGRIDS: Mapping global peat thickness and carbon stock via digital soil mapping approach, dataset*. Version 2.0.1. Zenodo. DOI: 10.5281/zenodo.15070037.

Wiesmeier, M., L. Urbanski, E. Hobley, B. Lang, M. von Lützow, E. Marin-Spiotta, B. van Wesemael, E. Rabot, M. Ließ, N. Garcia-Franco, et al. (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.

Witjes, M., L. Parente, J. Križan, T. Hengl, and L. Antonić (2023). "Ecodatacube. eu: analysis-ready open environmental data cube for Europe". *PeerJ* 11, e15478. DOI: 10.7717/peerj.15478.

Wu, J., L. Lin, T. Li, Q. Cheng, C. Zhang, and H. Shen (2022). "Fusing Landsat 8 and Sentinel-2 data for 10-m dense time-series imagery using a degradation-term constrained deep network". *International Journal of Applied Earth Observation and Geoinformation* 108, 102738. DOI: 10.1016/j.jag.2022.102738.

Wulder, M. A., T. R. Loveland, D. P. Roy, C. J. Crawford, J. G. Masek, C. E. Woodcock, R. G. Allen, M. C. Anderson, A. S. Belward, W. B. Cohen, et al. (2019). "Current status of Landsat program, science, and applications". *Remote sensing of environment* 225, 127–147. DOI: 10.1016/j.rse.2019.02.015.

Wulder, M. A., D. P. Roy, V. C. Radeloff, T. R. Loveland, M. C. Anderson, D. M. Johnson, S. Healey, Z. Zhu, T. A. Scambos, N. Pahlevan, M. Hansen, N. Gorelick, C. J. Crawford, J. G. Masek, T. Hermosilla, J. C. White, A. S. Belward, C. Schaaf, C. E. Woodcock, J. L. Huntington, L. Lymburner, P. Hostert, F. Gao, A. Lyapustin, J.-F. Pekel, P. Strobl, and B. D. Cook (2022). "Fifty years of Landsat science and impacts". *Remote Sensing of Environment* 280, 113195. DOI: 10.1016/j.rse.2022.113195.

Xiang, X., J. Du, P.-A. Jacinthe, B. Zhao, H. Zhou, H. Liu, and K. Song (2022). "Integration of tillage indices and textural features of Sentinel-2A multispectral images for maize residue cover estimation". *Soil and Tillage Research* 221, 105405. DOI: 10.1016/j.still.2022.105405.

Xue, Z., P. Du, and L. Feng (2014). "Phenology-driven land cover classification and trend analysis based on long-term remote sensing image series". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.4, 1142–1156. DOI: 10.1109/JSTARS.2013.2294956.

Yaalon, D. (1975). "Conceptual models in pedogenesis: can soil-forming functions be solved?" *Geoderma* 14.3, 189–205. DOI: 10.1016/0016-7061(75)90001-4.

Yaalon, D. (1989). "The earliest soil maps and their logic". *Bulletin of the International Society of Soil Science.–International Society of Soil Science, Wageningen p* 24.

Yang, L., Y. Cai, L. Zhang, M. Guo, A. Li, and C. Zhou (2021). "A deep learning method to predict soil organic carbon content at a regional scale using satellite-based phenology variables". *International Journal of Applied Earth Observation and Geoinformation* 102, 102428. DOI: 10.1016/j.jag.2021.102428.

Yang, L., X. He, F. Shen, C. Zhou, A.-X. Zhu, B. Gao, Z. Chen, and M. Li (2020a). "Improving prediction of soil organic carbon content in croplands using phenological parameters extracted from NDVI time series data". *Soil and Tillage Research* 196, 104465. DOI: 10.1016/j.still.2019.104465.

Yang, R.-M., L.-A. Liu, X. Zhang, R.-X. He, C.-M. Zhu, Z.-Q. Zhang, and J.-G. Li (2022). "The effectiveness of digital soil mapping with temporal variables in modeling soil organic carbon changes". *Geoderma* 405, 115407. DOI: 10.1016/j.geoderma.2021.115407.

Yang, T., K. H. Siddique, and K. Liu (2020b). "Cropping systems in agriculture and their impact on soil health-A review". *Global Ecology and Conservation* 23, e01118. DOI: 10.1016/j.gecco.2020.e01118.

Yang, X., J. Blower, L. Bastin, V. Lush, A. Zabala, J. Masó, D. Cornford, P. Díaz, and J. Lumsden (2013). "An integrated view of data quality in Earth observation". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371.1983, 20120072. DOI: 10.1098/rsta.2012.0072.

Yang, Y., J. Fang, Y. Tang, C. Ji, C. Zheng, J. He, and B. Zhu (2008). "Storage, patterns and controls of soil organic carbon in the Tibetan grasslands". *Global change biology* 14.7, 1592–1599. DOI: 10.1111/j.1365-2486.2008.01591.x.

Yigini, Y. and P. Panagos (2016). "Assessment of soil organic carbon stocks under future climate and land cover changes in Europe". *Science of the Total Environment* 557, 838–850. DOI: 10.1016/j.scitotenv.2016.03.085.

Yuzugullu, O., N. Fajraoui, A. Don, and F. Liebisch (2024). "Satellite-based soil organic carbon mapping on European soils using available datasets and support sampling". *Science of Remote Sensing* 9, 100118. DOI: 10.1016/j.srs.2024.100118.

Zepp, S., U. Heiden, M. Bachmann, M. Möller, M. Wiesmeier, and B. van Wesemael (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: 10.3390/rs13030474.

Zeraatpisheh, M., Y. Garosi, H. R. Owliaie, S. Ayoubi, R. Taghizadeh-Mehrjardi, T. Scholten, and M. Xu (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.

Zhang, L., Y. Cai, H. Huang, A. Li, L. Yang, and C. Zhou (2022). "A CNN-LSTM model for soil organic carbon content prediction with long time series of MODIS-based phenological variables". *Remote sensing* 14.18, 4441. DOI: 10.3390/rs14184441.

Zhang, L., G. B. Heuvelink, V. L. Mulder, S. Chen, X. Deng, and L. Yang (2024a). "Using process-oriented model output to enhance machine learning-based soil organic carbon prediction in space and time". *Science of the Total Environment* 922, 170778. DOI: 10.1016/j.scitotenv.2024.170778.

Zhang, T., L.-M. Huang, and R.-M. Yang (2024b). "Evaluation of digital soil mapping projection in soil organic carbon change modeling". *Ecological Informatics* 79, 102394. DOI: 10.1016/j.ecoinf.2023.102394.

Zhang, X., S. Chen, J. Xue, N. Wang, Y. Xiao, Q. Chen, Y. Hong, Y. Zhou, H. Teng, B. Hu, et al. (2023a). "Improving model parsimony and accuracy by modified greedy feature selection in digital soil mapping". *Geoderma* 432, 116383. DOI: 10.1016/j.geoderma.2023.116383.

Zhang, X., L. Liu, X. Chen, Y. Gao, S. Xie, and J. Mi (2020). "GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery". *Earth System Science Data Discussions* 2020, 1–31. DOI: 10.5194/essd-13-2753-2021.

Zhang, X., L. Liu, Y. Liu, S. Jayavelu, J. Wang, M. Moon, G. M. Henebry, M. A. Friedl, and C. B. Schaaf (2018). "Generation and evaluation of the VIIRS land surface phenology product". *Remote Sensing of Environment* 216, 212–229. DOI: 10.1016/j.rse.2018.06.047.

Zhang, X., E. Xie, J. Chen, Y. Peng, G. Yan, and Y. Zhao (2023b). "Modelling the spatiotemporal dynamics of cropland soil organic carbon by integrating process-based models differing in structures with machine learning". *Journal of Soils and Sediments* 23.7, 2816–2831. DOI: 10.1007/s11368-023-03516-9.

Zhang, Z., J. Ding, C. Zhu, J. Wang, X. Ge, X. Li, L. Han, X. Chen, and J. Wang (2023c). "Historical and future variation of soil organic carbon in China". *Geoderma* 436, 116557. DOI: 10.1016/j.geoderma.2023.116557.

Zheng, B., J. Campbell, G. Serbin, and C. Daughtry (2013). "Multitemporal remote sensing of crop residue cover and tillage practices: A validation of the minNDTI strategy in the United States". *Journal of Soil and Water Conservation* 68.2, 120–131. DOI: 10.2489/jswc.68.2.120.

Zheng, B., J. B. Campbell, and K. M. de Beurs (2012). "Remote sensing of crop residue cover using multi-temporal Landsat imagery". *Remote Sensing of Environment* 117, 177–183. DOI: 10.1016/j.rse.2011.09.016.

Zhou, J., L. Jia, M. Menenti, and B. Gorte (2016). "On the performance of remote sensing time series reconstruction methods–A spatial comparison". *Remote Sensing of Environment* 187, 367–384. DOI: 10.1016/j.rse.2016.10.025.

Zhou, T., Y. Geng, C. Ji, X. Xu, H. Wang, J. Pan, J. Bumberger, D. Haase, and A. Lausch (2021). "Prediction of soil organic carbon and the C: N ratio on a national scale using machine learning and satellite data: A comparison between Sentinel-2, Sentinel-3 and Landsat-8 images". *Science of the Total Environment* 755, 142661. DOI: 10.1016/j.scitotenv.2020.142661.

Ziche, D., E. Grüneberg, W. Riek, and N. Wellbrock (2022). "Comparison of the LUCAS 2015 inventory with the second National Forest Soil Inventory". *Thünen Report*. DOI: 10.3220/REP1651759791000.

Acknowledgements

This PhD has not made my life better in the way I once expected, but it has made me live more. And I cherish it especially for the people and moments that made the journey vivid, textured, and alive.

I was lucky to be guided by a supervisory team remarkable both in size and in substance. My promotor, Sytze, was there from beginning to end. However difficult things became, it always helped to know that I would soon talk to him. That alone was often enough to keep me going. With him, I never felt afraid to say that I was confused, to test a new idea, or to follow an intuition I could not yet defend. Without his guidance, I would not be here.

Kirsten has been my role model. No matter how busy she is, whenever I need help, she responds promptly, precisely, and concisely, often with a spark of humor and wisdom. Although I did not get to work closely with Martin very often, he consistently offered the kind of perspective and encouragement that helped me keep moving forward.

As someone who started with no soil science background at all, I owe a great deal to Florian and Bernhard. They patiently explained even the simplest concepts, and why some of my maps made no sense whatsoever. I am grateful not only for their scientific guidance, but also for welcoming me into their institutes during my research visits, and for sharing with me their affection for their work, cities, and food. The two weeks I spent in Braunschweig and Jena remain among the warmest memories.

Once upon a time at OpenGeoHub, the lunch talks, work travels, board game nights, and midnight drives truly lit up my daily PhD life. My three fellow PhDs at OpenGeoHub, Carmelo, Martijn, and Julia, were my closest companions. We cooked, talked, worked, and joked our way through this period together. Carmelo and Martijn were not only colleagues and friends, but often half-advisors too; because of them, I avoided many unnecessary detours. Julia had the wisdom to know herself, the courage to choose accordingly, and the resolve to act on it. Now, at the point of graduating, I feel I may finally be ready to do the same.

Before ChatGPT, Leandro was my go-to person for programming. From Davide, I learned life lessons ranging from how to manage Git tokens, how to deal with sore arms after

climbing to whether one should quit a PhD. Murat gave me several masterclasses in geo-visualization, as well as crash courses in Turkish tea culture. Rolf is a teacher with great patience and logic, who once tried to teach me software engineering in one afternoon (the failure was entirely my fault). Traveling to Italy with Robert was always a joy; he generously shared his knowledge of wine, food, and why we should focus on ourselves rather than be consumed by pressure. Yufeng tried very hard to make us love nature; I am not sure he succeeded, but he did make me respect its force. Nathalie never dismissed a question with a quick answer, but instead made time to help me understand how things actually worked. Lea's way of facing uncertainty and stepping beyond herself encouraged me greatly. Bea has always been there, empowering me in countless ways. I sometimes think she is a superhero, but in fact, she is even better. With a rare sensitivity to language, Viola always seemed able to capture the absurdity of a moment with surgical precision and humor. Gulten often said the thing I had not yet managed to say, with the energy, clarity, and a deeply engaged way of moving through life. Wherever I call for help, Serkan always manages to put out the fire. Wibi's optimism always made its way to the people around him, and I remember especially fondly the days we spent drifting through Brussels together. When I was once stuck in the office late at night, Mateo stayed online with me past midnight, a calm and steady presence when I needed one most. Although I did not work long with Giulio and Octavian, they supported my thesis from their perspective as excellent soil scientists and reliable collaborators. Fee's perspectives on social justice and his strong sense of action have also influenced me greatly. I am also grateful to Qin and Dong, whose support meant a great deal to me in the short time we worked together. Among them, I am especially grateful to Tianyu and Teacher Gao for all the drives, conversations, and support-day and night, on weekdays and weekends.

I was also fortunate to belong to the GRS community. Although I was not always on campus, it gave me a strong sense of rhythm, companionship, and belonging. Hannah is, in many ways, my connection to the GRS group and to the life of the flexroom. I still remember us cycling 20 km together after the run at our absolute most exhausted, as well as all the times she let me drag her to Eastern Express for lunch. Since Gaia stopped providing cups, Yang kept one for me so I would not dehydrate in the flexroom; we ended up sharing not only a cup, but also life and research, joys and worries. Dainius shared his knowledge, experience, and writing sessions generously, and always in a very inspiring way. I also thank Robert for always greeting me with warmth, and for generously making time to help me with questions when I was still finding my footing. Thank you, Dainius and Arno, for giving me the opportunity to be a teaching assistant in Geoscripting, from which I also learned a great deal. Thanks also to Nandika and Mirjam for pulling me into Veluweloop, which gave me the chance to run together with GRS.

I am also grateful to the PE&RC Graduate School for organizing events that connected PhD students not only to knowledge and research, but also to one another, and for always being there when we needed support.

Outside work and academia, my friends carried me more than they probably know. Xiaowei, Ying, and Yue listened to me over many meals and helped me take my troubles apart until they no longer seemed so formidable. Yitong and Maoran have been by my side since my master's, and I hope our movie dates will continue without end. Manze, your home has felt like a second home to me in the Netherlands, and your seriousness about both life and research has always stayed with me. Fanhao and Yuqi, we came out of the same master's track during the pandemic and seemed to meet the same obstacles at the same moments; somehow, we kept going. Meng and Jiaqi, far from home, we shared trust, support, and the comfort of hot dumplings in a foreign country. My childhood friends, Xuesong, Guannan, Yujie, Bingxu, and Jia, thank you for all the conversations across time zones. Lixia, thank you for believing so steadily in my ability and resilience through so many stages of my PhD. Ou and Jingwen, thank you for helping me in so many practical and generous ways since I first arrived in the Netherlands.

Last, and most of all, my thanks go to my family. Their support has been unconditional, unquestioning, and beyond measure. My grandparents, parents, aunt and uncle, and cousin are, simply, the best family I could have had. And Elvis, Potchi, and Pidan, you held our life together and took care of our home without ever being asked while I was commuting, traveling, and disappearing into my work.

I had the strange good fortune of finishing this degree alongside the rise of AI, with ChatGPT being born just two months after I started my PhD. Now, I do not know where either I or this era is heading. But I think it is time to step into the next stage.

About the author



Xuemeng was born and raised in a small town in northern China. As the youngest child in her family and a member of a close-knit community, she was always surrounded by love and companionship. She spent many summer holidays traveling to other parts of the country because of her father's work. Perhaps that was where she first learned to keep moving. In the end, the journey brought her to the other end of the continent, only with fewer mountains.

She studied Remote Sensing track in Ocean Detection Technology for her bachelor's degree in Qingdao, a beautiful coastal city with seafood she still misses. During her master's, she was supervised by Prof. Susan Steele-Dunne, who strongly encouraged her curiosity and gave shape to her idea of pursuing academia. By a turn of fate, she then began her PhD at Wageningen University and OpenGeoHub, where she met many people who changed her life in one way or another. She is now exploring what might come next.

Peer-reviewed Journal Publications

Tian, **X.**, D. Consoli, M. Witjes, F. Schneider, L. Parente, M. Şahin, Y.-F. Ho, R. Minařík, and T. Hengl (2025a). "Time series of Landsat-based bimonthly and annual spectral indices for continental Europe for 2000–2022". *Earth System Science Data* 17.2, 741–772. DOI: 10.5194/essd-17-741-2025.

Tian, **X.**, S. de Bruin, R. Simoes, M. S. Isik, R. Minarik, Y.-F. Ho, M. Şahin, M. Herold, D. Consoli, and T. Hengl (2025b). "Spatiotemporal prediction of soil organic carbon density in Europe (2000–2022) using earth observation and machine learning". *PeerJ* 13, e19605. DOI: 10.7717/peerj.19605.

Tian, **X.**, S. de Bruin, F. Schneider, M. Herold, and K. de Beurs (2026a). "Modeling Soil Organic Carbon Changes Using Signal-To-Noise Analysis: A Case Study Using European Soil Survey Datasets". *Global Change Biology* 32.4, e70813. DOI: 10.1111/gcb.70813.

Consoli, D., L. Parente, R. Simoes, M. Şahin, **X. Tian**, M. Witjes, L. Sloat, and T. Hengl (2024). "A computational framework for processing time-series of earth observation data based on discrete convolution: global-scale historical Landsat cloud-free aggregates at 30 m spatial resolution". *PeerJ* 12, e18585. DOI: 10.7717/peerj.18585.

Hackländer, J., L. Parente, Y.-F. Ho, T. Hengl, R. Simoes, D. Consoli, M. Şahin, **X. Tian**, M. Jung, M. Herold, G. Duveiller, M. Weynants, and I. Wheeler (2024). "Land potential assessment and trend-analysis using 2000–2021 FAPAR monthly time-series at 250 m spatial resolution". *PeerJ* 12, e16972. DOI: 10.7717/peerj.16972.

Other Scientific Publications

Tian, **X.**, B. Ahrens, L. Rossdeutscher, L. Alonso, and L. Parente (2026b). "Soil science-informed neural networks for soil organic carbon density modelling under scarce bulk density data". In: *EGUsphere*.

Overview of completed training activities

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 46 ECTS (≈ 30 weeks of activities)



Review/project proposal (4.5 ECTS)

- Spatial-temporal modeling of soil organic carbon at multiscale

Post-graduate courses (9.3 ECTS)

- Parallel programming in python; eScience center (2023)
- Summer school; OpenGeoHub (2023, 2025)
- Uncertainty Analysis and Statistical Validation of Spatial Environmental Models; PE&RC (2024)
- Resilience of living system; PE&RC (2025)
- Bayesian Statistics; PE&RC (2025)
- Geostatistics; PE&RC (2025)

Deficiency, refresh, brush-up courses (2 ECTS)

- Cornell soil health training; Cornell University (2024)

Laboratory training and working visits (3 ECTS)

- Challenges in soil organic carbon dynamics monitoring, on both sampling and modeling; Johann Heinrich von Thünen Institute (2025)

- Hybrid modelling: leverage the advantages of machine learning and process-based model; Max Planck Institute Biogeochemistry (2025)

Invited review of journal manuscripts (2 ECTS)

- Statistical Atlas of European Agriculture (Earth System Science Data, 2025)
- Drivers shaping the global pattern of soil organic carbon (Earth's Future, 2025)

Competence, skills and career-oriented activities (2.5 ECTS)

- Project and time management; WGS (2023)
- The Choice: Un-box your PhD process and take charge of your performance; WGS (2025)
- Career assessment; WGS (2026)

Scientific Integrity/Ethics in science activities (0.6 ECTS)

- Scientific Integrity; WGS (2023)

PE&RC Annual meetings, seminars and PE&RC weekend/retreat (1.5 ECTS)

- PE&RC First year's retreat; PE&RC (2022)
- PE&RC Midterm retreat; PE&RC (2025)

Discussion groups/local seminars or scientific meetings (4.5 ECTS)

- Digital Twins for Earth Sciences; SURF (2022)
- Open Geodata symposium; ITC (2023)
- WUR soil day; WUR (2025)
- EuroGEO workshop; Netherlands Space Office (2025)
- European Research and Innovation Days; European Commission (2025)
- OpenEarthMonitor seminars; OEMC (2022–2025)

International symposia, workshops and conferences (13.9 ECTS)

- EO 4 SOIL protection; ESA-ESRIN (2024)
- Global Symposium on Soil Information and Data 2024; FAO and Chinese Academy of Science (2024)
- 2nd EUSO young soil scientist; JRC (2024)
- Open Earth Monitor Global Workshop; OpenGeoHub and IIASA (2024)
- Soil Health Now!; OpenGeoHub (2025)
- EGU General Assembly; EGU (2025)
- Eurosoil; European Confederation of Soil Science Societies (2025)

Lecturing/supervision of practicals/tutorials (3 ECTS)

- Geoscripting (2025)

Authorship statement

Chapter 1 The overarching research question and its scientific and societal context were proposed by my promotors and co-promotors. I further delineated the research questions, positioned the work within the existing scientific literature. I wrote the first draft of the chapter and revised it based on comments from my promotor. I used ChatGPT for language support to improve clarity and readability. I am fully responsible for the final content and interpretation of the research.

Chapter 2 The research question and methodology were initially defined by my co-promotor, Tom. Together with my co-authors, I refined the research questions and experimental design. I carried out the experiments and performed the data analysis and visualization in collaboration with my co-authors. I wrote the first draft of the manuscript. My co-authors, particularly my promotors, provided comments on the draft, which I subsequently addressed. I used ChatGPT for language support to enhance clarity. I am fully responsible for the final content and interpretation of the research.

Chapter 3 The research question and methodology were initially defined by my co-promotor, Tom. During the course of the study, I further refined the experimental design and research questions in close consultation with my promotor, Sytze. I carried out the experiments and conducted the data analysis and visualization together with my co-authors. I wrote the first draft of the manuscript and revised it based on comments from my co-authors. I used ChatGPT for language support to enhance clarity. I am fully responsible for the final content and interpretation of the research.

Chapter 4 All co-authors contributed to the development of the research question. I proposed the methodology and carried out the experiments. With contributions from two of my co-authors, I conducted the data analysis, visualization, and interpretation. I wrote the first draft of the manuscript and revised it based on comments from my co-authors. I used ChatGPT for language support to enhance clarity. I am fully responsible for the final content and interpretation of the research.

Chapter 5 My advisor and co-author, Bernhard, proposed the research question, methodology, and experimental design. With contributions from all co-authors, I conducted the experiments and performed the data analysis. One co-author contributed to the visualization of the results. I wrote the first draft of the manuscript and revised it based on comments from my co-authors. I used ChatGPT for language support to enhance clarity. I am fully responsible for the final content and interpretation of the research.

Chapter 6 I wrote the first draft of the chapter, incorporating suggestions from my promotors and co-promotors regarding the topics and arguments to be included. I revised the text based on comments from my promotors. I used ChatGPT for language support to enhance clarity and readability.

This research received funding from the European Union under the Horizon Europe programme through "AI4SoilHealth" (Grant Agreement No. 101086179), "Open-Earth-Monitor Cyberinfrastructure" (Grant Agreement No. 101059548) and "Intergenerational Open Geospatial Carbon Registry" (Grant Agreement No. 101218854). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Cover design by Viola Coppi, Beatriz Sarabia and Xuemeng Tian.

Printed by ProefschriftMaken.

