International Journal of Plant & Soil Science

Volume 38, Issue 10, Page 55-81, 2026; Article no.IJPSS.166542
ISSN: 2320-7035

Digital Monitoring, Reporting and Verification of Soil Carbon in Carbon Farming: A Critical Review of Integrating Remote Sensing, Proximal Sensing, Artificial Intelligence and Process-Based Models

Aditya V. Machnoor [a++*], K. G. Rosin [a#], Rakesh Kokatnoor [a†] and Arun S. Kalasad [a†]

[a] *Water Technology Center, ICAR-IARI, Pusa, New Delhi, India.*

Authors' contributions

This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/ijpss/2026/v38i106331

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://pr.sdiarticle5.com/review-history/166542

Review Article

Received: 24/07/2026
Accepted: 16/09/2026
Published: 17/09/2026

Abstract

Carbon farming schemes increasingly pay land managers for increases in soil organic carbon, yet the credibility of these payments depends on monitoring, reporting and verification systems that can quantify small changes in a large and spatially variable carbon pool. Digital monitoring, reporting and verification promises to lower costs by combining satellite and airborne remote sensing, proximal soil sensing, artificial intelligence and process-based biogeochemical models, but the evidence that such combinations yield unbiased, verifiable estimates of soil organic carbon change remains uneven. This critical narrative review evaluates what each evidence stream can and cannot contribute to credible soil carbon accounting, how their integration propagates or reduces error, and which claims are established rather than preliminary. Literature was identified through structured searches of multidisciplinary

---

[++] *Ph.D. Scholar;* [#] *Scientist;* [†] *M.Sc. Scholar*
**Corresponding author: E-mail: adimachnoor07@gmail.com;*

Cite as: *Machnoor, A. V., Rosin, K. G., Kokatnoor, R., & Kalasad, A. S. (2026). Digital Monitoring, Reporting and Verification of Soil Carbon in Carbon Farming: A Critical Review of Integrating Remote Sensing, Proximal Sensing, Artificial Intelligence and Process-Based Models. International Journal of Plant & Soil Science, 38(10), 55–81. https://doi.org/10.9734/ijpss/2026/v38i106331*

scholarly indexes, backward and forward citation tracing and examination of institutional methodologies, and was appraised for design, validation strategy, uncertainty treatment and transferability. The synthesis indicates that direct measurement remains the only independent reference for soil organic carbon change, but that field-level detection is statistically weak unless sampling is dense, paired and replicated across many fields. Remote sensing is most reliable for documenting management practices and crop biomass rather than for retrieving topsoil carbon change, which is confounded by moisture, residue cover and roughness and is restricted to the surface layer. Proximal spectroscopy can reduce analytical costs substantially, provided that calibration transfer, bulk density and measurement error are handled explicitly. Machine learning improves spatial prediction but can overstate accuracy when validation ignores spatial dependence and extrapolation. Calibrated process-based models remain the principal engine for estimating change at scale, although independent time-series validation is scarce. Hybrid architectures that assimilate observations into mechanistic models are promising but have been tested mainly in a few temperate regions. Credibility therefore depends less on any single technology than on design-based verification, counterfactual baselines, transparent uncertainty propagation and explicit validity domains. Priorities include randomised multi-field measure-and-remeasure networks, open benchmark datasets, standardised uncertainty reporting and evaluation in smallholder and tropical systems.

Keywords: Soil organic carbon; carbon credits; measurement, reporting and verification; earth observation; soil spectroscopy; machine learning; biogeochemical modelling; uncertainty quantification.

1. Introduction

Soil organic carbon (SOC) is the largest actively cycling terrestrial carbon pool, and its management has become central to strategies for land-based climate change mitigation. Soil carbon has been estimated to represent about one quarter of the mitigation potential of natural climate solutions, with the larger share of that potential attributed to rebuilding depleted stocks rather than protecting existing ones (Bossio *et al.*, 2020). The launch of the "4 per mille Soils for Food Security and Climate" initiative at the twenty-first Conference of the Parties in 2015 crystallised the expectation that agricultural soils could offset a meaningful fraction of anthropogenic emissions (Minasny *et al.*, 2017). These expectations have been tempered by analyses showing that saturation, nutrient limitation, reversibility and inconsistent yield benefits substantially reduce the realistic contribution of soil carbon to mitigation (Moinet *et al.*, 2023; Schlesinger, 2022). Terminological confusion has compounded the problem: in an examination of 100 recent soil carbon papers, only 4% used the term carbon sequestration correctly, and many conflated stocks with removals or loss mitigation with net atmospheric withdrawal (Don *et al.*, 2024).

Carbon farming, understood here as the deliberate adoption of land management practices intended to increase carbon storage or reduce emissions from soils and biomass, has nonetheless moved rapidly from concept to market instrument. Private certificate providers and voluntary carbon standards issue tradable credits for SOC increases (Paul *et al.*, 2023), and the European Union has adopted a certification framework that covers carbon farming activities of at least five years' duration and requires quantification, additionality, long-term storage and sustainability to be demonstrated (European Parliament & Council of the European Union, 2024). Every such scheme rests on a system of monitoring, reporting and verification (MRV) that must convert observations of a slowly changing, spatially heterogeneous soil property into auditable claims of atmospheric benefit (Oldfield *et al.*, 2022; Smith *et al.*, 2020).

The difficulty of this conversion is well documented. Changes in SOC stocks attributable to management are typically small relative to the background stock and to its spatial variability, so that detection requires many samples, long intervals or both (Poeplau *et al.*, 2022; Smith, 2004). An evaluation of 45 intensively sampled cropland fields found that stock change estimates for individual fields were inaccurate and frequently indicated marked gains or losses where none had occurred, whereas estimates pooled across many fields became reliable only with dense sampling and paired controls (Bradford *et al.*, 2023). Protocols used in voluntary markets differ widely in how they estimate initial stocks, measure change and apply models, and some rely on regional default values that can overestimate initial stocks by up to 2.5 times (Dupla *et al.*, 2024). Wider evidence from carbon crediting outside agriculture has shown systematic over-crediting where baselines and quantification were weak (Badgley *et al.*, 2022; Probst *et al.*, 2024; West *et al.*, 2020), which raises the stakes for soil carbon accounting.

Digital MRV (dMRV) has emerged as the principal proposed remedy. The term refers here to MRV systems in which remotely sensed observations, proximal soil sensors, statistical and machine learning (ML) models, process-based biogeochemical models and digital data infrastructures are combined to reduce the cost and increase the frequency and spatial resolution of soil carbon quantification. The conceptual antecedents are long-standing. Digital soil mapping (DSM) formalised the prediction of soil properties from environmental covariates two decades ago (McBratney *et al.*, 2003), integrated global frameworks linking long-term experiments, monitoring networks and models were proposed a decade later (Smith *et al.*, 2012), and the Intergovernmental Panel on Climate Change (IPCC) inventory guidance has long distinguished default-factor, country-specific and model-based or measurement-based tiers (Intergovernmental Panel on Climate Change [IPCC], 2019). What has changed is the scale of available data and computation. Freely available decametric satellite time series, large soil spectral libraries, deep learning and computationally efficient model–data fusion have made it technically feasible to generate field-level estimates for millions of fields (Guan *et al.*, 2023; Paustian *et al.*, 2019). Recent syntheses document the expanding use of remote sensing and machine learning for SOC estimation, while also emphasising the continuing importance of calibration, validation and adequate reference data (Li *et al.*, 2024; Lima *et al.*, 2025).

Technical feasibility is not the same as evidential credibility. Existing reviews have largely examined individual components: remote sensing of topsoil SOC (Angelopoulou *et al.*, 2019; Vaudour *et al.*, 2022), proximal sensing for carbon accounting (England & Viscarra Rossel, 2018), machine learning for digital soil mapping (Wadoux *et al.*, 2020) or the combined predictive performance of sensing, artificial intelligence (AI) and biogeochemical models (Ding *et al.*, 2025). Others have compared MRV frameworks, protocols or costs (Batjes *et al.*, 2024; Dupla *et al.*, 2024; Ihasusta *et al.*, 2026; Vanderheyden *et al.*, 2026), and a recent systematic review has emphasised uncertainty as a core component of MRV credibility (Wu & Lin, 2026). These contributions leave three problems insufficiently resolved. First, predictive accuracy for SOC stocks is frequently reported as if it implied accuracy for SOC stock change, which is the quantity that credits require. Second, integration is often assumed to reduce uncertainty, whereas the conditions under which error from one evidence stream propagates into, or is corrected by, another have rarely been examined critically. Third, questions of counterfactual baselines, validation domains and verification independence, which determine whether a quantified change can legitimately be credited, are usually treated separately from the technical literature on sensing and modelling. The present review addresses these problems by evaluating the four evidence streams jointly against the requirements of a credible soil carbon claim.

Recent methodological evidence further shows that uncertainty in digital soil mapping can depend on the chosen inference pathway and error-propagation assumptions, while newer Earth-observation approaches are being evaluated for croplands in which conservation practices reduce bare-soil availability (Szatmári & Pásztor, 2025; Castaldi *et al.*, 2026).

1.1 Scope, Research Gap and Objectives

The review concerns mineral agricultural soils under cropland and managed grassland in which carbon farming practices such as reduced tillage, cover cropping, residue retention, organic amendments and grazing or rotation changes are intended to increase SOC stocks. Its central object is the quantification of SOC stock change for project-level and programme-level carbon accounting, including the estimation of baselines and the verification of reported outcomes. Four evidence streams are examined: satellite, airborne and unmanned aerial remote sensing; laboratory and in-field proximal soil sensing; AI, including machine learning and deep learning; and process-based biogeochemical and crop–carbon models, together with the hybrid architectures that combine them. The knowledge gap addressed is the absence of a critical, integrative appraisal of whether and under what conditions these streams, individually and in combination, produce estimates of SOC stock change that are unbiased, accompanied by defensible uncertainty and independently verifiable. The objectives are to define the evidential requirements of a credible soil carbon claim; to appraise the strength and limitations of evidence for each stream against those requirements; to evaluate the principal integration architectures and the validation evidence supporting them; to examine the governance, cost and equity conditions under which digital MRV operates; and to identify prioritised research needs. The intended contribution is an evidence-tiered framework that distinguishes established from hypothesised relationships within digital MRV. Organic soils and peatlands, forest and agroforestry biomass carbon, enhanced rock weathering, biochar-specific accounting, nitrous oxide and methane quantification and the economic design of carbon markets are excluded except where they bear directly on SOC accounting.

2. Methods for Literature Selection

2.1 Review Design and Rationale

A critical narrative review design was adopted because the review question spans soil science, remote sensing, spectroscopy, statistics, computer science, biogeochemical modelling and environmental governance, and because the relevant evidence comprises heterogeneous designs, including field experiments, sampling simulations, model intercomparisons, algorithm benchmarks, protocol analyses and policy evaluations. These designs report non-commensurable outcomes, so that pooled effect estimation of the kind performed in a meta-analysis would be inappropriate, and a scoping review would map the literature without the interpretive appraisal that the question requires. Narrative reviews can offer interpretive depth and theoretical integration that systematic designs do not provide, provided that the selection and appraisal process is transparent and reflexive (Greenhalgh *et al.*, 2018; Sukhera, 2022). The conduct and reporting of the review were guided by the criteria of the Scale for the Assessment of Narrative Review Articles, which emphasises justification of importance, explicit aims, description of the literature search, appropriate referencing, scientific reasoning and presentation of relevant data (Baethge *et al.*, 2019). The narrative design does not remove the need for transparency or appraisal, and the procedures below were applied consistently.

2.2 Information Sources

Searches were conducted in Crossref, OpenAIRE, CORE, the Directory of Open Access Journals (DOAJ), Europe PMC, Semantic Scholar and Google Scholar. These indexes were selected because together they cover the environmental, agricultural, remote sensing and computational literatures relevant to the topic, including open-access and repository-held versions of articles. Institutional sources were searched directly, namely the IPCC Task Force on National Greenhouse Gas Inventories, the Food and Agriculture Organization of the United Nations (FAO), the official legal database of the European Union and the methodology pages of a major voluntary carbon standard. Database searching was supplemented by backward citation searching of recent high-quality reviews and methodological papers, forward citation searching of influential studies, related-article searching and searches for corrections, expressions of concern and retractions.

2.3 Search Strategy and Search Period

The search strategy combined six concept blocks, summarised in Table 1: carbon farming and accounting; remote sensing; proximal sensing; artificial intelligence; process-based and hybrid modelling; and credibility, covering uncertainty, validation, sampling design, baselines, additionality, permanence and cost. A representative combined string was ("soil organic carbon" OR "soil carbon" OR SOC) AND ("carbon farming" OR "carbon credit*" OR "carbon offset*" OR "monitoring, reporting and verification" OR MRV OR "soil carbon accounting") AND ("remote sensing" OR "Earth observation" OR Sentinel-2 OR hyperspectral OR spectroscopy OR "machine learning" OR "deep learning" OR "digital soil mapping" OR "process-based model" OR DayCent OR RothC OR "data assimilation") AND (uncertainty OR validation OR "sampling design" OR baseline OR additionality OR permanence). Stream-specific strings paired each technology block with the SOC and credibility blocks, and syntax was adapted to the query conventions of each index. The search period began on 1 January 2015, the year in which the 4 per mille initiative was launched (Minasny *et al.*, 2017) and in which freely available decametric multispectral time series from the Sentinel-2 constellation began to reshape remote sensing of cropland soils (Vaudour *et al.*, 2022), and it ended on 13 July 2026. Foundational publications before 2015 were included when they were necessary to explain the development of sampling theory, spectroscopy, digital soil mapping or soil carbon modelling.

Table 1 condenses the concept blocks and their function in the review. The blocks were deliberately constructed so that technology terms were never searched in isolation from SOC and credibility terms, because the review concerns the evidential value of these technologies for accounting rather than their general performance. The table also shows that the credibility block, rather than any technology block, is the organising principle of the synthesis.

Table 1. Search concept blocks, condensed free-text terms and their role in the review

Concept block	Condensed free-text terms	Role in the review	Supporting references
Carbon farming and soil carbon accounting	soil organic carbon; soil carbon; carbon farming; carbon credit*; carbon offset*; monitoring, reporting and verification; MRV; soil carbon accounting	Defines the accounting object: SOC stock change attributable to management and its conversion into claims	Oldfield *et al.* (2022); Paul *et al.* (2023); Smith *et al.* (2020)
Remote sensing	remote sensing; Earth observation; Sentinel-2; Landsat; hyperspectral; imaging spectroscopy; bare soil composite; synthetic aperture radar; cover crop detection; tillage mapping	Identifies evidence on direct SOC retrieval, practice detection and biomass observation	Angelopoulou *et al.* (2019); Chabrillat *et al.* (2019); Vaudour *et al.* (2022)
Proximal sensing	proximal soil sensing; soil spectroscopy; vis–NIR; mid-infrared; spectral library; calibration transfer; LIBS; inelastic neutron scattering; bulk density sensing	Identifies evidence on low-cost analytical and in-field alternatives to laboratory measurement	England and Viscarra Rossel (2018); Viscarra Rossel *et al.* (2016)
Artificial intelligence	machine learning; deep learning; artificial intelligence; digital soil mapping; random forest; quantile regression forest; knowledge-guided machine learning	Identifies evidence on data-driven prediction of SOC stocks and change and on validation practice	Heuvelink *et al.* (2021); Wadoux *et al.* (2020)
Process-based and hybrid modelling	process-based model; biogeochemical model; DayCent; RothC; Century; ecosys; data assimilation; model–data fusion; hybrid model	Identifies evidence on mechanistic estimation of change, calibration and integration with observations	Campbell and Paustian (2015); Guan *et al.* (2023)
Credibility	uncertainty; validation; sampling design; equivalent soil mass; baseline; additionality; permanence; leakage; over-crediting; cost	Links technical evidence to the requirements of a creditable claim	Dupla *et al.* (2024); Le Noë *et al.* (2023); Stanley *et al.* (2023)

Note. Abbreviations: LIBS, laser-induced breakdown spectroscopy; MRV, monitoring, reporting and verification; SOC, soil organic carbon; vis–NIR, visible–near-infrared. The asterisk denotes truncation. Terms were combined with Boolean operators and adapted to the syntax of each index

2.4 Eligibility Criteria

Peer-reviewed journal articles, peer-reviewed reviews and syntheses, and methodological or regulatory documents from recognised intergovernmental, governmental and standard-setting bodies were eligible. Studies were included when they addressed the quantification of SOC stocks or stock change in mineral agricultural soils, the performance or validation of a sensing, AI or modelling approach relevant to such quantification, or the design, uncertainty, cost or governance of soil carbon MRV. Field, laboratory, simulation, modelling and protocol-analysis designs were all eligible, provided that methods were reported in sufficient detail to judge the basis of the conclusions. Studies of forests, wetlands and organic soils were considered only when they supplied methodological evidence that transferred directly to agricultural SOC accounting, such as validation of large-scale mapping models or evidence of over-crediting. Publications in English were included. Preprints, conference abstracts, trade literature, promotional material from commercial MRV providers and sources whose bibliographic details could not be confirmed were excluded. No geographical restriction was imposed, although the resulting evidence base was geographically concentrated, as discussed in Section 7.

2.5 Screening, Duplicate Handling and Study Selection

Records retrieved from each index were screened first by title and abstract for relevance to the review question and then by full text or, where full text was inaccessible, by the abstract and publisher record, with inclusion restricted to studies whose central claims could be verified. Duplicate records arising from overlapping indexes and from multiple repository versions of the same article were identified through digital object identifiers and title matching, and the version of record was retained. Bibliographic details and digital object identifiers were verified against the registration agency record for every included source, and retraction and correction notices were checked. Screening was iterative, with citation tracing continued until additional searches yielded no new studies that altered the thematic synthesis. Because selection followed the logic of a critical narrative review rather than a protocol-registered systematic review, record counts are not reported and no flow diagram is presented.

2.6 Critical Appraisal and Evidence Synthesis

Studies were appraised using criteria appropriate to their designs. For sampling and field studies, attention was given to the sampling design, replication, depth and mass conventions, and the presence of controls. For sensing and AI studies, the principal criteria were the independence of validation data, the treatment of spatial dependence and extrapolation, whether stocks or stock changes were evaluated, and the reporting of prediction uncertainty. For process-based and hybrid models, emphasis was placed on calibration strategy, independence of validation, the domain of validity and uncertainty propagation. Policy and protocol analyses were assessed for the transparency of their evidential basis. Formal risk-of-bias scores were not assigned, because no recognised tool is applicable across these heterogeneous designs. Themes were developed around the requirements of a credible soil carbon claim, and each evidence stream was evaluated against those requirements. Disagreements between studies were interpreted by examining differences in scale, soil and climate, validation design, the quantity being predicted and the reference data used, and conclusions were graded according to the consistency and independence of supporting evidence.

3. The Evidential Requirements of a Credible Soil Carbon Claim

Any assessment of digital MRV must begin with the quantity that a credit represents. A soil carbon credit does not certify that a field contains a certain amount of carbon. It certifies that management caused a change in the SOC stock, relative to what would otherwise have occurred, that this change can be expressed as a removal of carbon dioxide (CO_2) or an avoided loss with known uncertainty, and that the claimed benefit will persist or be accounted for if reversed. This section examines those requirements, because they define the standard against which each technology must be judged.

3.1 Stocks, Stock Changes and the Meaning of Removal

Three quantities are routinely conflated in the literature: the SOC stock at a point in time, the change in stock over a monitoring interval and the difference between the observed change and a counterfactual change. Only the last corresponds to a mitigation outcome. SOC storage, meaning an increase in stocks, should be differentiated from sequestration, which implies a net removal of atmospheric CO_2 (Chenu *et al.*, 2019). Don *et al.* (2024) showed that these and related terms, including loss mitigation and accrual, are frequently used interchangeably, and that practices which slow the decline of SOC in soils that are losing carbon reduce emissions but do not constitute net removals. The distinction matters for crediting because the same measured trajectory can represent either an atmospheric removal or merely a smaller loss, depending on the baseline. Mitigation claims must also account for non-CO_2 greenhouse gases and leakage, which can offset or exceed the SOC benefit (Don *et al.*, 2024; Paul *et al.*, 2023).

The realistic magnitude of change sets the detection problem. A meta-analysis of cover cropping across 139 plots at 37 sites estimated an annual SOC stock increase of 0.32 ± 0.08 Mg C ha^{-1} yr^{-1} in a mean depth of 22 cm, with linear accumulation observed for several decades (Poeplau & Don, 2015). Rates of this order typically represent less than 1% of the topsoil stock per year. The response of SOC to no-tillage is more contested. A meta-analysis of 69 paired experiments sampled below 40 cm found that no-tillage increased SOC in the surface 10 cm but decreased it in the 20–40 cm layer, with no significant increase in total stock to 40 cm (Luo *et al.*, 2010), and subsequent syntheses concluded that no-tillage stores additional carbon only under particular soil and climatic conditions and with large uncertainty (Ogle *et al.*, 2019; Powlson *et al.*, 2014). The evidence is therefore strongest for practices that increase carbon inputs, such as cover crops and organic amendments, and weakest for practices that mainly redistribute carbon within the profile, a pattern consistent with broader syntheses that call for reassessment of the storage potential of practices such as no-tillage and emphasise unresolved questions about stabilisation processes and the accurate quantification of soil-based mitigation (Chenu *et al.*, 2019; Paustian *et al.*, 2016). Saturation and diminishing returns further constrain long-term accrual (Moinet *et al.*, 2023). For MRV design, the implication is that credible systems must be able to detect increments of a few tenths of a megagram of carbon per hectare per year against background stocks that are two orders of magnitude larger.

3.2 Detectability: Spatial Variability, Sampling Design and Time

The feasibility of measuring such increments has been examined through power analysis, simulation and empirical resampling. Early modelling showed that an increase in carbon inputs of 20–25% would become detectable only after about 6–10 years with a sampling regime able to resolve a 3% change in background SOC, and would not be detectable at all with a regime able to resolve only a 15% change (Smith, 2004). Analyses of European soil monitoring networks similarly concluded that intervals of about 10 years would be needed to detect large simulated changes, and that minimum detectable changes generally exceeded annual national greenhouse gas emissions, precluding annual accounting from monitoring networks alone (Saby *et al.*, 2008). At the plot scale, resampling a single profile pit by a single pit in German croplands and grasslands produced mean absolute errors of 5.1 and 7.6 Mg C ha^{-1} in the 0–30 cm layer, even with minimal spatial displacement, and replication with three profiles reduced these errors by approximately half (Poeplau *et al.*, 2022).

The most direct evaluation of feasibility for crediting is the resampling analysis of Bradford *et al.* (2023), which used high-density sampling of 45 cropland fields to simulate realistic monitoring designs. Individual-field estimates of change were inaccurate even at higher sampling densities, whereas project-level means across about 30 fields yielded approximately 80% of estimates within 20% of the simulated change. When a dynamic baseline was introduced through paired control fields, accurate estimation of management effects required higher sampling densities, around 30 field pairs and effect sizes of about 3 Mg C ha^{-1} over 10 years. A subsequent comment argued that paired resampling of the same georeferenced locations can make field-level detection more tractable than the independent resampling assumed in that analysis (von Haden *et al.*, 2024). The disagreement is instructive rather than contradictory. Paired designs reduce the variance of change estimates by removing much of the persistent spatial component, but they depend on accurate relocation, consistent depth and mass conventions and the absence of disturbance at sampled points, conditions that are demanding in commercial settings.

Design-based statistical inference provides the logical foundation for verification. Stanley *et al.* (2023) emphasised that valid inferences about SOC in heterogeneous landscapes depend on probability sampling and explicit accounting for measurement error, rather than on convenience or purposive sampling whose uncertainty cannot be defined. Stratified and balanced sampling designs that exploit auxiliary information, including satellite-derived indices and terrain attributes, improve efficiency over simple random sampling at the field scale (de Gruijter *et al.*, 2016; Potash *et al.*, 2022), and sampling theory for mapping and for estimating means has been set out for DSM applications (Brus, 2019). Potash *et al.* (2025) extended this reasoning to the project scale, showing that measure-and-remeasure projects with randomised assignment of practices, two-stage cluster sampling and measurement of about 10% of fields can achieve a competitive return on investment within five years when projects comprise thousands of fields and literature-derived treatment effects are realised commercially. These studies converge on a clear conclusion: direct measurement can quantify average management effects across many fields with acceptable accuracy, but it cannot reliably quantify change on individual fields over typical crediting intervals. This asymmetry defines the role that digital methods must play.

3.3 Soil Mass, Depth and Analytical Conventions

SOC stock is the product of carbon concentration, bulk density and depth, corrected for coarse fragments. Changes in bulk density induced by tillage or by organic matter accumulation alter the mass of soil in a fixed-depth layer and can bias estimated stock changes. Fixed-depth sampling to 30 cm in which bulk density declines from 1.5 to 1.1 g cm^{-3} was shown to underestimate stock change by 17% (Fowler *et al.*, 2023), and the equivalent soil mass (ESM) approach consistently produced lower errors than fixed-depth comparisons in simulations applied to bioenergy cropping systems (von Haden *et al.*, 2020). Procedures for applying ESM across multiple layers are well established (Wendt & Hauser, 2013), and a synthesis of no-tillage studies showed that fixed-depth calculations overestimated surface stocks, because bulk density increased under no-tillage, while underestimating stocks in the 30–70 cm layer relative to ESM (Xiao *et al.*, 2020). ESM is not free of error. Variation in sampled core lengths and the choice of reference mass introduce additional variance, and alternatives based on the mean sampled mass or on equivalent bulk density reduced variance by 38–86% relative to the reference-mass method implemented in an Australian soil carbon method (Rutjens *et al.*, 2026).

Depth is an equally consequential convention. In a 19-year Mediterranean trial, winter cover crops increased SOC by 1.44 Mg C ha^{-1} in the 0–30 cm layer but decreased it by 14.86 Mg C ha^{-1} between 30 and 200 cm, producing a net profile loss that topsoil sampling would have concealed (Tautges *et al.*, 2019). Power analyses of whole-profile data have shown that the probability of detecting even 50% differences in carbon stocks between treatments can be as low as 20–60% because of natural variability, which cautions against interpreting non-significant subsoil differences as evidence of no change (Kravchenko & Robertson, 2011). Redistribution under no-tillage provides a further illustration of why the depth of assessment governs the inferred outcome (Luo *et al.*, 2010). Analytical conventions add a third layer of uncertainty. In calcareous soils, inorganic carbon must be separated from organic carbon, and combustion methods that measure both simultaneously are being developed to reduce the cost and error of this separation (Carter *et al.*, 2024). These conventions are directly relevant to digital MRV because satellite reflectance senses only the upper few millimetres of bare soil, most proximal sensors characterise concentration rather than stock, and most process-based models are calibrated to 0–30 cm stocks. Any digital estimate inherits the depth, mass and analytical assumptions of its reference data.

3.4 Counterfactuals, Additionality, Permanence and Leakage

The counterfactual baseline is the least observable component of a soil carbon claim, yet it can dominate the credited quantity. Where regional SOC stocks are rising or falling for reasons unrelated to the project, such as climate trends, historical land-use legacies or the prior diffusion of conservation practices, a static baseline will attribute these background trends to the project. National-scale counterfactual modelling in the United States estimated that cropland SOC stocks increased every year from 1995 to 2015, largely because of the earlier adoption of climate-smart practices, which implies that part of the remaining potential has already been realised (Ogle *et al.*, 2023). Dynamic baselines constructed from legacy data or control sites are being developed to represent such trends (Somenahally *et al.*, 2025), and paired control fields are the empirical equivalent at the project scale (Bradford *et al.*, 2023).

Additionality raises behavioural as well as biophysical questions. Interviews with farmers participating and not participating in United States soil carbon programmes found that payments were largely reaching farmers who were already implementing, or strongly inclined to implement, the incentivised practices, and that payments were regarded as marginal to adoption decisions (Barbato & Strong, 2023). Satellite evidence points in a similar direction: cover crops in the Midwestern United States were preferentially adopted on fields with lower yields and poorer soils, consistent with self-selection (Seifert *et al.*, 2018). Such selection effects complicate both additionality and the estimation of treatment effects from observational data. Policy analyses have identified additionality, leakage and permanence as persistent challenges for soil carbon sequestration policy (Thamo & Pannell, 2016), and an evaluation of private soil carbon certificates concluded that they are unlikely to deliver the emission offsets attributed to them, chiefly because permanence cannot be guaranteed and long-term monitoring and accountability are lacking (Paul *et al.*, 2023).

The permanence debate is not settled. SOC is dynamically stable, persisting through continuous turnover rather than inertness, and its vulnerability varies among fractions and environments (Dynarski *et al.*, 2020). Leifeld (2023) argued that non-permanent sinks deliver a real and quantifiable climate benefit that can be accommodated through ex ante biophysical discounting, while Minasny and McBratney (2024) proposed a tonne-year unit that credits the time-integrated amount of carbon stored across pools with different mean residence times. These proposals shift the MRV burden from a single end-of-period measurement to repeated or continuous estimation of stocks over time, which is precisely where digital methods are expected to contribute. Evidence from other crediting sectors shows the consequences of weak baselines and quantification: a synthesis of rigorous evaluations covering 2,346 projects estimated that less than 16% of the credits issued to the investigated projects represented real emission reductions (Probst *et al.*, 2024), and independent analyses of forest and avoided-deforestation offsets documented systematic over-crediting (Badgley *et al.*, 2022; West *et al.*, 2020). Although these analyses did not examine soil carbon projects, they establish that crediting integrity fails most often at the baseline and quantification stages, which are the stages digital MRV seeks to transform.

Table 2 summarises the components of a soil carbon claim, the dominant sources of error for each and the present status of evidence. Its main implication is that the components with the weakest evidential support, namely field-level change, counterfactual baselines and permanence, are those that digital methods are most frequently expected to resolve. The table also shows that several error sources, including bulk density change, depth truncation and baseline drift, are systematic rather than random and therefore cannot be reduced simply by increasing the number of observations or model runs.

Table 2. Components of a credible soil carbon claim, dominant sources of error and the status of supporting evidence

Claim component	What must be demonstrated	Dominant error or bias sources	Status of evidence	Supporting references
Baseline SOC stock	Stock at project start with known uncertainty, to the declared depth	Spatial heterogeneity; regional default values; bulk density and coarse fragments; discrepancies among gridded datasets	Well established that direct sampling is required; defaults can overestimate stocks substantially	Dupla *et al.* (2024); Poeplau *et al.* (2022); W. Zhou *et al.* (2023)
SOC stock change	Change over the monitoring interval, distinguishable from sampling and analytical noise	Low signal-to-noise ratio; relocation error; short intervals	Feasible for project-level means with dense, replicated sampling; unreliable for single fields	Bradford *et al.* (2023); Smith (2004); von Haden *et al.* (2024)
Soil mass and depth conventions	Comparison on an equivalent soil mass and across the depth affected by management	Bulk density change; redistribution with depth; inorganic carbon	ESM consistently reduces bias; subsoil losses can reverse topsoil gains	Fowler *et al.* (2023); Tautges *et al.* (2019); von Haden *et al.* (2020)
Counterfactual baseline and additionality	Change relative to what would have occurred without the project, under practices that would not otherwise have been adopted	Background trends; prior adoption; self-selection	Methods emerging; dynamic baselines and control fields rarely implemented at scale	Barbato and Strong (2023); Ogle *et al.* (2023); Somenahally *et al.* (2025)
Permanence and reversal	Persistence of the stored carbon or accounting for its release	Practice reversal; climate-driven losses; fraction-specific turnover	Contested; alternative accounting units proposed but untested in markets	Dynarski *et al.* (2020); Leifeld (2023); Minasny and McBratney (2024)
Leakage and non-CO_2 gases	Absence of displaced emissions and inclusion of other greenhouse gases	Displacement of production; nitrous oxide from inputs	Recognised conceptually; rarely quantified in SOC projects	Don *et al.* (2024); Paul *et al.* (2023); Thamo and Pannell (2016)
Uncertainty and conservativeness	Credited quantity adjusted for the full uncertainty of the estimate	Omitted error sources; overconfident validation	Deduction methods exist but vary widely among protocols	Mathers *et al.* (2023); Stanley *et al.* (2023); Wadoux *et al.* (2026)

Note. Abbreviations: CO_2, carbon dioxide; ESM, equivalent soil mass; SOC, soil organic carbon. Status assessments reflect the synthesis in Section 3 and are not exhaustive

4. Remote and Proximal Sensing as Evidence Streams

Sensing technologies enter soil carbon MRV in three distinct roles that are often conflated: as direct estimators of SOC concentration or stock, as sources of covariates that improve spatial prediction or sampling efficiency, and as observers of the management activities and crop growth that drive SOC change. The strength of evidence differs markedly among these roles, and the distinction is essential for judging what sensing can contribute to crediting.

4.1 Direct Retrieval of Topsoil SOC from Satellite and Airborne Reflectance

The physical basis for optical retrieval is the absorption of visible and shortwave infrared radiation by organic matter, which darkens soil and produces characteristic spectral features, but the signal originates from the upper few millimetres of exposed soil and is modulated by moisture, roughness, crusting, crop residues and green vegetation (Angelopoulou *et al.*, 2019; Chabrillat *et al.*, 2019). Reviews of satellite retrieval over cultivated land have documented a rapid expansion of studies using Sentinel-2 and Landsat data, together with wide variation in accuracy among sites and dates (Vaudour *et al.*, 2022). In a comparison across croplands in Germany, Luxembourg and Belgium, Sentinel-2 predictions were generally only slightly less accurate than airborne hyperspectral predictions, but the ratio of performance to deviation ranged from 2.6 in Luxembourg to 1.1 in Belgium, indicating that site conditions rather than sensor characteristics often limited performance (Castaldi *et al.*, 2019). A time-series analysis in France showed that the cross-validated coefficient of determination for topsoil SOC varied from near zero to 0.58 depending solely on the acquisition date, with soil roughness and moisture the principal controls (Vaudour *et al.*, 2019). Comparisons in the Czech Republic found Sentinel-2 predictions to be marginally less accurate than laboratory spectroscopy, with both airborne and satellite maps performing better where SOC contents were relatively high (Gholizadeh *et al.*, 2018).

The dominant methodological response has been compositing, in which many acquisitions are filtered for bare and dry conditions and combined into a synthetic soil reflectance image. Processors built on Landsat archives (Rogge *et al.*, 2018) and synthetic soil images derived from multi-temporal data in Brazil (Demattê *et al.*, 2018) established the approach, and Sentinel-2 composites have since been refined by screening surface conditions. Using 303 field photographs of surface states, Dvorakova *et al.* (2023) identified a Normalized Burn Ratio 2 (NBR2) threshold that separated dry bare soils from moist or residue-covered soils and showed that prediction uncertainty declined as the number of valid scenes per pixel increased, reaching a minimum at six or more scenes. Crop residues were shown in airborne imaging spectroscopy to degrade predictions unless residue-affected pixels were excluded using cellulose absorption thresholds (Dvorakova *et al.*, 2020). Multi-year bare-soil mosaics combined with soil moisture products and terrain and geophysical covariates improved regional mapping in central France (Urbina-Salazar *et al.*, 2023). The most extensive operational test is a pre-operational European monitoring system that uses three-year Sentinel-2 reflectance composites and the Land Use/Cover Area frame statistical Survey soil database. Its continental coefficient of determination was 0.41 for croplands and 0.28 for permanently vegetated land, and performance was poor in Mediterranean regions where the frequency of bare soil was low (van Wesemael *et al.*, 2024).

These results support a measured conclusion. Satellite reflectance composites can map spatial patterns of topsoil SOC concentration over large areas with moderate accuracy in temperate arable regions with frequent bare-soil exposure. They have not been shown to detect management-induced changes in SOC stock at field scale over crediting intervals. Three reasons explain this gap. The typical prediction errors of several grams of carbon per kilogram of soil exceed the expected annual change in concentration by an order of magnitude; the signal refers to the soil surface rather than the 0–30 cm or deeper layer relevant for accounting; and the practices that increase SOC, such as cover cropping and residue retention, are precisely those that reduce bare-soil exposure and therefore degrade retrieval. Compositing over multiple years, which stabilises the spectral signal, simultaneously blurs the temporal resolution needed to observe change. The apparent contradiction between promising mapping studies and weak change detection is therefore largely explained by the difference between the quantity being predicted and the quantity required for crediting.

4.2 Imaging Spectroscopy and Emerging Sensors

Imaging spectroscopy resolves narrow absorption features associated with organic matter, clay minerals and cellulose, and has been advocated as the next step beyond multispectral retrieval (Chabrillat *et al.*, 2019). Airborne hyperspectral data consistently outperform broadband data in controlled comparisons, but the advantage is frequently modest once surface conditions are controlled, and spectral resampling of airborne data to Sentinel-2 bands did not change prediction accuracy in four of five sub-regions in one comparison (Castaldi *et al.*, 2019). The European monitoring system reproduced spatial patterns obtained from regional algorithms applied to new-generation hyperspectral satellites, while underestimating very high values in kettle holes of a morainic landscape (van Wesemael *et al.*, 2024). Spaceborne imaging spectroscopy remains constrained by revisit frequency, signal-to-noise ratio and the same surface disturbances that affect multispectral data. Synthetic aperture radar (SAR) contributes information on moisture and roughness that can be used to screen or correct optical data (Urbina-Salazar *et al.*, 2023), but it provides limited direct information on SOC and added relatively little to tillage classification beyond Landsat in one large evaluation (Azzari *et al.*, 2019). The evidence therefore supports a supporting rather than a primary role for these sensors in SOC change estimation.

4.3 Remote Sensing of Management Activity and Crop Growth

The strongest contribution of remote sensing to MRV lies not in measuring SOC but in documenting the management and productivity that determine SOC change. Satellite classifiers have mapped winter cover crops in the Midwestern United States with an accuracy of 91.5% (Seifert *et al.*, 2018), and fusion of multi-source satellite data with machine learning quantified field-level cover crop adoption from 2000 to 2021, showing that adoption in 2021 was four times that of 2011 but still covered only 7.2% of maize (*Zea mays* L.) and soybean (*Glycine max* (L.) Merr.) fields (Q. Zhou *et al.*, 2022). Tillage intensity was mapped at 30 m resolution across the North Central United States with accuracies of 75–79%, depending on whether validation data were split by field, year or state (Azzari *et al.*, 2019). Crop residue cover, a direct indicator of tillage intensity and carbon return, was estimated from narrow shortwave infrared residue indices on WorldView-3 imagery with a coefficient of determination of 0.94, compared with 0.84 for broadband indices available from Landsat-type sensors, although estimation was limited to fields with little green vegetation (Hively *et al.*, 2018).

Satellite time series also constrain the carbon inputs that drive SOC change. Assimilation of high-resolution green area index observations into a simple crop model reproduced daily CO_2 fluxes and components of the annual net ecosystem carbon budget of winter wheat, evaluated against eight years of eddy covariance measurements in south-western France (Pique *et al.*, 2020). The AgriCarbon-EO processing chain extended this approach to 10 m resolution over regional extents such as 10,000 km², propagating uncertainty from reflectance to carbon budget components and reproducing net ecosystem exchange measured by flux towers with coefficients of determination of 0.77–0.87 (Wijmer *et al.*, 2024). Eddy covariance measurements have long shown that harvest exports and management strongly affect cropland net ecosystem carbon budgets (Ceschia *et al.*, 2010), which is why remote estimation of biomass and yield is a prerequisite for credible input-driven estimates of SOC change.

Practice detection has three limitations that matter for verification. Classification accuracies of 75–92% imply that a substantial fraction of fields are misclassified, and errors may be spatially clustered rather than random, so that aggregate adoption estimates can be more reliable than field-level attributions. Accuracy estimates depend on how validation data are split, and evaluations that separated training and testing by field, year and state gave lower but more realistic accuracies than random splits (Azzari *et al.*, 2019). Detection of an activity is not evidence of its effect on SOC, which depends on soil, climate, biomass and duration (Ogle *et al.*, 2019; Poeplau & Don, 2015). Remote sensing of management is therefore best regarded as verification of activity data and as an input to models, not as a substitute for measurement of outcomes.

4.4 Laboratory and Field Proximal Spectroscopy

Proximal sensing addresses the cost of the analytical step rather than the sampling step. Diffuse reflectance spectroscopy in the visible–near-infrared (vis–NIR) and mid-infrared (MIR) ranges has been evaluated for more than two decades as a rapid alternative to dry combustion, with MIR generally providing more accurate predictions of organic carbon than vis–NIR under laboratory conditions (Viscarra Rossel *et al.*, 2006). The development of large spectral libraries was advanced early through work on African soils (Shepherd & Walsh, 2002) and has since expanded to continental and global scales. A continental vis–NIR library enabled prediction of SOC across Europe (Stevens *et al.*, 2013), a global vis–NIR library was assembled from contributions across many countries (Viscarra Rossel *et al.*, 2016), and the national MIR library of the United States Kellogg Soil Survey Laboratory has supported accurate prediction of organic carbon and related soil health indicators (Sanderman *et al.*, 2020; Wijewardane *et al.*, 2018). A review of proximal sensing for carbon accounting concluded that vis–NIR spectroscopy was the most suitable method for SOC concentration and active gamma-ray attenuation for bulk density, while noting that sensors for gravel content were lacking and that standardised methods and procedural guidelines were needed before proximal sensing could support reporting and verification (England & Viscarra Rossel, 2018).

The transfer of this laboratory evidence to in-field and commercial settings is less secure. An independent field evaluation of a commercial multi-sensor probe on 100 European arable samples found that default algorithms overestimated SOC by 0.20–0.27 percentage points; a pH correction reduced bias to 0.11 percentage points, but no in-field model met all equivalence criteria against laboratory methods, even though per-measurement costs were less than one tenth of laboratory costs (Kohl *et al.*, 2025). Accuracy thresholds that are acceptable for soil fertility advice are not necessarily acceptable for crediting, where a systematic bias of a tenth of a percentage point of carbon concentration can exceed the entire expected change over a monitoring period. The evidence supports laboratory MIR spectroscopy, calibrated and quality-controlled against dry combustion, as a credible substitute for part of the conventional analytical workload, whereas in-field sensing is better suited to dense mapping and stratification than to certification-grade measurement.

4.5 Spectral Libraries, Calibration Transfer and Inter-Instrument Comparability

Spectral predictions are only as transferable as the calibrations that underpin them. The lack of calibration libraries representing local soils has been identified as a principal constraint on the wider uptake of soil spectroscopy, motivating the development of a global calibration library and estimation service (Shepherd *et al.*, 2022). Regional libraries often perform poorly at local scales. When local samples were available, localised modelling approaches, including memory-based learning, outperformed spiking of regional libraries with local samples once more than about 20 local samples were available (Ng *et al.*, 2022), and deep transfer learning has been used to adapt global spectra to local soil carbon monitoring (Shen *et al.*, 2022). A review of calibration set optimisation and library transfer emphasised that calibration choices can materially alter soil carbon estimates (Dorantes *et al.*, 2022).

Inter-instrument comparability is a further source of systematic error. A large international ring trial of MIR instruments found that all instruments predicted well when calibrated internally, but that transfer of models from the Kellogg Soil Survey Laboratory library to instruments with contrasting configurations required spectral standardisation, and that non-linear models were less sensitive to spectral differences than global partial least squares regression (Safanelli *et al.*, 2023). A separate test of calibration transfer between two spectrometers obtained good predictions of total carbon after pre-processing, with further improvement mainly through bias removal when about 50 transfer samples were used (Sanderman *et al.*, 2023). For MRV, the implication is that spectroscopic measurements from different laboratories, instruments or times cannot be assumed to be comparable. Repeated measurements for change detection must either use the same instrument and calibration or include explicit transfer and bias checks, and these procedures must be documented for verification.

4.6 Bulk Density, Profile Sensing and Nuclear Methods

Because stock depends on bulk density and depth, sensing of concentration alone is insufficient. Combining gamma-ray attenuation and vis–NIR spectroscopy allowed bulk density to be measured on fresh, 1 m soil cores in the field with accuracy similar to that of the conventional single-ring method, enabling stocks to be calculated on either a fixed-depth or cumulative soil mass basis (Lobsey & Viscarra Rossel, 2016). An integrated core sensing system extended this approach to fine-resolution profiles of bulk density and carbon fractions, with Kalman smoothing used to derive complete profiles with propagated uncertainties (Viscarra Rossel *et al.*, 2017). Such systems directly address the mass and depth conventions identified in Section 3.3, although their deployment remains limited.

Nuclear and laser-based techniques measure carbon in situ or with minimal preparation. Inelastic neutron scattering (INS) has been compared with dry combustion for in situ determination of the soil carbon pool (Wielopolski *et al.*, 2011), and benchmarking of a mobile INS system showed agreement with dry combustion within experimental error for the upper soil layer (Yakubova *et al.*, 2016), while laser-induced breakdown spectroscopy (LIBS) has been reviewed as a quantitative method for soil carbon with particular attention to matrix effects and the separation of organic from inorganic carbon (Senesi & Senesi, 2016). The listing of infrared spectroscopy, LIBS and INS alongside dry combustion among the measurement techniques recognised in the current version of a widely used voluntary methodology indicates growing regulatory acceptance (Verra, 2025). Acceptance does not in itself demonstrate equivalence across soils and instruments, and the evidence base for in-field INS and LIBS remains small, sensitive to the depth sensed and dependent on site-specific calibration.

4.7 Measurement Error as a Component of the Uncertainty Budget

Proximal sensing replaces a small analytical error with a larger but cheaper prediction error, and this trade-off must be represented in subsequent modelling. A geostatistical study in western Cameroon that incorporated measurement error variances from conventional and MIR analyses into DSM found that predicted values changed little, but prediction uncertainties were quantified more realistically when measurement error was acknowledged (Takoutsing *et al.*, 2022). Uncertainty in initial soil data also propagates into model-based estimates: in the United States Midwest, two widely used SOC stock datasets differed by a normalised root mean square error of 48% for 0–30 cm stocks, and simulated stock changes driven by the two datasets were only weakly correlated, with initial SOC concentration a larger source of uncertainty than bulk density (W. Zhou *et al.*, 2023). These findings imply that the choice of analytical method and baseline data is not a neutral technical detail but a determinant of credited quantities. Table 3 compares the sensing streams discussed in this section with respect to the quantity observed, representative demonstrated performance, principal limitations and the MRV role that the evidence can support. The comparison shows that no sensing stream directly observes field-level SOC stock change. The most defensible roles are the verification of management activity, the constraint of carbon inputs, the stratification of sampling and the reduction of analytical cost, each of which strengthens rather than replaces design-based measurement.

Table 3. Comparison of remote and proximal sensing streams for soil carbon MRV

Sensing stream	Quantity observed	Representative demonstrated performance	Principal limitations for MRV	Defensible MRV role	Supporting references
Multispectral bare-soil composites	Topsoil reflectance related to SOC concentration in the upper millimetres	Continental R^2 of 0.41 in croplands; site-level RPD 1.1–2.6; strong dependence on date and surface state	Surface signal only; confounded by moisture, residues and roughness; low bare-soil frequency under carbon farming practices	Spatial covariate; sampling stratification; regional pattern mapping	Castaldi *et al.* (2019); Dvorakova *et al.* (2023); van Wesemael *et al.* (2024); Vaudour *et al.* (2019)
Imaging spectroscopy	Narrow absorption features of organic matter, clays and cellulose	Modest gains over broadband data once surface conditions are controlled	Revisit frequency; signal-to-noise ratio; same surface disturbances	Improved covariates; residue screening	Chabrillat *et al.* (2019); Dvorakova *et al.* (2020); Gholizadeh *et al.* (2018)
Practice and residue detection	Cover crops, tillage intensity, residue cover	Accuracy of 75–92% for practices; residue R^2 up to 0.94 with narrow SWIR indices	Misclassification; validation-split dependence; activity is not outcome	Verification of activity data; model inputs; adoption baselines	Azzari *et al.* (2019); Hively *et al.* (2018); Seifert *et al.* (2018); Q. Zhou *et al.* (2022)
Crop biomass and phenology for data assimilation	Green area index, biomass, yield and derived carbon inputs	NEE R^2 of 0.77–0.87 against flux towers at 10 m resolution	Tested mainly for cereals in temperate regions; soil processes simplified	Constraint of carbon inputs in process-based models	Pique *et al.* (2020); Wijmer *et al.* (2024)
Laboratory vis–NIR and MIR spectroscopy	Spectra calibrated to SOC concentration	Accurate predictions from large national libraries; MIR generally superior to vis–NIR	Calibration transfer; inter-instrument bias; local representativeness	Reduction of analytical cost under quality control	Safanelli *et al.* (2023); Sanderman *et al.* (2023); Viscarra Rossel *et al.* (2006); Wijewardane *et al.* (2018)
In-field multi-sensor probes	SOC concentration estimated in situ	Bias of 0.11–0.27 percentage points; precision criteria not met in one independent evaluation	Systematic bias relative to laboratory methods; limited independent evaluation	Dense mapping and stratification rather than certification	Kohl *et al.* (2025)

Sensing stream	Quantity observed	Representative demonstrated performance	Principal limitations for MRV	Defensible MRV role	Supporting references
Gamma-ray attenuation and core sensing	Bulk density and fine-resolution profile properties	Accuracy similar to single-ring method for bulk density	Limited deployment; equipment cost	Stock calculation on an equivalent soil mass basis	Lobsey and Viscarra Rossel (2016); Viscarra Rossel *et al.* (2017)
INS and LIBS	Elemental carbon in situ or with minimal preparation	Agreement with dry combustion within experimental error in benchmark tests	Small evidence base; shallow sensing depth; matrix effects	Emerging measurement option subject to site calibration	Senesi and Senesi (2016); Wielopolski *et al.* (2011); Yakubova *et al.* (2016)

Note. Abbreviations: INS, inelastic neutron scattering; LIBS, laser-induced breakdown spectroscopy; MIR, mid-infrared; MRV, monitoring, reporting and verification; NEE, net ecosystem exchange; R^2, coefficient of determination; RPD, ratio of performance to deviation; SOC, soil organic carbon; SWIR, shortwave infrared; vis–NIR, visible–near-infrared. Performance values are drawn from the cited studies and are not directly comparable across rows because of differences in sites, reference methods and validation designs

5. Artificial Intelligence: Predictive Power, Validation and the Limits of Data-Driven Inference

AI enters soil carbon MRV mainly through machine learning models that predict SOC from spatial and temporal covariates, through deep learning architectures that exploit spatial context or spectral information, and increasingly through hybrid schemes that embed mechanistic knowledge. This section evaluates the purely data-driven approaches. Hybrid approaches are considered in Section 6 because their credibility depends on the process-based component.

5.1 Machine Learning for Mapping SOC Stocks in Space and Time

Machine learning has become the dominant approach in DSM because it captures non-linear relationships between SOC and covariates describing climate, terrain, parent material, vegetation and land use (Wadoux *et al.*, 2020). Global products such as SoilGrids 2.0 use quantile regression forests (QRF) to deliver predictions with spatially explicit uncertainty (Poggio *et al.*, 2021). Deep learning architectures can exploit landscape context: a convolutional neural network that used neighbourhoods of covariate pixels reduced prediction error by 30% relative to conventional point-based models in a national SOC mapping example for Chile (Padarian *et al.*, 2019). A recent review concluded that deep learning often outperforms classical machine learning for SOC prediction, but that no single algorithm is consistently best and that the inclusion of remote sensing covariates improves accuracy (Ding *et al.*, 2025).

Extending machine learning from space to time is the step that matters for MRV. Heuvelink *et al.* (2021) used QRF with dynamic vegetation covariates to map annual SOC stocks in Argentina from 1982 to 2017. The model predicted modest temporal change, with a larger decline in the Pampa region, and its predicted temporal variation for 2001–2015 was seven times larger than that obtained with the IPCC Tier 1 approach. Prediction uncertainties were nevertheless substantial and were attributed to the limited number and poor temporal distribution of calibration data and the limited explanatory power of covariates. Space-time mapping of SOC from remote sensing and machine learning has since been developed further (Bartsch *et al.*, 2025), and landscape-scale studies have combined repeated monitoring with DSM to map stock change (Ellili *et al.*, 2019). The shared finding is that machine learning captures spatial variation well but is less able to represent temporal dynamics, especially where management rather than land-cover change drives SOC trajectories. This interpretation is supported by a regional comparison in eastern China, in which temporal trends of SOC stock predicted by machine learning differed markedly from those of a process-oriented model (Zhang *et al.*, 2024).

A recent end-to-end application of DSM to MRV illustrates both the potential and the conditions of the approach. Using 1,019 samples from seven fields in two Midwestern states of the United States, Wadoux *et al.* (2026) calibrated a QRF model with a cross-validated coefficient of determination of 0.61 and a prediction interval coverage of 91%, aggregated field predictions to the project level with geostatistical methods that account for spatial correlation, and derived uncertainty deductions. For an illustrative 5-year project of about 174 ha, the deduction was 12%, whereas projects larger than about 20,000 ha over five or more years generally had deductions below 5%. The analysis showed that increasing the number of samples per stratum changed mean estimates little but reduced project-level variance. The favourable economics of DSM for MRV thus depend on spatial averaging over large projects, which is consistent with the sampling evidence reviewed in Section 3.2.

5.2 Validation, Spatial Dependence and the Area of Applicability

The accuracy statistics reported for machine learning maps depend critically on how validation is performed. In a large forest inventory in central Africa, a random forest model appeared to explain more than half of the variation in aboveground biomass under non-spatial cross-validation, but showed almost no predictive power when spatial autocorrelation was accounted for (Ploton *et al.*, 2020). Although that study concerned biomass rather than soil, the same mechanism applies to SOC maps calibrated with clustered samples. The appropriate remedy remains contested. Wadoux *et al.* (2021) argued that spatial cross-validation lacks a theoretical basis for estimating map accuracy, that standard cross-validation is deficient when data are clustered, and that probability sampling with design-based inference provides the preferred basis for accuracy assessment. Subsequent work proposed weighting and model-based procedures to reduce bias when only clustered samples are available (de Bruin *et al.*, 2022). Meyer and Pebesma (2021) introduced the area of applicability (AOA), defined through a dissimilarity index in predictor space, beyond which cross-validation error estimates no longer hold, and later argued that many global ecological maps are presented without adequate assessment of where their predictions are reliable (Meyer & Pebesma, 2022).

These debates have direct consequences for MRV. A crediting model is typically calibrated on research sites, soil surveys or a subset of project fields and then applied to other fields, years and practices. If the fields to be credited lie outside the covariate and management space of the calibration data, reported validation statistics do not apply. The disagreement between design-based and model-based validation is best resolved by recognising that they answer different questions. Design-based validation with probability samples estimates the accuracy of a map or project total, which is what verification requires, whereas AOA-type diagnostics identify where a model should not be used. A credible dMRV system requires both, and neither is routinely reported in commercial applications.

5.3 Uncertainty Quantification and Spatial Aggregation

Credits are issued for aggregated quantities, such as project totals, rather than for individual pixels, and the uncertainty of an aggregate is not the average of pixel uncertainties. Szatmári *et al.* (2021) showed that predicting spatial totals and averages of SOC stock change with quantified uncertainty requires geostatistical treatment of spatially correlated prediction errors and cannot be achieved by machine learning alone. Using regression cokriging with random forest and stochastic cosimulation, they estimated an increase in Hungarian topsoil SOC stock between 1992 and 2010 with a 90% prediction interval, and found that larger spatial supports reduced both predictions and uncertainties of average change and made statistically significant change easier to demonstrate. The same principle underlies the project-level uncertainty analysis of Wadoux *et al.* (2026). Measurement error in calibration data must also be included, since ignoring it produces unrealistically narrow prediction intervals (Takoutsing *et al.*, 2022). The evidence is therefore consistent in indicating that pixel-level machine learning uncertainty cannot be summed or averaged naively for crediting, and that aggregation methods must be specified and validated.

5.4 Interpretability, Causality and the Risk of Spurious Signals

Machine learning models for SOC are trained on observational data in which management is correlated with soil, climate and farm characteristics. Interpretation methods such as variable importance, partial dependence and Shapley values can reveal how a model uses covariates, but they describe the model rather than causal processes in the soil (Wadoux & Molnar, 2022). The satellite evidence that cover crops were adopted preferentially on less productive fields (Seifert *et al.*, 2018) shows how selection can confound associations between practices and outcomes. A model trained on such data might attribute differences in SOC to cover cropping that actually reflect pre-existing soil conditions, or the reverse. For crediting, where the quantity of interest is a causal treatment effect relative to a counterfactual, purely associative models are therefore insufficient unless combined with designs that support causal inference, such as randomised or matched controls. Data-driven Earth system science has recognised this limitation and has advocated hybrid models that combine the pattern recognition of deep learning with process understanding (Reichstein *et al.*, 2019), which leads to the modelling approaches examined in Section 6.

6. Process-Based Models and Hybrid Model–Data Fusion

Because direct measurement cannot resolve change on individual fields and data-driven models struggle with temporal dynamics, most agricultural carbon crediting has relied on process-based biogeochemical models to estimate SOC accrual (Potash *et al.*, 2025). These models represent the inputs, decomposition and stabilisation of organic matter as functions of climate, soil properties and management, and they can simulate the counterfactual management scenario that no measurement can observe. Their credibility therefore rests on calibration, validation and uncertainty quantification within the domain of application.

6.1 Process-Based Models as Accounting Engines

Soil organic matter models range from simple multi-pool first-order models such as RothC to ecosystem models such as Century and DayCent and to microbially explicit formulations, and their application has expanded from research to national inventories and project-level accounting (Campbell & Paustian, 2015). The IPCC guidance recognises model-based estimation as a higher-tier approach for national inventories, provided that models are evaluated against measurements representative of national conditions (IPCC, 2019). The FAO protocol for MRV of SOC in agricultural landscapes similarly combines baseline measurement, repeated sampling and modelling (Food and Agriculture Organization of the United Nations [FAO], 2020).

The magnitude of model uncertainty depends strongly on scale. An early national analysis using the Century model estimated increases in United States cropland SOC during the 1990s with uncertainties of about ±16–22% at the national scale, but median uncertainties exceeded ±100% at the regional scale and several hundred per cent at the site scale, with most uncertainty attributable to model structure (Ogle *et al.*, 2010). This scale dependence is central to digital MRV, which seeks field-level estimates precisely where model uncertainty is largest. Bayesian calibration offers a partial remedy. Calibration of DayCent with 491 SOC measurements from 19 long-term experiments reduced model uncertainty by a factor of 6.6 relative to prior parameter distributions (Gurung *et al.*, 2020). The most directly relevant evaluation for crediting calibrated and validated a credit-oriented version of DayCent with 668 measured SOC stock changes from 41 research sites across 14 combinations of crops and practices, using Bayesian Markov chain Monte Carlo methods and k-fold cross-validation (Mathers *et al.*, 2023). The model met the performance criteria of a voluntary soil enrichment protocol, with at least 90% of prediction intervals covering measured values and mean bias smaller than pooled measurement uncertainty, and the posterior distributions allowed uncertainty deductions to be calculated. The authors stated explicitly that the calibrated parameters are valid only within the domain of the validation dataset.

Two qualifications limit the generality of these results. First, validation data from long-term experiments represent well-managed research plots with accurate records of management, and their transferability to commercial fields with less precise management data is not established. Potash *et al.* (2025) noted that measure-and-remeasure projects could provide the independent validation on commercial farms that research trials cannot supply. Second, model inputs themselves carry uncertainty. Simulated stock changes in the United States Midwest were only weakly correlated when the same model was driven by two different soil datasets, and uncertainty in initial SOC concentration dominated that of bulk density (W. Zhou *et al.*, 2023). Model-based crediting therefore depends on field-specific baseline measurements, which links it back to the sampling requirements discussed in Section 3.

6.2 Structural Uncertainty, Model Ensembles and Independent Validation

Different models forced with identical inputs can produce divergent projections. Three soil biogeochemical models produced similar global SOC stocks under common forcing but diverged in their projections of twentieth-century change, with some gaining and others losing more than 20 Pg C, reflecting differences in the temperature sensitivity of turnover and in the representation of freeze–thaw processes and textural controls on stabilisation (Wieder *et al.*, 2018). In German croplands, the mean absolute error of simulated SOC stock trends varied by almost two orders of magnitude among 30 combinations of six models and five carbon input methods, the choice of model mattered more than the choice of input method, and a multi-model ensemble outperformed individual combinations (Riggers *et al.*, 2019). An ensemble of 26 models tested against long-term bare-fallow experiments showed that generic, knowledge-based parameterisation described SOC decline adequately, and highlighted the dependence of performance on calibration strategy and initialisation (Farina *et al.*, 2021).

The deeper problem is that few models are validated independently against observed time series. A review of about 250 SOC models found a critical lack of independent validation using observed time series and noted that approximately 60% of the models were designed for conceptual understanding rather than prediction (Le Noë *et al.*, 2023). Model structure is also evolving. Frameworks that represent measurable fractions, such as particulate and mineral-associated organic matter, have been proposed to replace conceptual pools that cannot be observed directly (Abramoff *et al.*, 2018; Lavallee *et al.*, 2020). Constraining DayCent pools with measured carbon fractions improved initialisation and increased projected SOC losses under land-cover change and strong warming by about 30% relative to the default model (Dangal *et al.*, 2022). Such structural changes can alter credited quantities substantially, which implies that model versions, parameter sets and domains of validity must be fixed and documented for each crediting period.

6.3 Hybrid Architectures: Data Assimilation, Knowledge-Guided Learning and Machine Learning–Process Fusion

Hybrid approaches seek to combine the temporal and counterfactual reasoning of process-based models with the spatial detail and observational constraint of sensing and machine learning. The general case for hybrid modelling in Earth system science rests on the complementary weaknesses of physical and data-driven models (Reichstein *et al.*, 2019). In agriculture, three families of hybrid architecture can be distinguished.

The first family assimilates remotely sensed observations into crop–carbon models. Data assimilation of remote sensing into crop models has been reviewed extensively for yield and biomass estimation (Jin *et al.*, 2018), and its extension to carbon budgets has been demonstrated for winter wheat in south-western France (Pique *et al.*, 2020; Wijmer *et al.*, 2024). These systems constrain photosynthesis, biomass and residue return with observations, which reduces the uncertainty of carbon inputs to the soil. They do not directly constrain decomposition or stabilisation, and their authors have identified the need to represent more soil processes (Wijmer *et al.*, 2024). A comparative analysis of SOC monitoring methodologies for croplands concluded that no method suits all purposes and recommended, where possible, measure-and-remeasure or parsimonious Tier 3 models with assimilation of Earth observation data, on the grounds that biomass assimilation improves the estimation of carbon returns and model accuracy at low cost (Ihasusta *et al.*, 2026). A Finnish field observatory network has similarly combined field measurements, weather data, satellite imagery and models for near-real-time model–data synthesis on research sites and pilot farms (Nevalainen *et al.*, 2022).

The second family embeds process knowledge into machine learning. Knowledge-guided machine learning (KGML) uses outputs and structures of process-based models to train and constrain neural networks. In the United States Corn Belt, a KGML framework integrating knowledge embedded in a process-based agroecosystem model, high-resolution remote sensing and machine learning outperformed both the process-based model and black-box machine learning in quantifying carbon cycle dynamics and resolved 86% more spatial detail in SOC change than coarse-resolution approaches (Liu *et al.*, 2024). A related framework coupled KGML with data assimilation for efficient agroecosystem prediction in the same region (Yang *et al.*, 2023). At the global scale, an approach combining global-scale datasets, a microbially explicit model, data assimilation and deep learning indicated that microbial carbon use efficiency was at least four times as important as other evaluated factors in determining SOC storage (Tao *et al.*, 2023). These studies demonstrate the capacity of hybrid models to extract process-level information from large datasets, but their evaluation has focused on predictive skill for fluxes, crop variables and SOC stocks, rather than on independently measured SOC stock changes in commercial fields.

The third family uses process-based simulations to inform machine learning in space and time. In an eastern Chinese cropland region, a hybrid model that used process-oriented predictions in unsampled years as additional training data reduced root mean square error by 19% relative to machine learning alone and produced temporal trends consistent with the process model (Zhang *et al.*, 2024). In another heavily modified region, sequential updating of random forest and RothC predictions with an ensemble Kalman filter (EnKF) yielded a coefficient of determination of 0.52 for 2015, compared with 0.47 for random forest and 0.13 for RothC (Xie *et al.*, 2024). The integrated system-of-systems framework proposed for field-level agricultural carbon outcomes combines these elements with cross-scale sensing of environment, management and crop conditions, model–data fusion, AI for scaling to millions of fields and multi-tier validation (Guan *et al.*, 2023).

The evidence on hybrids is encouraging but asymmetric. Improvements are typically demonstrated relative to weaker baselines, often in the region where the method was developed, using cross-validation within the same dataset. Independent evaluation against remeasured SOC stock change on fields not used in development is rare. Hybrid models can also transmit the structural assumptions of the process model into apparently data-driven predictions, so that agreement between hybrid and process model trends, as reported by Zhang *et al.* (2024), may indicate consistency rather than independent confirmation. The claim that integration reduces uncertainty is plausible and supported for fluxes and biomass, but it remains to be demonstrated for credited SOC stock changes.

6.4 Measure-and-Model Versus Measure-and-Remeasure

Voluntary protocols distinguish between approaches that estimate SOC change primarily from calibrated models initialised with measured baselines and approaches that estimate change directly from repeated measurements (Dupla *et al.*, 2024; Oldfield *et al.*, 2022). The two are often presented as competing alternatives. The evidence reviewed here suggests that they are complementary and that their relationship is hierarchical with respect to verification. Model-based approaches provide field-level estimates, counterfactual scenarios and annual reporting, but their validity is bounded by the calibration domain and by the fidelity of management data (Mathers *et al.*, 2023). Measure-and-remeasure approaches rely on fewer assumptions and provide an independent check on model-based estimates, but are statistically reliable only for aggregated estimates across many fields and require randomised or matched controls to address dynamic baselines (Bradford *et al.*, 2023; Potash *et al.*, 2025). A digital MRV architecture that uses models and sensing to allocate credits among fields while using design-based remeasurement to verify and, where necessary, correct the aggregate would exploit the strengths of both. Such true-up designs are conceptually straightforward but have not yet been evaluated empirically at commercial scale.

Table 4 compares the principal integration architectures in terms of their mechanism, the strongest available supporting evidence, their main unresolved weaknesses and their current maturity for crediting. The comparison indicates that architectures with the strongest validation evidence for SOC stock change, namely calibrated process-based models and design-based remeasurement, are the least digital, whereas the most technologically advanced architectures have been validated mainly for stocks, fluxes or crop variables. This inversion is the central tension of digital MRV.

Table 4. Integration architectures for soil carbon accounting and their evidential status for crediting SOC stock change

Architecture	Integration mechanism	Strongest supporting evidence	Principal unresolved weakness	Maturity for crediting	Supporting references
Design-based measure-and-remeasure with sensing-informed stratification	Probability sampling at baseline and remeasurement; satellite and terrain covariates used for stratification	Accurate project-level means with dense sampling across about 30 fields or field pairs; economic feasibility for projects of thousands of fields	Unreliable for individual fields; requires controls for dynamic baselines; long intervals	High for aggregated verification	Bradford *et al.* (2023); Potash *et al.* (2025); Potash *et al.* (2022)
Calibrated process-based model (measure-and-model)	Measured baseline stocks; Bayesian calibration on long-term experiments; simulated counterfactual	Validation against 668 measured stock changes from 41 sites with at least 90% interval coverage	Domain-limited; management data fidelity; structural uncertainty	High within validated domain; uncertain outside it	Gurung *et al.* (2020); Mathers *et al.* (2023); Ogle *et al.* (2010)
Multi-model ensemble	Averaging or weighting of several process models	Ensembles outperform individual models for regional SOC trends	Does not remove shared structural biases; computational and governance complexity	Moderate for inventories; limited for projects	Farina *et al.* (2021); Riggers *et al.* (2019)
Machine learning DSM with geostatistical aggregation	QRF or deep learning on covariates; aggregation accounting for spatially correlated errors	Project-level uncertainty and deductions derived from calibrated QRF; valid uncertainty of spatial totals	Captures spatial variation better than temporal change; extrapolation beyond calibration domain	Moderate for stocks; low for change	Szatmári *et al.* (2021); Wadoux *et al.* (2026)
Space-time machine learning	Dynamic covariates such as	National space-time maps with	Large uncertainty where temporal	Low for crediting	Bartsch *et al.* (2025);

Architecture	Integration mechanism	Strongest supporting evidence	Principal unresolved weakness	Maturity for crediting	Supporting references
	vegetation indices predict annual stocks	quantified uncertainty	calibration data are sparse; management effects weakly represented		Heuvelink *et al.* (2021)
Machine learning augmented with process-model output or EnKF updating	Process simulations used as training data or sequentially fused with machine learning predictions	Error reduction of 19% relative to machine learning; improved R^2 relative to each component	Validated within development regions; potential transmission of model bias	Low to moderate; research stage	Xie *et al.* (2024); Zhang *et al.* (2024)
Earth observation data assimilation into crop–carbon models	Satellite green area index or reflectance constrains biomass, fluxes and carbon inputs	NEE and biomass reproduced against flux towers and destructive samples	Soil processes simplified; mainly temperate cereals	Moderate for carbon inputs; low for SOC change	Ihasusta *et al.* (2026); Pique *et al.* (2020); Wijmer *et al.* (2024)
Knowledge-guided machine learning	Neural networks trained on and constrained by process-model structure and outputs, with remote sensing inputs	Outperformed process-based and black-box models for carbon cycle variables in the United States Corn Belt	Independent validation against remeasured SOC change lacking; regional concentration	Low; promising research stage	Guan *et al.* (2023); Liu *et al.* (2024); Yang *et al.* (2023)

Note. Abbreviations: DSM, digital soil mapping; EnKF, ensemble Kalman filter; NEE, net ecosystem exchange; QRF, quantile regression forest; R^2, coefficient of determination; SOC, soil organic carbon. Maturity ratings are qualitative judgements derived from the synthesis in Sections 3–6 and indicate the strength of evidence for crediting SOC stock change, not general scientific merit

7. Credibility as a Socio-Technical Property: Protocols, Costs, Equity and Data Governance

The technical evidence reviewed above does not determine credibility on its own. Credibility also depends on how protocols translate estimates into credits, on who can afford to participate, on whether data and models can be scrutinised and on the geographical representativeness of the evidence base.

7.1 Protocol Heterogeneity and the Treatment of Uncertainty

Analyses of voluntary protocols have found wide variation in scientific rigour. Some protocols do not require in situ baseline sampling, protocols relying on models differ in the practices and information they require, and protocols relying on sampling differ in design, tools, analytical methods and stock calculation (Dupla *et al.*, 2024). A scoring of MRV methodologies across croplands, grasslands and forests concluded that the main obstacle to a unified MRV system lies in the monitoring component, and proposed a modular architecture in which methods can be combined according to context (Batjes *et al.*, 2024). Regulatory frameworks are beginning to converge on principles rather than methods. The European Union certification framework requires quantified net benefits, additionality, long-term storage and sustainability and provides for liability mechanisms in the event of release, while leaving detailed methodologies to subsequent delegated acts (European Parliament & Council of the European Union, 2024). A widely used voluntary methodology now lists several accepted analytical techniques and recommends reassessment of baselines every five years (Verra, 2025).

Uncertainty deductions are the principal mechanism by which protocols convert uncertain estimates into conservative credits. Their credibility depends on whether all relevant error sources are included. The deductions derived for a calibrated process-based model reflected parameter and residual uncertainty within the validation domain (Mathers *et al.*, 2023), and those derived for a DSM approach reflected prediction and spatial aggregation uncertainty (Wadoux *et al.*, 2026). Neither framework, as reported, fully quantifies uncertainty arising from the counterfactual baseline, from misreported management or from structural differences among models, which are among the largest error sources identified in Sections 3 and 6. A systematic review of MRV systems for agricultural SOC emphasised that uncertainty is a core determinant of credibility and of the economic viability of projects (Wu & Lin, 2026). The present synthesis adds that uncertainty deductions provide only nominal conservativeness when they omit systematic error sources, because larger projects reduce random error but not bias.

7.2 Costs, Scale and the Economics of Verification

Cost is the principal argument for digital MRV. A cost model comparing decision-support tools, process-based models and direct soil measurement found substantial variation in MRV costs depending on the monitoring approach, project design and protocol standards, with potential burdens on farmers, and suggested that digitalisation and the use of long-term experimental sites could reduce costs, although policy support is likely to be needed for smaller farms (Vanderheyden *et al.*, 2026). The economics of both measurement and DSM improve strongly with project size, because spatial averaging reduces the uncertainty per unit area and fixed costs are distributed across more fields (Potash *et al.*, 2025; Wadoux *et al.*, 2026). In-field sensors can reduce analytical costs by an order of magnitude, but at the expense of accuracy that may not meet certification requirements (Kohl *et al.*, 2025).

The consequence is a structural bias in who can participate. Large, consolidated projects achieve lower deductions and lower costs per credit, whereas small and fragmented holdings face higher costs and greater uncertainty. An evaluation of nine SOC accounting approaches for a smallholder carbon project in western Kenya found that remote sensing could reduce the cost of direct and model-based measurement, but that its wider integration required transparency, investment in spectral libraries and user-friendly tools, and attention to the balance between community benefits and the detached nature of remote techniques (Okoli & Birkenberg, 2024). The same study suggested that handheld sensors and community participation in monitoring could lower costs and generate local benefits, provided that protocols were consistent. Research on farmer perceptions in the United States indicated that participants often regarded soil carbon programmes as convoluted, burdensome and unpredictable (Barbato & Strong, 2023). Digital MRV that lowers costs without simplifying participation or clarifying data ownership may therefore do little to broaden access.

7.3 Transparency, Data Infrastructure and Verification Independence

Verification in the strict sense requires that an independent party can reproduce or test the claimed quantity. For measurement-based MRV, this can be achieved by independent resampling. For digital MRV, it requires access to input data, model code, parameter sets, calibration and validation datasets and processing chains, many of which are proprietary in commercial systems. Proposals for a global soil information system have outlined integrated data–model frameworks supported by expanded measurement and monitoring networks, remote sensing and crowd-sourced management activity data (Paustian *et al.*, 2019), and a global MRV framework has been envisaged around repeat surveys, long-term experiments, benchmark sites and models (Smith *et al.*, 2020). Distributed-ledger solutions for carbon markets have been assessed as approaching technical maturity (Sipthorpe *et al.*, 2022). Such registries can secure the provenance and immutability of records, but they cannot validate the scientific estimates recorded in them, and they should not be mistaken for verification of SOC change.

Scientific infrastructure for independent evaluation is developing unevenly. Large public spectral libraries and calibration services (Shepherd *et al.*, 2022; Wijewardane *et al.*, 2018), open global soil maps with quantified uncertainty (Poggio *et al.*, 2021), open processing chains (Wijmer *et al.*, 2024) and field observatory networks (Nevalainen *et al.*, 2022) provide models for transparent practice. Ecosystem-scale flux observations from eddy covariance towers, together with remote sensing products and land-surface models, have been identified as underused sources of evidence for assessing nature-based climate solutions (Novick *et al.*, 2022), and long-term experiments underpin model calibration (Gurung *et al.*, 2020; Mathers *et al.*, 2023). What is largely missing is an open benchmark of remeasured SOC stock change on commercial fields, against which competing digital MRV systems could be tested independently.

7.4 Geographical Concentration of the Evidence Base

The evidence reviewed is concentrated in temperate regions, particularly the maize and soybean systems of the Midwestern United States and arable systems in western and central Europe, with smaller bodies of work from Australia, China and Brazil. Remote sensing retrieval of SOC performs worst where bare soil is rarely exposed, as in Mediterranean systems dominated by tree crops and winter cereals (van Wesemael *et al.*, 2024), and many tropical and smallholder systems share these characteristics together with small field sizes and frequent cloud cover. Continental-scale mapping in Africa using coarse-resolution satellite reflectance and spatially balanced field survey data has demonstrated the feasibility of large-area soil property mapping (Vågen *et al.*, 2016), and measurement error has been incorporated into DSM in western Cameroon (Takoutsing *et al.*, 2022), but evidence on field-level SOC change detection, model validation and hybrid MRV in smallholder systems remains scarce (Okoli & Birkenberg, 2024). Calibration domains for process-based models are similarly concentrated in regions with long-term experiments. The generalisability of performance metrics derived in data-rich temperate systems to soils with low initial SOC stocks, for which the highest relative sequestration rates have been reported (Minasny *et al.*, 2017), is therefore uncertain.

8. Synthesis: An Evidence-Tiered Architecture for Digital Soil Carbon MRV

The preceding sections support an integrated interpretation that can be summarised as an evidence-tiered architecture (Fig. 1). The architecture is developed from the present synthesis rather than adapted from an existing framework, although it is consistent with the system-of-systems concept (Guan *et al.*, 2023), the modular MRV approach (Batjes *et al.*, 2024) and the global MRV vision (Smith *et al.*, 2020). Its organising principle is that each evidence stream should be assigned the role for which the evidence is strongest, and that relationships among streams should be distinguished according to whether they have been demonstrated or are hypothesised.

At the foundation lies design-based ground reference: probability sampling at baseline and remeasurement, analysed by dry combustion or by quality-controlled spectroscopy with explicit calibration transfer, and calculated on an equivalent soil mass basis to the depth affected by management. This layer is the only independent reference for SOC stock change, and its reliability for aggregated project-level change is well established, provided that sampling is sufficiently dense and replicated and includes controls (Bradford *et al.*, 2023; Potash *et al.*, 2025). The sensing layer provides three distinct inputs: covariates for stratification and mapping, observations of management activity, and constraints on crop growth and carbon inputs. The evidence is strongest for the second and third of these roles and weakest for direct detection of SOC change from reflectance (Azzari *et al.*, 2019; van Wesemael *et al.*, 2024; Wijmer *et al.*, 2024). The modelling layer converts measurements and sensing into field-level and counterfactual estimates. Calibrated process-based models are established within their validation domains (Mathers *et al.*, 2023), machine learning is established for spatial prediction of stocks (Wadoux *et al.*, 2026), and hybrid models are promising but not yet independently validated for credited stock change (Liu *et al.*, 2024; Zhang *et al.*, 2024). The accounting layer combines estimates with baselines, uncertainty deductions and permanence provisions. The verification layer closes the loop through independent remeasurement, validation within declared domains of applicability and auditable data provenance.

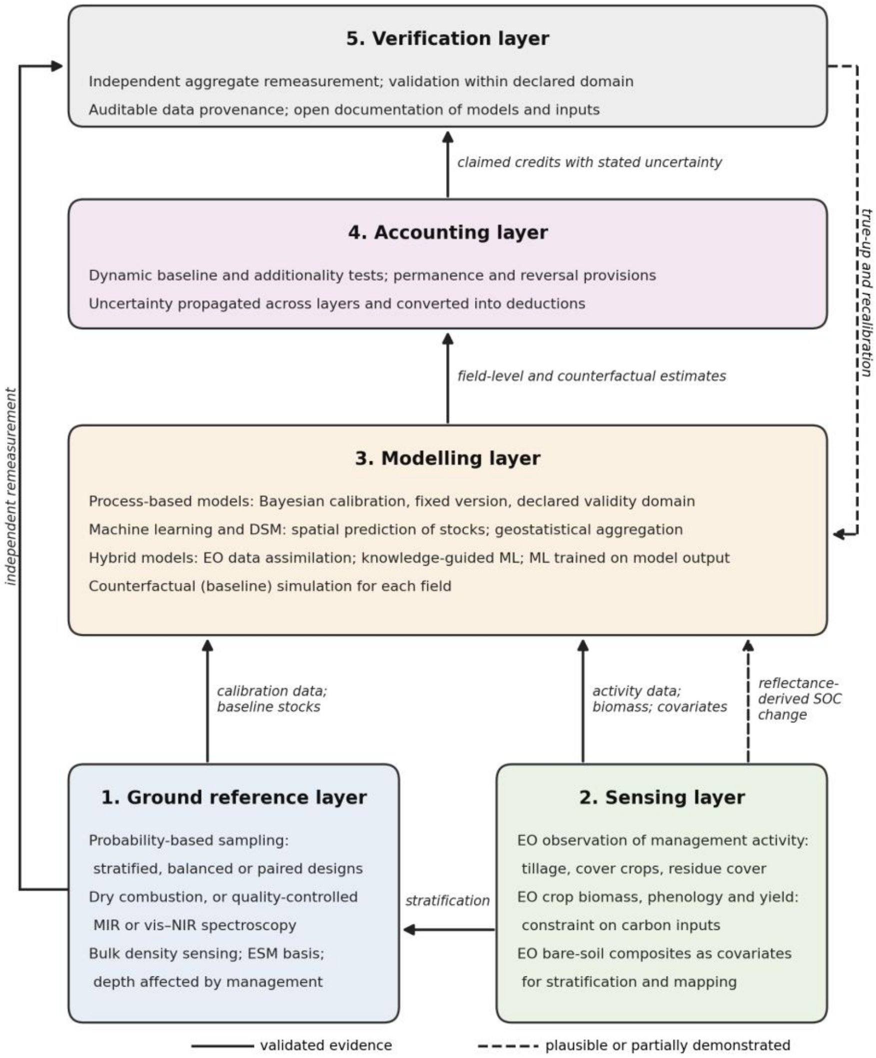

Fig. 1. Evidence-tiered architecture for digital monitoring, reporting and verification (MRV) of soil organic carbon (SOC), developed from the present synthesis. Boxes represent evidence layers and their principal components. Solid arrows denote relationships supported by validated evidence in the reviewed literature; dashed arrows denote relationships that are plausible or partially demonstrated but have not been validated consistently for crediting SOC stock change. The feedback arrow from verification to modelling represents recalibration or true-up of model-based estimates using design-based remeasurement, a relationship that is conceptually established but not yet evaluated at commercial scale. Abbreviations: DSM, digital soil mapping; EO, Earth observation; ESM, equivalent soil mass; MIR, mid-infrared; ML, machine learning; vis–NIR, visible–near-infrared

Three implications follow from this architecture. First, digital methods strengthen MRV most when they reduce the cost and increase the efficiency of design-based measurement, for example through stratification, spectroscopy and bulk density sensing, rather than when they replace it. Second, the credibility of field-level credits produced by models depends on periodic aggregate verification by remeasurement, with the ability to correct model-based estimates when discrepancies arise. Third, uncertainty must be propagated across layers rather than reported separately for each, because systematic errors in baselines, management data and model structure can dominate the random errors that current deductions address. Confidence in the architecture as a whole is limited by the scarcity of studies that have implemented all layers simultaneously and compared their outputs against independent remeasurement.

9. Future Directions

The research priorities below follow from the gaps identified in the synthesis and are ordered from immediate needs, which could be addressed with existing methods within the next few years, to longer-term needs that require methodological development or sustained observation. Table 5 summarises the priorities, the designs most likely to resolve them and the outcome measures by which progress should be judged.

Table 5. Prioritised research needs for credible digital soil carbon MRV.

Priority and horizon	Unresolved problem	Recommended design	Scale and timescale	Key outcome measures	Supporting references
Independent benchmark networks (immediate)	Model-based and hybrid estimates of SOC change lack independent validation on commercial fields	Randomised or matched practice assignment; paired georeferenced resampling; ESM calculation; deeper sampling on a subset	Several soil–climate regions; thousands of fields; remeasurement after 5–10 years	Bias and interval coverage of predicted change; minimum detectable change; cost per verified tonne	Bradford *et al.* (2023); Le Noë *et al.* (2023); Potash *et al.* (2025)
Complete uncertainty budgets (immediate)	Deductions omit baseline, management-data and structural errors	Factorial perturbation of inputs; multi-model comparison; propagation to project totals	Existing projects and research sites; 2–5 years	Variance share by source; realised coverage against remeasurement	Mathers *et al.* (2023); Riggers *et al.* (2019); W. Zhou *et al.* (2023)
Validation standards for AI (immediate)	Accuracy inflated by spatial dependence and extrapolation; validation of stocks rather than change	Design-based validation; temporal hold-out; area of applicability reporting; open benchmarks	Regional to continental datasets; ongoing	Accuracy of predicted change; proportion of fields inside area of applicability	Meyer and Pebesma (2021); Ploton *et al.* (2020); Wadoux *et al.* (2021)
Spectroscopic quality assurance (immediate)	Inter-instrument and in-field biases comparable to expected change	Ring trials; reference materials; documented calibration transfer	Multi-laboratory; 1–3 years	Inter-laboratory bias; equivalence with dry combustion	Kohl *et al.* (2025); Safanelli *et al.* (2023); Shepherd *et al.* (2022)
Process-constrained hybrid models (longer term)	Hybrids constrain inputs better than decomposition and subsoil dynamics	Assimilation of measurable fractions and profile data; out-of-region evaluation	Multiple regions; 10 years or more	Accuracy of change outside development domain; profile-level bias	Dangal *et al.* (2022); Liu *et al.* (2024); Wijmer *et al.* (2024)
Empirical counterfactual baselines (longer term)	Background trends and prior adoption inflate credited change	Satellite-derived adoption histories; matched controls; dynamic baselines; quasi-experimental analysis	Regional; historical reconstruction plus 5–10 years	Difference between static and dynamic baseline credits; share of change attributable to background trends	Ogle *et al.* (2023); Somenahally *et al.* (2025); Q. Zhou *et al.* (2022)
Under-represented systems (longer term)	Evidence concentrated in temperate, mechanised agriculture	Co-designed MRV; community sampling; handheld sensing validated against laboratory data	Smallholder and tropical landscapes; 5–10 years	Cost per participant; accuracy of change; participation and benefit distribution	Okoli and Birkenberg (2024); Takoutsing *et al.* (2022); Vågen *et al.* (2016)
Permanence and reversal monitoring (longer term)	Persistence after programme exit unknown; alternative accounting units untested	Long-term monitoring after exit; empirical tests of time-integrated accounting	Programme cohorts; decades	Reversal frequency and magnitude; divergence among accounting units	Dynarski *et al.* (2020); Leifeld (2023); Minasny and McBratney (2024)

Note. Abbreviations: AI, artificial intelligence; ESM, equivalent soil mass; MRV, monitoring, reporting and verification; SOC, soil organic carbon. Horizons distinguish needs addressable with existing methods (immediate) from those requiring methodological development or sustained observation (longer term)

The most immediate priority is an independent empirical benchmark for digital MRV. At present, model-based and hybrid estimates of SOC stock change are validated mainly against research plots, cross-validation within development datasets or other models, and independent time-series validation is scarce (Le Noë *et al.*, 2023; Mathers *et al.*, 2023). Networks of commercial fields with randomised or matched assignment of practices, paired georeferenced baseline and remeasurement sampling, analysis on an equivalent soil mass basis and deeper sampling on a representative subset would provide the reference against which competing systems could be tested (Bradford *et al.*, 2023; Potash *et al.*, 2025; von Haden *et al.*, 2024). Such networks should span several soil and climate regions and remeasure at intervals of five to ten years. The relevant outcomes are the bias and interval coverage of predicted project-level change, the minimum detectable change and the cost per verified tonne of carbon. If data and protocols were openly available, these networks would allow regulators to set evidence-based uncertainty deductions and would convert the current debate between measurement-based and model-based approaches into an empirical comparison.

A second immediate priority is the development of complete uncertainty budgets. Existing deductions reflect parameter, residual and spatial aggregation uncertainty (Mathers *et al.*, 2023; Wadoux *et al.*, 2026), whereas baseline data, management records and model structure have been shown to contribute large and partly systematic errors (Riggers *et al.*, 2019; W. Zhou *et al.*, 2023). Studies that deliberately perturb each input, compare multiple models and management data sources and propagate the resulting uncertainty to project-level credits would reveal which error sources dominate in practice. The appropriate outcome is the share of total uncertainty attributable to each source and the realised coverage of stated intervals against remeasured change. Standards could then require that deductions reflect the dominant sources rather than only those easiest to quantify.

A third immediate priority concerns validation standards for AI. The evidence that validation design can inflate apparent accuracy is strong (Ploton *et al.*, 2020; Wadoux *et al.*, 2021), and the need to delimit the area in which a model can be applied is well argued (Meyer & Pebesma, 2021). What is lacking is agreement on validation protocols specific to SOC change, including temporal hold-out, design-based validation with probability samples, reporting of the area of applicability and evaluation of change rather than stock. Community benchmarks with withheld test regions and years would allow these protocols to be tested. A parallel need exists for spectroscopic quality assurance, because inter-instrument and in-field biases can exceed expected changes (Kohl *et al.*, 2025; Safanelli *et al.*, 2023). Ring trials, reference materials and documented calibration transfer procedures specific to repeated MRV measurements would address this gap.

Longer-term priorities concern the process representation and scope of hybrid systems. Current hybrid models constrain carbon inputs and fluxes more effectively than decomposition and stabilisation, and they rarely represent subsoil dynamics (Wijmer *et al.*, 2024). Incorporating measurable fractions (Dangal *et al.*, 2022; Lavallee *et al.*, 2020) and deeper profiles into knowledge-guided and data assimilation frameworks (Liu *et al.*, 2024), and evaluating these frameworks against remeasured change outside their development regions, would test whether their apparent advantages survive transfer. This work requires observation periods of a decade or more and should be designed alongside the benchmark networks described above.

Counterfactual baselines represent a second longer-term need. Remote sensing now permits reconstruction of historical practice adoption at field scale (Seifert *et al.*, 2018; Q. Zhou *et al.*, 2022), and national counterfactual modelling has demonstrated the influence of past adoption on current SOC trends (Ogle *et al.*, 2023). Combining satellite-derived adoption histories with matched control fields, dynamic baselines from legacy data (Somenahally *et al.*, 2025) and quasi-experimental designs could provide empirically grounded baselines and tests of additionality. The key outcomes would be the difference between static and dynamic baseline credits and the proportion of credited change attributable to background trends.

Evaluation in under-represented systems is a third longer-term need and arguably the most consequential for equity. The evidence base is concentrated in temperate, mechanised agriculture, whereas smallholder and tropical systems face distinct constraints of field size, cloud cover, crop cover, data scarcity and cost (Okoli & Birkenberg, 2024). Studies that co-design MRV with farming communities, test handheld sensing and community-based sampling against laboratory reference data, and incorporate measurement error into mapping (Takoutsing *et al.*, 2022) would indicate whether digital MRV can extend participation or whether it will reinforce the advantages of large projects. Finally, the permanence debate would benefit from long-term monitoring of reversals after programme exit and from empirical evaluation of time-integrated accounting units (Leifeld, 2023; Minasny & McBratney, 2024), which require monitoring well beyond typical crediting periods and which digital methods are well placed to support once their accuracy for change has been established.

10. Conclusions

The review set out to define what a credible soil carbon claim must demonstrate, to appraise the contribution of remote sensing, proximal sensing, artificial intelligence and process-based models against those requirements and to evaluate the architectures that combine them. The most defensible conclusion is that credibility is a property of the whole MRV system rather than of any technology within it. A creditable claim requires an estimate of management-induced change in SOC stock relative to a counterfactual, calculated on a consistent soil mass and depth basis, accompanied by uncertainty that includes systematic as well as random error, and open to independent verification. None of the four evidence streams satisfies these requirements alone.

The evidence is strongest for a small number of findings. Direct measurement with probability sampling remains the only independent reference for SOC stock change and can estimate average effects across many fields reliably, but it cannot resolve change on individual fields over typical crediting intervals. Remote sensing is most reliable as an observer of management activity and crop growth and as a source of covariates for stratification, whereas direct retrieval of topsoil SOC from reflectance is confounded by surface conditions and has not been shown to detect management-induced change. Laboratory mid-infrared spectroscopy with appropriate calibration transfer can reduce analytical costs without unacceptable loss of accuracy, while in-field sensors currently suit mapping better than certification. Calibrated process-based models can meet stringent validation criteria within the domains in which they have been tested, but their validity outside those domains and on commercial fields remains uncertain.

Other conclusions are more tentative. Machine learning improves spatial prediction of SOC stocks, but reported accuracies are sensitive to validation design, and its capacity to represent management-driven temporal change is limited. Hybrid architectures that assimilate observations into process-based models or embed process knowledge in machine learning have shown clear gains for fluxes, biomass and stock prediction, yet their advantage for credited stock change has not been demonstrated against independent remeasurement. Claims that integration inherently reduces uncertainty should therefore be treated as hypotheses to be tested rather than established properties of digital MRV.

The broader significance of the synthesis lies in its reframing of the role of digital methods. Their greatest immediate value is to make design-based measurement cheaper, better targeted and more frequent, to document management reliably and to allocate aggregate verified change among fields, rather than to replace measurement with prediction. The central gap is the absence of open, independently remeasured benchmarks of SOC stock change on commercial fields in diverse regions. Until such benchmarks exist, confidence in digital MRV should remain proportional to the strength of design-based verification that accompanies it, and the geographic concentration of current evidence should temper expectations that performance achieved in data-rich temperate systems will transfer to smallholder and tropical agriculture.

11. Limitations

The first limitation is intrinsic to the narrative review design. Although the search, selection and appraisal procedures were described transparently and applied consistently, the identification of themes, the selection of studies to illustrate them and the grading of evidence involved judgement, and selection and interpretive bias cannot be excluded. Different reviewers might have weighted the same evidence differently, particularly for contested questions such as the feasibility of field-level detection and the permanence of soil carbon.

The search relied on openly accessible scholarly indexes and institutional sources, and subscription bibliographic databases were not searched. Some relevant studies indexed only in such databases may therefore have been missed, although citation tracing of recent reviews was used to reduce this risk. Only publications in English were included, which may have under-represented research published in other languages, including regional studies from Latin America, China and francophone Africa. Full texts of some articles were not accessible, and in those cases appraisal relied on abstracts and publisher records, which limited the depth of methodological evaluation for a minority of studies.

The evidence base itself imposes further limitations. Studies differed widely in design, spatial scale, depth, reference methods, performance metrics and validation strategies, which precluded quantitative synthesis and limited direct comparison of reported accuracies. The literature is geographically concentrated in temperate, mechanised agricultural systems, so that conclusions about smallholder and tropical systems rest on sparse evidence. Algorithm and model development studies may be subject to publication bias favouring positive results, and many commercial digital MRV systems are not documented in the peer-reviewed literature and could not be evaluated. Long-term evidence on SOC change under carbon farming, on reversals after programme exit and on the performance of digital MRV over complete crediting periods is scarce. Finally, the field is developing rapidly, and methodologies, regulations and technologies may have changed since the end of the search period.

Declaration of AI Use

This manuscript was prepared through the intellectual contributions of the author(s) to the study design, data analysis, interpretation of results, and scholarly content. Grammarly and ChatGPT were used solely to assist with grammatical refinement and reference formatting during manuscript revision. The author(s) accept full responsibility for the accuracy, integrity, and final content of the manuscript.

Competing Interests

Authors have declared that no competing interests exist.

References

Abramoff, R., Xu, X., Hartman, M., O'Brien, S., Feng, W., Davidson, E., Finzi, A., Moorhead, D., Schimel, J., Torn, M., & Mayes, M. A. (2018). The Millennial model: In search of measurable pools and transformations for modeling soil carbon in the new century. *Biogeochemistry, 137*(1–2), 51–71. https://doi.org/10.1007/s10533-017-0409-7

Angelopoulou, T., Tziolas, N., Balafoutis, A., Zalidis, G., & Bochtis, D. (2019). Remote sensing techniques for soil organic carbon estimation: A review. *Remote Sensing, 11*(6), Article 676. https://doi.org/10.3390/rs11060676

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

Badgley, G., Freeman, J., Hamman, J. J., Haya, B., Trugman, A. T., Anderegg, W. R. L., & Cullenward, D. (2022). Systematic over-crediting in California's forest carbon offsets program. *Global Change Biology, 28*(4), 1433–1445. https://doi.org/10.1111/gcb.15943

Baethge, C., Goldbeck-Wood, S., & Mertens, S. (2019). SANRA—A scale for the quality assessment of narrative review articles. *Research Integrity and Peer Review, 4*, Article 5. https://doi.org/10.1186/s41073-019-0064-8

Barbato, C. T., & Strong, A. L. (2023). Farmer perspectives on carbon markets incentivizing agricultural soil carbon sequestration. *npj Climate Action, 2*, Article 26. https://doi.org/10.1038/s44168-023-00055-4

Bartsch, B. dos A., Rosin, N. A., Rosas, J. T. F., Poppiel, R. R., Makino, F. Y., Vogel, L. G., Novais, J. J. M., Falcioni, R., Alves, M. R., & Demattê, J. A. M. (2025). Space-time mapping of soil organic carbon through remote sensing and machine learning. *Soil and Tillage Research, 248*, Article 106428. https://doi.org/10.1016/j.still.2024.106428

Batjes, N. H., Ceschia, E., Heuvelink, G. B. M., Demenois, J., le Maire, G., Cardinael, R., Arias-Navarro, C., & van Egmond, F. (2024). Towards a modular, multi-ecosystem monitoring, reporting and verification (MRV) framework for soil organic carbon stock change assessment. *Carbon Management, 15*(1), Article 2410812. https://doi.org/10.1080/17583004.2024.2410812

Bossio, D. A., Cook-Patton, S. C., Ellis, P. W., Fargione, J., Sanderman, J., Smith, P., Wood, S., Zomer, R. J., von Unger, M., Emmer, I. M., & Griscom, B. W. (2020). The role of soil carbon in natural climate solutions. *Nature Sustainability, 3*(5), 391–398. https://doi.org/10.1038/s41893-020-0491-z

Bradford, M. A., Eash, L., Polussa, A., Jevon, F. V., Kuebbing, S. E., Hammac, W. A., Rosenzweig, S., & Oldfield, E. E. (2023). Testing the feasibility of quantifying change in agricultural soil carbon stocks through empirical sampling. *Geoderma, 440*, Article 116719. https://doi.org/10.1016/j.geoderma.2023.116719

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

Campbell, E. E., & Paustian, K. (2015). Current developments in soil organic matter modeling and the expansion of model applications: A review. *Environmental Research Letters, 10*(12), Article 123004. https://doi.org/10.1088/1748-9326/10/12/123004

Carter, T. L., Schaecher, C., Monteith, S., & Ferguson, R. (2024). Using combustion analysis to simultaneously measure soil organic and inorganic carbon. *Geoderma, 451*, Article 117066. https://doi.org/10.1016/j.geoderma.2024.117066

Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., & van Wesemael, B. (2019). Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. *ISPRS Journal of Photogrammetry and Remote Sensing, 147*, 267–282. https://doi.org/10.1016/j.isprsjprs.2018.11.026

Castaldi, F., Syu, C.-H., Valdez, M. C., Yen, C.-C., Chen, C.-F., Bertinaria, F., & Toscano, P. (2026). Monitoring soil organic carbon from earth observation in the era of cover cropping and no-tillage. *Smart Agricultural Technology, 13*, Article 101935. https://doi.org/10.1016/j.atech.2026.101935

Ceschia, E., Béziat, P., Dejoux, J. F., Aubinet, M., Bernhofer, C., Bodson, B., Buchmann, N., Carrara, A., Cellier, P., Di Tommasi, P., Elbers, J. A., Eugster, W., Grünwald, T., Jacobs, C. M. J., Jans, W. W. P., Jones, M., Kutsch, W., Lanigan, G., Magliulo, E., … Wattenbach, M. (2010). Management effects on net ecosystem carbon and GHG budgets at European crop sites. *Agriculture, Ecosystems & Environment, 139*(3), 363–383. https://doi.org/10.1016/j.agee.2010.09.020

Chabrillat, S., Ben-Dor, E., Cierniewski, J., Gomez, C., Schmid, T., & van Wesemael, B. (2019). Imaging spectroscopy for soil mapping and monitoring. *Surveys in Geophysics, 40*(3), 361–399. https://doi.org/10.1007/s10712-019-09524-0

Chenu, C., Angers, D. A., Barré, P., Derrien, D., Arrouays, D., & Balesdent, J. (2019). Increasing organic stocks in agricultural soils: Knowledge gaps and potential innovations. *Soil and Tillage Research, 188*, 41–52. https://doi.org/10.1016/j.still.2018.04.011

Dangal, S. R. S., Schwalm, C., Cavigelli, M. A., Gollany, H. T., Jin, V. L., & Sanderman, J. (2022). Improving soil carbon estimates by linking conceptual pools against measurable carbon fractions in the DAYCENT model version 4.5. *Journal of Advances in Modeling Earth Systems, 14*(5), Article e2021MS002622. https://doi.org/10.1029/2021MS002622

de Bruin, S., Brus, D. J., Heuvelink, G. B. M., van Ebbenhorst Tengbergen, T., & Wadoux, A. M. J.-C. (2022). Dealing with clustered samples for assessing map accuracy by cross-validation. *Ecological Informatics, 69*, Article 101665. https://doi.org/10.1016/j.ecoinf.2022.101665

de Gruijter, J. J., McBratney, A. B., Minasny, B., Wheeler, I., Malone, B. P., & Stockmann, U. (2016). Farm-scale soil carbon auditing. *Geoderma, 265*, 120–130. https://doi.org/10.1016/j.geoderma.2015.11.010

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

Ding, Z., Liu, K., Grunwald, S., Smith, P., Ciais, P., Wang, B., Wadoux, A. M. J.-C., Ferreira, C., Karunaratne, S., Shurpali, N., Yin, X., Roberts, D., Madgett, O., Duncan, S., Zhou, M., Liu, Z., & Harrison, M. T. (2025). Advancing soil organic carbon prediction: A comprehensive review of technologies, AI, process-based and hybrid modelling approaches. *Advanced Science, 12*(31), Article e04152. https://doi.org/10.1002/advs.202504152

Don, A., Seidel, F., Leifeld, J., Kätterer, T., Martin, M., Pellerin, S., Emde, D., Seitz, D., & Chenu, C. (2024). Carbon sequestration in soils and climate change mitigation—Definitions and pitfalls. *Global Change Biology, 30*(1), Article e16983. https://doi.org/10.1111/gcb.16983

Dorantes, M. J., Fuentes, B. A., & Miller, D. M. (2022). Calibration set optimization and library transfer for soil carbon estimation using soil spectroscopy—A review. *Soil Science Society of America Journal, 86*(4), 879–903. https://doi.org/10.1002/saj2.20435

Dupla, X., Bonvin, E., Deluz, C., Lugassy, L., Verrecchia, E., Baveye, P. C., Grand, S., & Boivin, P. (2024). Are soil carbon credits empty promises? Shortcomings of current soil carbon quantification methodologies and improvement avenues. *Soil Use and Management, 40*(3), Article e13092. https://doi.org/10.1111/sum.13092

Dvorakova, K., Heiden, U., Pepers, K., Staats, G., van Os, G., & van Wesemael, B. (2023). Improving soil organic carbon predictions from a Sentinel–2 soil composite by assessing surface conditions and uncertainties. *Geoderma, 429*, Article 116128. https://doi.org/10.1016/j.geoderma.2022.116128

Dvorakova, K., Shi, P., Limbourg, Q., & van Wesemael, B. (2020). Soil organic carbon mapping from remote sensing: The effect of crop residues. *Remote Sensing, 12*(12), Article 1913. https://doi.org/10.3390/rs12121913

Dynarski, K. A., Bossio, D. A., & Scow, K. M. (2020). Dynamic stability of soil carbon: Reassessing the "permanence" of soil carbon sequestration. *Frontiers in Environmental Science, 8*, Article 514701. https://doi.org/10.3389/fenvs.2020.514701

Ellili, Y., Walter, C., Michot, D., Pichelin, P., & Lemercier, B. (2019). Mapping soil organic carbon stock change by soil monitoring and digital soil mapping at the landscape scale. *Geoderma, 351*, 1–8. https://doi.org/10.1016/j.geoderma.2019.03.005

England, J. R., & Viscarra Rossel, R. A. (2018). Proximal sensing for soil carbon accounting. *SOIL, 4*(2), 101–122. https://doi.org/10.5194/soil-4-101-2018

European Parliament & Council of the European Union. (2024). *Regulation (EU) 2024/3012 of the European Parliament and of the Council of 27 November 2024 establishing a Union certification framework for permanent carbon removals, carbon farming and carbon storage in products*. *Official Journal of the European Union*. http://data.europa.eu/eli/reg/2024/3012/oj

Farina, R., Sándor, R., Abdalla, M., Álvaro-Fuentes, J., Bechini, L., Bolinder, M. A., Brilli, L., Chenu, C., Clivot, H., De Antoni Migliorati, M., Di Bene, C., Dorich, C. D., Ehrhardt, F., Ferchaud, F., Fitton, N., Francaviglia, R., Franko, U., Giltrap, D. L., Grant, B. B., … Bellocchi, G. (2021). Ensemble modelling, uncertainty and robust predictions of organic carbon in long-term bare-fallow soils. *Global Change Biology, 27*(4), 904–928. https://doi.org/10.1111/gcb.15441

Food and Agriculture Organization of the United Nations. (2020). *A protocol for measurement, monitoring, reporting and verification of soil organic carbon in agricultural landscapes: GSOC-MRV protocol*. FAO. https://doi.org/10.4060/cb0509en

Fowler, A. F., Basso, B., Millar, N., & Brinton, W. F. (2023). A simple soil mass correction for a more accurate determination of soil carbon stock changes. *Scientific Reports, 13*, Article 2242. https://doi.org/10.1038/s41598-023-29289-2

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

Greenhalgh, T., Thorne, S., & Malterud, K. (2018). Time to challenge the spurious hierarchy of systematic over narrative reviews? *European Journal of Clinical Investigation, 48*(6), Article e12931. https://doi.org/10.1111/eci.12931

Guan, K., Jin, Z., Peng, B., Tang, J., DeLucia, E. H., West, P. C., Jiang, C., Wang, S., Kim, T., Zhou, W., Griffis, T., Liu, L., Yang, W. H., Qin, Z., Yang, Q., Margenot, A., Stuchiner, E. R., Kumar, V., Bernacchi, C., … Yang, S.-J. (2023). A scalable framework for quantifying field-level agricultural carbon outcomes. *Earth-Science Reviews, 243*, Article 104462. https://doi.org/10.1016/j.earscirev.2023.104462

Gurung, R. B., Ogle, S. M., Breidt, F. J., Williams, S. A., & Parton, W. J. (2020). Bayesian calibration of the DayCent ecosystem model to simulate soil organic carbon dynamics and reduce model uncertainty. *Geoderma, 376*, Article 114529. https://doi.org/10.1016/j.geoderma.2020.114529

Heuvelink, G. B. M., Angelini, M. E., Poggio, L., Bai, Z., Batjes, N. H., van den Bosch, R., Bossio, D., Estella, S., Lehmann, J., Olmedo, G. F., & Sanderman, J. (2021). Machine learning in space and time for modelling soil organic carbon change. *European Journal of Soil Science, 72*(4), 1607–1623. https://doi.org/10.1111/ejss.12998

Hively, W. D., Lamb, B. T., Daughtry, C. S. T., Shermeyer, J., McCarty, G. W., & Quemada, M. (2018). Mapping crop residue and tillage intensity using WorldView-3 satellite shortwave infrared residue indices. *Remote Sensing, 10*(10), Article 1657. https://doi.org/10.3390/rs10101657

Ihasusta, A., Al Bitar, A., Batjes, N. H., van Egmond, F., Cardinael, R., Karunaratne, S., Barré, P., Arnaud, L., Heuvelink, G., Macdonald, B., Wijmer, T., Demenois, J., Paul-Victor, C., Reynders, S., & Ceschia, E. (2026). Choosing the appropriate methodology to monitor soil organic carbon (SOC) in croplands: Aligning methods with evolving monitoring reporting verification (MRV) frameworks. *Carbon Management, 17*(1), Article 2638317. https://doi.org/10.1080/17583004.2026.2638317

Intergovernmental Panel on Climate Change. (2019). *2019 refinement to the 2006 IPCC guidelines for national greenhouse gas inventories: Vol. 4. Agriculture, forestry and other land use* (E. Calvo Buendia, K. Tanabe, A. Kranjc, J. Baasansuren, M. Fukuda, S. Ngarize, A. Osako, Y. Pyrozhenko, P. Shermanau, & S. Federici, Eds.). IPCC. https://www.ipcc-nggip.iges.or.jp/public/2019rf/vol4.html

Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., & Wang, J. (2018). A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy, 92*, 141–152. https://doi.org/10.1016/j.eja.2017.11.002

Kohl, L., Vielhauer, C., Öztürk, A., Minarsch, E.-M. L., Ahl, C., Niether, W., Clifton-Brown, J., & Gattinger, A. (2025). Field evaluation of a portable multi-sensor soil carbon analyzer: Performance, precision, and limitations under real-world conditions. *Soil Systems, 9*(3), Article 67. https://doi.org/10.3390/soilsystems9030067

Kravchenko, A. N., & Robertson, G. P. (2011). Whole-profile soil carbon stocks: The danger of assuming too much from analyses of too little. *Soil Science Society of America Journal, 75*(1), 235–240. https://doi.org/10.2136/sssaj2010.0076

Lavallee, J. M., Soong, J. L., & Cotrufo, M. F. (2020). Conceptualizing soil organic matter into particulate and mineral-associated forms to address global change in the 21st century. *Global Change Biology, 26*(1), 261–273. https://doi.org/10.1111/gcb.14859

Le Noë, J., Manzoni, S., Abramoff, R., Bölscher, T., Bruni, E., Cardinael, R., Ciais, P., Chenu, C., Clivot, H., Derrien, D., Ferchaud, F., Garnier, P., Goll, D., Lashermes, G., Martin, M., Rasse, D., Rees, F., Sainte-Marie, J., Salmon, E., … Guenet, B. (2023). Soil organic carbon models need independent time-series validation for reliable prediction. *Communications Earth & Environment, 4*, Article 158. https://doi.org/10.1038/s43247-023-00830-5

Leifeld, J. (2023). Carbon farming: Climate change mitigation via non-permanent carbon sinks. *Journal of Environmental Management, 339*, Article 117893. https://doi.org/10.1016/j.jenvman.2023.117893

Li, T., Cui, L., Kuhnert, M., McLaren, T. I., Pandey, R., Liu, H., Wang, W., Xu, Z., Xia, A., Dalal, R. C., & Dang, Y. P. (2024). A comprehensive review of soil organic carbon estimates: Integrating remote sensing and machine learning technologies. *Journal of Soils and Sediments, 24*, 3556–3571. https://doi.org/10.1007/s11368-024-03913-8

Lima, A. A. J., Lopes, J. C., Lopes, R. P., de Figueiredo, T., Vidal-Vázquez, E., & Hernández, Z. (2025). Soil organic carbon assessment using remote-sensing data and machine learning: A systematic literature review. *Remote Sensing, 17*(5), Article 882. https://doi.org/10.3390/rs17050882

Liu, L., Zhou, W., Guan, K., Peng, B., Xu, S., Tang, J., Zhu, Q., Till, J., Jia, X., Jiang, C., Wang, S., Qin, Z., Kong, H., Grant, R., Mezbahuddin, S., Kumar, V., & Jin, Z. (2024). Knowledge-guided machine learning can improve carbon cycle quantification in agroecosystems. *Nature Communications, 15*, Article 357. https://doi.org/10.1038/s41467-023-43860-5

Lobsey, C. R., & Viscarra Rossel, R. A. (2016). Sensing of soil bulk density for more accurate carbon accounting. *European Journal of Soil Science, 67*(4), 504–513. https://doi.org/10.1111/ejss.12355

Luo, Z., Wang, E., & Sun, O. J. (2010). Can no-tillage stimulate carbon sequestration in agricultural soils? A meta-analysis of paired experiments. *Agriculture, Ecosystems & Environment, 139*(1–2), 224–231. https://doi.org/10.1016/j.agee.2010.08.006

Mathers, C., Black, C. K., Segal, B. D., Gurung, R. B., Zhang, Y., Easter, M. J., Williams, S., Motew, M., Campbell, E. E., Brummitt, C. D., Paustian, K., & Kumar, A. A. (2023). Validating DayCent-CR for cropland soil carbon offset reporting at a national scale. *Geoderma, 438*, Article 116647. https://doi.org/10.1016/j.geoderma.2023.116647

McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma, 117*(1–2), 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

Meyer, H., & Pebesma, E. (2021). Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods in Ecology and Evolution, 12*(9), 1620–1633. https://doi.org/10.1111/2041-210X.13650

Meyer, H., & Pebesma, E. (2022). Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nature Communications, 13*, Article 2208. https://doi.org/10.1038/s41467-022-29838-9

Minasny, B., & McBratney, A. B. (2024). Soil carbon tonne-year accounting: Crediting the additional time-integrated amount of carbon captured in soil. *Soil Security, 15*, Article 100153. https://doi.org/10.1016/j.soisec.2024.100153

Minasny, B., Malone, B. P., McBratney, A. B., Angers, D. A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B. S., Field, D. J., Gimona, A., Hedley, C. B., Hong, S. Y., Mandal, B., Marchant, B. P., Martin, M., McConkey, B. G., Mulder, V. L., … Winowiecki, L. (2017). Soil carbon 4 per mille. *Geoderma, 292*, 59–86. https://doi.org/10.1016/j.geoderma.2017.01.002

Moinet, G. Y. K., Hijbeek, R., van Vuuren, D. P., & Giller, K. E. (2023). Carbon for soils, not soils for carbon. *Global Change Biology, 29*(9), 2384–2398. https://doi.org/10.1111/gcb.16570

Nevalainen, O., Niemitalo, O., Fer, I., Juntunen, A., Mattila, T., Koskela, O., Kukkamäki, J., Höckerstedt, L., Mäkelä, L., Jarva, P., Heimsch, L., Vekuri, H., Kulmala, L., Stam, Å., Kuusela, O., Gerin, S., Viskari, T., Vira, J., Hyväluoma, J., … Liski, J. (2022). Towards agricultural soil carbon monitoring, reporting, and verification through the Field Observatory Network (FiON). *Geoscientific Instrumentation, Methods and Data Systems, 11*(1), 93–109. https://doi.org/10.5194/gi-11-93-2022

Ng, W., Minasny, B., Jones, E., & McBratney, A. (2022). To spike or to localize? Strategies to improve the prediction of local soil properties using regional spectral library. *Geoderma, 406*, Article 115501. https://doi.org/10.1016/j.geoderma.2021.115501

Novick, K. A., Metzger, S., Anderegg, W. R. L., Barnes, M., Cala, D. S., Guan, K., Hemes, K. S., Hollinger, D. Y., Kumar, J., Litvak, M., Lombardozzi, D., Normile, C. P., Oikawa, P., Runkle, B. R. K., Torn, M., & Wiesner, S. (2022). Informing nature-based climate solutions for the United States with the best-available science. *Global Change Biology, 28*(12), 3778–3794. https://doi.org/10.1111/gcb.16156

Ogle, S. M., Alsaker, C., Baldock, J., Bernoux, M., Breidt, F. J., McConkey, B., Regina, K., & Vazquez-Amabile, G. G. (2019). Climate and soil characteristics determine where no-till management can store carbon in soils and mitigate greenhouse gas emissions. *Scientific Reports, 9*, Article 11665. https://doi.org/10.1038/s41598-019-47861-7

Ogle, S. M., Breidt, F. J., Del Grosso, S., Gurung, R., Marx, E., Spencer, S., Williams, S., & Manning, D. (2023). Counterfactual scenarios reveal historical impact of cropland management on soil organic carbon stocks in the United States. *Scientific Reports, 13*, Article 14564. https://doi.org/10.1038/s41598-023-41307-x

Ogle, S. M., Breidt, F. J., Easter, M., Williams, S., Killian, K., & Paustian, K. (2010). Scale and uncertainty in modeled soil organic carbon stock changes for US croplands using a process-based model. *Global Change Biology, 16*(2), 810–822. https://doi.org/10.1111/j.1365-2486.2009.01951.x

Okoli, A. O., & Birkenberg, A. (2024). Monitoring soil carbon in smallholder carbon projects: Insights from Kenya. *Climatic Change, 177*(9), Article 143. https://doi.org/10.1007/s10584-024-03796-1

Oldfield, E. E., Eagle, A. J., Rubin, R. L., Rudek, J., Sanderman, J., & Gordon, D. R. (2022). Crediting agricultural soil carbon sequestration. *Science, 375*(6586), 1222–1225. https://doi.org/10.1126/science.abl7991

Padarian, J., Minasny, B., & McBratney, A. B. (2019). Using deep learning for digital soil mapping. *SOIL, 5*(1), 79–89. https://doi.org/10.5194/soil-5-79-2019

Paul, C., Bartkowski, B., Dönmez, C., Don, A., Mayer, S., Steffens, M., Weigl, S., Wiesmeier, M., Wolf, A., & Helming, K. (2023). Carbon farming: Are soil carbon certificates a suitable tool for climate change mitigation? *Journal of Environmental Management, 330*, Article 117142. https://doi.org/10.1016/j.jenvman.2022.117142

Paustian, K., Collier, S., Baldock, J., Burgess, R., Creque, J., DeLonge, M., Dungait, J., Ellert, B., Frank, S., Goddard, T., Govaerts, B., Grundy, M., Henning, M., Izaurralde, R. C., Madaras, M., McConkey, B., Porzig, E., Rice, C., Searle, R., … Jahn, M. (2019). Quantifying carbon for agricultural soil management: From the current status toward a global soil information system. *Carbon Management, 10*(6), 567–587. https://doi.org/10.1080/17583004.2019.1633231

Paustian, K., Lehmann, J., Ogle, S., Reay, D., Robertson, G. P., & Smith, P. (2016). Climate-smart soils. *Nature, 532*(7597), 49–57. https://doi.org/10.1038/nature17174

Pique, G., Fieuzal, R., Al Bitar, A., Veloso, A., Tallec, T., Brut, A., Ferlicoq, M., Zawilski, B., Dejoux, J.-F., Gibrin, H., & Ceschia, E. (2020). Estimation of daily CO_2 fluxes and of the components of the carbon budget for winter wheat by the assimilation of Sentinel 2-like remote sensing data into a crop model. *Geoderma, 376*, Article 114428. https://doi.org/10.1016/j.geoderma.2020.114428

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., Dormann, C., Cornu, G., Viennois, G., Bayol, N., Lyapustin, A., Gourlet-Fleury, S., & Pélissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications, 11*, Article 4540. https://doi.org/10.1038/s41467-020-18321-y

Poeplau, C., & Don, A. (2015). Carbon sequestration in agricultural soils via cultivation of cover crops – A meta-analysis. *Agriculture, Ecosystems & Environment, 200*, 33–41. https://doi.org/10.1016/j.agee.2014.10.024

Poeplau, C., Prietz, R., & Don, A. (2022). Plot-scale variability of organic carbon in temperate agricultural soils—Implications for soil monitoring. *Journal of Plant Nutrition and Soil Science, 185*(3), 403–416. https://doi.org/10.1002/jpln.202100393

Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *SOIL, 7*(1), 217–240. https://doi.org/10.5194/soil-7-217-2021

Potash, E., Bradford, M. A., Oldfield, E. E., & Guan, K. (2025). Measure-and-remeasure as an economically feasible approach to crediting soil organic carbon at scale. *Environmental Research Letters, 20*(2), Article 024025. https://doi.org/10.1088/1748-9326/ada16c

Potash, E., Guan, K., Margenot, A., Lee, D., DeLucia, E., Wang, S., & Jang, C. (2022). How to estimate soil organic carbon stocks of agricultural fields? Perspectives using ex-ante evaluation. *Geoderma, 411*, Article 115693. https://doi.org/10.1016/j.geoderma.2021.115693

Powlson, D. S., Stirling, C. M., Jat, M. L., Gerard, B. G., Palm, C. A., Sanchez, P. A., & Cassman, K. G. (2014). Limited potential of no-till agriculture for climate change mitigation. *Nature Climate Change, 4*(8), 678–683. https://doi.org/10.1038/nclimate2292

Probst, B. S., Toetzke, M., Kontoleon, A., Díaz Anadón, L., Minx, J. C., Haya, B. K., Schneider, L., Trotter, P. A., West, T. A. P., Gill-Wiehl, A., & Hoffmann, V. H. (2024). Systematic assessment of the achieved emission reductions of carbon crediting projects. *Nature Communications, 15*, Article 9562. https://doi.org/10.1038/s41467-024-53645-z

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature, 566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Riggers, C., Poeplau, C., Don, A., Bamminger, C., Höper, H., & Dechow, R. (2019). Multi-model ensemble improved the prediction of trends in soil organic carbon stocks in German croplands. *Geoderma, 345*, 17–30. https://doi.org/10.1016/j.geoderma.2019.03.014

Rogge, D., Bauer, A., Zeidler, J., Mueller, A., Esch, T., & Heiden, U. (2018). Building an exposed soil composite processor (SCMaP) for mapping spatial and temporal characteristics of soils with Landsat imagery (1984–2014). *Remote Sensing of Environment, 205*, 1–17. https://doi.org/10.1016/j.rse.2017.11.004

Rutjens, R. J. L., Reinke, K., Addinsall, M., Bremner, A., & Grover, S. (2026). Alternatives to equivalent soil mass in monitoring, reporting and verification of changes in soil carbon. *Soil Use and Management, 42*(2), Article e70213. https://doi.org/10.1111/sum.70213

Saby, N. P. A., Bellamy, P. H., Morvan, X., Arrouays, D., Jones, R. J. A., Verheijen, F. G. A., Kibblewhite, M. G., Verdoodt, A., Berényi Üveges, J., Freudenschuß, A., & Simota, C. (2008). Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content? *Global Change Biology, 14*(10), 2432–2442. https://doi.org/10.1111/j.1365-2486.2008.01658.x

Safanelli, J. L., Sanderman, J., Bloom, D., Todd-Brown, K., Parente, L. L., Hengl, T., Adam, S., Albinet, F., Ben-Dor, E., Boot, C. M., Bridson, J. H., Chabrillat, S., Deiss, L., Demattê, J. A. M., Demyan, M. S., Dercon, G., Doetterl, S., van Egmond, F., Ferguson, R., … Żelazny, W. R. (2023). An interlaboratory comparison of mid-infrared spectra acquisition: Instruments and procedures matter. *Geoderma, 440*, Article 116724. https://doi.org/10.1016/j.geoderma.2023.116724

Sanderman, J., Gholizadeh, A., Pittaki-Chrysodonta, Z., Huang, J., Safanelli, J. L., & Ferguson, R. (2023). Transferability of a large mid-infrared soil spectral library between two Fourier-transform infrared spectrometers. *Soil Science Society of America Journal, 87*(3), 586–599. https://doi.org/10.1002/saj2.20513

Sanderman, J., Savage, K., & Dangal, S. R. S. (2020). Mid-infrared spectroscopy for prediction of soil health indicators in the United States. *Soil Science Society of America Journal, 84*(1), 251–261. https://doi.org/10.1002/saj2.20009

Schlesinger, W. H. (2022). Biogeochemical constraints on climate change mitigation through regenerative farming. *Biogeochemistry, 161*(1), 9–17. https://doi.org/10.1007/s10533-022-00942-8

Seifert, C. A., Azzari, G., & Lobell, D. B. (2018). Satellite detection of cover crops and their effects on crop yield in the Midwestern United States. *Environmental Research Letters, 13*(6), Article 064033. https://doi.org/10.1088/1748-9326/aac4c8

Senesi, G. S., & Senesi, N. (2016). Laser-induced breakdown spectroscopy (LIBS) to measure quantitatively soil carbon with emphasis on soil organic carbon. A review. *Analytica Chimica Acta, 938*, 7–17. https://doi.org/10.1016/j.aca.2016.07.039

Shen, Z., Ramirez-Lopez, L., Behrens, T., Cui, L., Zhang, M., Walden, L., Wetterlind, J., Shi, Z., Sudduth, K. A., Baumann, P., Song, Y., Catambay, K., & Viscarra Rossel, R. A. (2022). Deep transfer learning of global spectra for local soil carbon monitoring. *ISPRS Journal of Photogrammetry and Remote Sensing, 188*, 190–200. https://doi.org/10.1016/j.isprsjprs.2022.04.009

Shepherd, K. D., & Walsh, M. G. (2002). Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal, 66*(3), 988–998. https://doi.org/10.2136/sssaj2002.9880

Shepherd, K. D., Ferguson, R., Hoover, D., van Egmond, F., Sanderman, J., & Ge, Y. (2022). A global soil spectral calibration library and estimation service. *Soil Security, 7*, Article 100061. https://doi.org/10.1016/j.soisec.2022.100061

Sipthorpe, A., Brink, S., Van Leeuwen, T., & Staffell, I. (2022). Blockchain solutions for carbon markets are nearing maturity. *One Earth, 5*(7), 779–791. https://doi.org/10.1016/j.oneear.2022.06.004

Smith, P. (2004). How long before a change in soil organic carbon can be detected? *Global Change Biology, 10*(11), 1878–1883. https://doi.org/10.1111/j.1365-2486.2004.00854.x

Smith, P., Davies, C. A., Ogle, S., Zanchi, G., Bellarby, J., Bird, N., Boddey, R. M., McNamara, N. P., Powlson, D., Cowie, A., van Noordwijk, M., Davis, S. C., Richter, D. deB., Kryzanowski, L., van Wijk, M. T., Stuart, J., Kirton, A., Eggar, D., Newton-Cross, G., … Braimoh, A. K. (2012). Towards an integrated global framework to assess the impacts of land use and management change on soil carbon: Current capability and future vision. *Global Change Biology, 18*(7), 2089–2101. https://doi.org/10.1111/j.1365-2486.2012.02689.x

Smith, P., Soussana, J.-F., Angers, D., Schipper, L., Chenu, C., Rasse, D. P., Batjes, N. H., van Egmond, F., McNeill, S., Kuhnert, M., Arias-Navarro, C., Olesen, J. E., Chirinda, N., Fornara, D., Wollenberg, E., Álvaro-Fuentes, J., Sanz-Cobena, A., & Klumpp, K. (2020). How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. *Global Change Biology, 26*(1), 219–241. https://doi.org/10.1111/gcb.14815

Somenahally, A. C., Bokati, L., & Kumar, S. (2025). Estimating soil organic carbon deficits at the continental scale using legacy-data-driven dynamic baseline and attainable projections. *Geoderma, 462*, Article 117515. https://doi.org/10.1016/j.geoderma.2025.117515

Stanley, P., Spertus, J., Chiartas, J., Stark, P. B., & Bowles, T. (2023). Valid inferences about soil carbon in heterogeneous landscapes. *Geoderma, 430*, Article 116323. https://doi.org/10.1016/j.geoderma.2022.116323

Stevens, A., Nocita, M., Tóth, G., Montanarella, L., & van Wesemael, B. (2013). Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy. *PLoS ONE, 8*(6), Article e66409. https://doi.org/10.1371/journal.pone.0066409

Sukhera, J. (2022). Narrative reviews: Flexible, rigorous, and practical. *Journal of Graduate Medical Education, 14*(4), 414–417. https://doi.org/10.4300/JGME-D-22-00480.1

Szatmári, G., & Pásztor, L. (2025). The impact of direct and indirect digital soil mapping approaches on spatial uncertainty. *Geoderma, 460*, Article 117448. https://doi.org/10.1016/j.geoderma.2025.117448

Szatmári, G., Pásztor, L., & Heuvelink, G. B. M. (2021). Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics. *Geoderma, 403*, Article 115356. https://doi.org/10.1016/j.geoderma.2021.115356

Takoutsing, B., Heuvelink, G. B. M., Stoorvogel, J. J., Shepherd, K. D., & Aynekulu, E. (2022). Accounting for analytical and proximal soil sensing errors in digital soil mapping. *European Journal of Soil Science, 73*(2), Article e13226. https://doi.org/10.1111/ejss.13226

Tao, F., Huang, Y., Hungate, B. A., Manzoni, S., Frey, S. D., Schmidt, M. W. I., Reichstein, M., Carvalhais, N., Ciais, P., Jiang, L., Lehmann, J., Wang, Y.-P., Houlton, B. Z., Ahrens, B., Mishra, U., Hugelius, G., Hocking, T. D., Lu, X., Shi, Z., … Luo, Y. (2023). Microbial carbon use efficiency promotes global soil carbon storage. *Nature, 618*(7967), 981–985. https://doi.org/10.1038/s41586-023-06042-3

Tautges, N. E., Chiartas, J. L., Gaudin, A. C. M., O'Geen, A. T., Herrera, I., & Scow, K. M. (2019). Deep soil inventories reveal that impacts of cover crops and compost on soil carbon sequestration differ in surface and subsurface soils. *Global Change Biology, 25*(11), 3753–3766.
https://doi.org/10.1111/gcb.14762

Thamo, T., & Pannell, D. J. (2016). Challenges in developing effective policy for soil carbon sequestration: Perspectives on additionality, leakage, and permanence. *Climate Policy, 16*(8), 973–992. https://doi.org/10.1080/14693062.2015.1075372

Urbina-Salazar, D., Vaudour, E., Richer-de-Forges, A. C., Chen, S., Martelet, G., Baghdadi, N., & Arrouays, D. (2023). Sentinel-2 and Sentinel-1 bare soil temporal mosaics of 6-year periods for soil organic carbon content mapping in central France. *Remote Sensing, 15*(9), Article 2410. https://doi.org/10.3390/rs15092410

Vågen, T.-G., Winowiecki, L. A., Tondoh, J. E., Desta, L. T., & Gumbricht, T. (2016). Mapping of soil properties and land degradation risk in Africa using MODIS reflectance. *Geoderma, 263*, 216–225. https://doi.org/10.1016/j.geoderma.2015.06.023

van Wesemael, B., Abdelbaki, A., Ben-Dor, E., Chabrillat, S., d'Angelo, P., Demattê, J. A. M., Genova, G., Gholizadeh, A., Heiden, U., Karlshoefer, P., Milewski, R., Poggio, L., Sabetizade, M., Sanz, A., Schwind, P., Tsakiridis, N., Tziolas, N., Yagüe, J., & Žížala, D. (2024). A European soil organic carbon monitoring system leveraging Sentinel 2 imagery and the LUCAS soil data base. *Geoderma, 452*, Article 117113.
https://doi.org/10.1016/j.geoderma.2024.117113

Vanderheyden, L., Gayot, N., & Van den Broeck, G. (2026). What monitoring, reporting & verification (MRV) systems can reduce costs and enhance scalability of carbon farming? *Journal of Environmental Management, 401*, Article 128885. https://doi.org/10.1016/j.jenvman.2026.128885

Vaudour, E., Gholizadeh, A., Castaldi, F., Saberioon, M., Borůvka, L., Urbina-Salazar, D., Fouad, Y., Arrouays, D., Richer-de-Forges, A. C., Biney, J., Wetterlind, J., & van Wesemael, B. (2022). Satellite imagery to map topsoil organic carbon content over cultivated areas: An overview. *Remote Sensing, 14*(12), Article 2917.
https://doi.org/10.3390/rs14122917

Vaudour, E., Gomez, C., Loiseau, T., Baghdadi, N., Loubet, B., Arrouays, D., Ali, L., & Lagacherie, P. (2019). The impact of acquisition date on the prediction performance of topsoil organic carbon from Sentinel-2 for croplands. *Remote Sensing, 11*(18), Article 2143. https://doi.org/10.3390/rs11182143

Verra. (2025). *VM0042 improved agricultural land management* (Version 2.2) [Methodology]. https://verra.org/methodologies/vm0042-improved-agricultural-land-management-v2-2/

Viscarra Rossel, R. A., Behrens, T., Ben-Dor, E., Brown, D. J., Demattê, J. A. M., Shepherd, K. D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B. G., Bartholomeus, H. M., Bayer, A. D., Bernoux, M., Böttcher, K., Brodský, L., Du, C. W., Chappell, A., … Ji, W. (2016). A global spectral library to characterize the world's soil. *Earth-Science Reviews, 155*, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012

Viscarra Rossel, R. A., Lobsey, C. R., Sharman, C., Flick, P., & McLachlan, G. (2017). Novel proximal sensing for monitoring soil organic C stocks and condition. *Environmental Science & Technology, 51*(10), 5630–5641. https://doi.org/10.1021/acs.est.7b00889

Viscarra Rossel, R. A., Walvoort, D. J. J., McBratney, A. B., Janik, L. J., & Skjemstad, J. O. (2006). Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma, 131*(1–2), 59–75. https://doi.org/10.1016/j.geoderma.2005.03.007

von Haden, A. C., Sanford, G. R., Cates, A. M., & Jackson, R. D. (2024). Paired resampling to detect field-level soil organic carbon stock change. Comment on "Testing the feasibility of quantifying change in agricultural soil carbon stocks through empirical sampling" by Bradford *et al. Geoderma, 448*, Article 116959. https://doi.org/10.1016/j.geoderma.2024.116959

von Haden, A. C., Yang, W. H., & DeLucia, E. H. (2020). Soils' dirty little secret: Depth-based comparisons can be inadequate for quantifying changes in soil organic carbon and other mineral soil properties. *Global Change Biology, 26*(7), 3759–3770. https://doi.org/10.1111/gcb.15124

Wadoux, A. M. J.-C., & Molnar, C. (2022). Beyond prediction: Methods for interpreting complex models of soil variation. *Geoderma, 422*, Article 115953. https://doi.org/10.1016/j.geoderma.2022.115953

Wadoux, A. M. J.-C., Donovan, M., Fu, P., Leach, N. R., Maddalena, J., Rustowicz, R., Schurman, D., & Kellner, J. R. (2026). A digital soil mapping approach to soil carbon monitoring, reporting and verification (MRV). *npj Sustainable Agriculture, 4*, Article 58. https://doi.org/10.1038/s44264-026-00125-0

Wadoux, A. M. J.-C., Heuvelink, G. B. M., de Bruin, S., & Brus, D. J. (2021). Spatial cross-validation is not the right way to evaluate map accuracy. *Ecological Modelling, 457*, Article 109692. https://doi.org/10.1016/j.ecolmodel.2021.109692

Wadoux, A. M. J.-C., Minasny, B., & McBratney, A. B. (2020). Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews, 210*, Article 103359. https://doi.org/10.1016/j.earscirev.2020.103359

Wendt, J. W., & Hauser, S. (2013). An equivalent soil mass procedure for monitoring soil organic carbon in multiple soil layers. *European Journal of Soil Science, 64*(1), 58–65. https://doi.org/10.1111/ejss.12002

West, T. A. P., Börner, J., Sills, E. O., & Kontoleon, A. (2020). Overstated carbon emission reductions from voluntary REDD+ projects in the Brazilian Amazon. *Proceedings of the National Academy of Sciences of the United States of America, 117*(39), 24188–24194. https://doi.org/10.1073/pnas.2004334117

Wieder, W. R., Hartman, M. D., Sulman, B. N., Wang, Y.-P., Koven, C. D., & Bonan, G. B. (2018). Carbon cycle confidence and uncertainty: Exploring variation among soil biogeochemical models. *Global Change Biology, 24*(4), 1563–1579. https://doi.org/10.1111/gcb.13979

Wielopolski, L., Chatterjee, A., Mitra, S., & Lal, R. (2011). In situ determination of soil carbon pool by inelastic neutron scattering: Comparison with dry combustion. *Geoderma, 160*(3–4), 394–399. https://doi.org/10.1016/j.geoderma.2010.10.009

Wijewardane, N. K., Ge, Y., Wills, S., & Libohova, Z. (2018). Predicting physical and chemical properties of US soils with a mid-infrared reflectance spectral library. *Soil Science Society of America Journal, 82*(3), 722–731. https://doi.org/10.2136/sssaj2017.10.0361

Wijmer, T., Al Bitar, A., Arnaud, L., Fieuzal, R., & Ceschia, E. (2024). AgriCarbon-EO v1.0.1: Large-scale and high-resolution simulation of carbon fluxes by assimilation of Sentinel-2 and Landsat-8 reflectances using a Bayesian approach. *Geoscientific Model Development, 17*(3), 997–1021. https://doi.org/10.5194/gmd-17-997-2024

Wu, W., & Lin, B. (2026). MRV systems for agricultural soil organic carbon in offset markets: Methods, uncertainty, and policy. *Mitigation and Adaptation Strategies for Global Change, 31*(5), Article 57. https://doi.org/10.1007/s11027-026-10327-z

Xiao, L., Zhou, S., Zhao, R., Greenwood, P., & Kuhn, N. J. (2020). Evaluating soil organic carbon stock changes induced by no-tillage based on fixed depth and equivalent soil mass approaches. *Agriculture, Ecosystems & Environment, 300*, Article 106982. https://doi.org/10.1016/j.agee.2020.106982

Xie, E., Chen, J., Peng, Y., Yan, G., & Zhao, Y. (2024). Integration of machine learning and process-based model outputs via ensemble Kalman filter enhanced space–time modelling of soil organic carbon in a highly human impacted area. *Soil Use and Management, 40*(4), Article e13127. https://doi.org/10.1111/sum.13127

Yakubova, G., Kavetskiy, A., Prior, S. A., & Torbert, H. A. (2016). Benchmarking the inelastic neutron scattering soil carbon method. *Vadose Zone Journal, 15*(2), 1–11. https://doi.org/10.2136/vzj2015.04.0056

Yang, Q., Liu, L., Zhou, J., Ghosh, R., Peng, B., Guan, K., Tang, J., Zhou, W., Kumar, V., & Jin, Z. (2023). A flexible and efficient knowledge-guided machine learning data assimilation (KGML-DA) framework for agroecosystem prediction in the US Midwest. *Remote Sensing of Environment, 299*, Article 113880. https://doi.org/10.1016/j.rse.2023.113880

Zhang, L., Heuvelink, G. B. M., Mulder, V. L., Chen, S., Deng, X., & Yang, L. (2024). Using process-oriented model output to enhance machine learning-based soil organic carbon prediction in space and time. *Science of the Total Environment, 922*, Article 170778. https://doi.org/10.1016/j.scitotenv.2024.170778

Zhou, Q., Guan, K., Wang, S., Jiang, C., Huang, Y., Peng, B., Chen, Z., Wang, S., Hipple, J., Schaefer, D., Qin, Z., Stroebel, S., Coppess, J., Khanna, M., & Cai, Y. (2022). Recent rapid increase of cover crop adoption across the U.S. Midwest detected by fusing multi-source satellite data. *Geophysical Research Letters, 49*(22), Article e2022GL100249. https://doi.org/10.1029/2022GL100249

Zhou, W., Guan, K., Peng, B., Margenot, A., Lee, D., Tang, J., Jin, Z., Grant, R., DeLucia, E., Qin, Z., Wander, M. M., & Wang, S. (2023). How does uncertainty of soil organic carbon stock affect the calculation of carbon budgets and soil carbon credits for croplands in the U.S. Midwest? *Geoderma, 429*, Article 116254. https://doi.org/10.1016/j.geoderma.2022.116254

Peer-review history:
The peer review history for this paper can be accessed here: https://pr.sdiarticle5.com/review-history/166542