AGRICULTURAL SOIL CARBON:

A CALL for Improved Evidence of Climate Mitigation

Environmental Defense Fund

YASSP

- We identify the critical need for causal approaches to be employed at the scale of commercial agriculture to build high-quality evidence for Measurement, Monitoring, Reporting and Verification (MMRV) to quantify the effectiveness of soil carbon farming.

- We emphasize that, contrary to arguments that have led to reliance on process-based biogeochemical models for carbon accounting, empirical measure-and-remeasure projects appear scientifically feasible at regional agricultural scales with current best practices for soil sampling and carbon analysis.

- Even if modeling approaches remain predominant, we make the case that project-scale empirical data are required to test, support and advance the ability of such scaling approaches to estimate real emission reductions and removals.

To increase confidence that the effects of carbon farming represent real carbon accrual and/or avoided emissions, we summarize the design principles that should underpin measure-and-remeasure approaches. We use the term "carbon farming" in this report since change in soil organic carbon is the primary outcome variable in agricultural soil climate accounting and the term is being increasingly used in policy frameworks, such as the European Union's certification for carbon removals (EU, 2024). However, we recognize that soil carbon accrual and avoided emissions are just one way that climate-smart and regenerative agricultural practices contribute to climate change mitigation, adaptation and sustainability. We do not address the effectiveness of soil carbon farming as a climate mitigation strategy, which necessitates quantification of potential tradeoffs with emissions of other greenhouse gases (GHGs) of concern such as N_2O. Instead, we focus on the need to expand soil MMRV to validate that climate mitigation claims represent reality. We propose practical ways forward to generate high-quality evidence at the scale of commercial agriculture that can help to inform, quantify and validate GHG outcomes, as well as support and advance the suitability of the varied MMRV approaches being used or proposed for scaling soil carbon farming. We emphasize that such high-quality evidence can, more broadly, help to identify and improve practices that have the greatest benefits for climate adaptation and mitigation.

CITATION AND FUNDING

Please cite this report as: *Bradford, M.A., Oldfield, E.E., Arredondo, M.G., Black, H.I.J., Forbes, E.S., Jevon, F.V., Kelly, C., Lavallee, J.M., Liu, S., Polussa, A., Potash, E., Rosenzweig, S.T., Sanderman, J., Sheban, K.C., Sheffer, M., Treakle, T., Tully, K., Kuebbing, S.E. and the Synthesizing Science Lab (2025) Agricultural soil carbon: A call for improved evidence of climate mitigation. Yale Applied Science Synthesis Program and Environmental Defense Fund white paper.*

The members of the "Synthesizing Science Lab" were Michael Culbertson, Emily Evans, Marshall McCall, Phoebe Merrick, Onolunose Oko-Ose, Kristen Tam and Ingrid Thyr. Development of the white paper was supported by these graduate students through a practicum class at the Yale School of the Environment. The class supported a workshop – bringing together participants from multiple sectors involved in practice, policy and research around carbon markets, carbon accounting, and agriculture – that contributed to the development of this white paper. Those sectors represented at the workshop included U.S. Federal agencies, climate registries, agribusinesses, food corporations, independent consultants, private climate investment companies, academia, growers, and environmental non-governmental organizations.

Funding for the workshop was obtained through a grant from the Yale Center for Natural Carbon Capture and the Environmental Defense Fund, with awards from King Philanthropies and Arcadia, a charitable fund of Lisbet Rausing and Peter Baldwin. The workshop was hosted by The Forest School, Yale School of the Environment.

The organizers of the workshop, class and white paper - Drs. Mark A. Bradford, Emily E. Oldfield and Sara E. Kuebbing - have no commercial interests in carbon markets.

Dr. Oldfield is Senior Scientist at the Environmental Defense Fund, one of the world's leading environmental nonprofit organizations. Guided by science and economics, EDF finds practical and lasting solutions to the most serious environmental problems.

Drs. Kuebbing and Bradford are Director and Faculty Lead, respectively, of the Yale Applied Science Synthesis Program (YASSP), whose mission is to generate science to support decision making. YASSP produces quantitative, reputable and open scientific syntheses that guide and inform direct actions around land stewardship. One aim of the Program is to be a neutral organizing space for socializing and advancing ideas that advance biodiversity, climate and human goals related to ecosystem management.

YASSP is a joint initiative of the Yale Center for Natural Carbon Capture and The Forest School, Yale School of the Environment, with ongoing support from these entities and extramural grants and gifts.

The organizers acknowledge Stefan Jirka, Alison Nord, Daniel Rath, Stephen Wood, Cristel Zoebisch and others, who cannot be named (nor authored or acknowledged) for political reasons, for their contributions to this synthesis effort.

The workshop followed Chatham House rules given that various attendees were limited for proprietary, conflict-of-interest, and regulatory reasons from having their perspectives attributed to them or their organizations. The report assessment and guidelines therefore do represent the expert synthesis of the authors but cannot be assumed to necessarily represent the expert viewpoints of unnamed attendees and nor the institutional affiliations of the authors or those acknowledged.

Building a Strong Data Foundation for MMRV

The opportunity to increase soil organic carbon stocks through agricultural management interventions, such as practices considered climate-smart (which commonly overlap with practices labeled as regenerative or sustainable), is recommended by the Intergovernmental Panel on Climate Change (IPCC) as a critical strategy for mitigating and adapting to climate change (IPCC, 2022). Yet there is low confidence in the estimated magnitude of climate mitigation benefit from soil "carbon farming". Despite the IPCC recommendation and substantial corporate and government investment, including the U.S. Inflation Reduction Act and EU Carbon Removals and Carbon Farming Regulation, confidence that agricultural soil carbon farming is an effective natural climate solution (NCS) ranges from very low to relatively high (Amundson and Biardeau, 2018; Janzen et al., 2022; Minasny et al., 2017; Oldfield et al., 2024; Paustian et al., 2016; Schlesinger, 2022; Zomer et al., 2017).

One cause of low confidence is the approaches to MMRV used to quantify soil carbon accrual or avoided emissions (Hayes et al., 2023). The prevailing approaches are rooted in "upstream" data from a limited set of small-plot research studies (Garsia et al., 2023; Le Noë et al., 2023; Ogle et al., 2023a) (Figure 1). Whether these highly controlled, experimental studies generate outcomes representative of those on working farms is unknown and, understandably, is questioned given their failure to reproduce conditions under which commercial-scale agriculture is practiced (Oldfield et al., 2024). Yet these data are used to inform and validate process-based models, emission factors and carbon intensity scores (for definitions, see "Glossary of Key Terms" in the "Additional Insights" at end of the main text) used to estimate soil carbon outcomes across a variety of use-cases, including national-level GHG inventory accounting, voluntary carbon market (VCM) credit generation, and scope 3 inventories and interventions. These accounting methods – either explicitly or implicitly – make causal claims about the impacts of an agricultural intervention (e.g., no-tillage, cover cropping, compost amendments, crop diversification) on GHG outcomes.

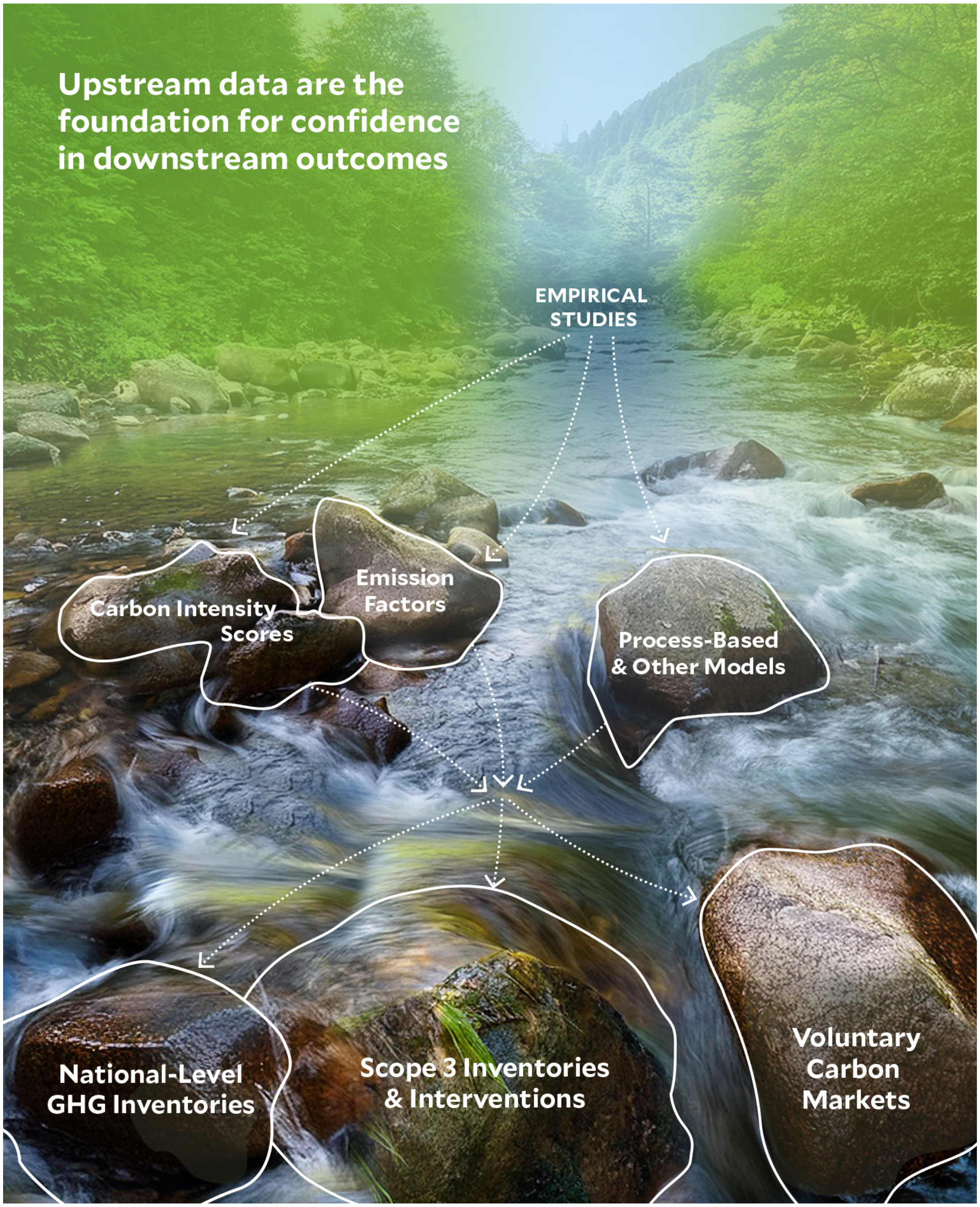

Figure 1 | Upstream evidence from empirical studies must be fit-to-purpose to build confidence that downstream approaches that scale on the basis of this evidence make reliable causal claims. *In soil carbon accounting, empirical upstream data are largely limited to small-scale experimental studies designed to test mechanisms and potential effect sizes but not the quantitative impacts of agricultural practices on soil carbon accrual or avoided emissions under real-world conditions. Our synthesis work suggests that the average plot size of the 20-30 such studies for the U.S. Midwest – used in downstream approaches – is ~200 m^2 (i.e. ~1/20th acre), with agricultural practices applied in ways that fail to recreate conditions of commercial agriculture (see main text). Given that most accounting approaches (e.g. emissions factors) rely on this same set of small-scale experiments, it should not be surprising that they estimate similar climate mitigation benefits of adopting practices such as cover cropping, regardless of the intended purpose (e.g. carbon markets and national inventories). Empirical studies conducted at the scales of commercial agriculture are needed to build confidence in the evidence to support causal claims of climate mitigation.*

Given the stakes – reducing GHGs, offsetting fossil fuel emissions, meeting corporate GHG targets – the reliance of measure-and-model approaches on small-scale experimental data undermines confidence in estimates of climate mitigation through soil carbon farming at the regional scales of agricultural land management (Oldfield et al., 2024). Calls to build confidence in soil carbon removal and avoided emissions claims are intensifying (Hayes et al., 2023), with demands to redress the paucity of data that are available from working farm fields across regions and for specific commodities (Bradford et al., 2023; Potash et al., 2025). Broad efforts – in the U.S. and internationally – are releasing billions of dollars of government funding for land management interventions to mitigate climate change, with the expectation that their effectiveness can be reliably quantified. Yet unless the limitations of current MMRV approaches for soil carbon farming are addressed, it seems likely that those funds, in addition to private investment, may be directed away from climate-smart agricultural management. Such actions would raise financial barriers for farmers to adopt practices that counter the degradation of agricultural soils, services and interconnected ecosystems. A report from the Stanford Law and Policy Lab (Hayes et al., 2023) underscores the need for data at the scale of working farms and ranches:

> “Forward progress in this area [i.e., climate-smart agriculture] is constrained by major data and analytical gaps that prevent accurate and verifiable quantification of how much climate-smart farming and ranching practices are growing carbon stocks in soils...the lack of practical and scientifically sound approaches for confirming that specified practices generate claimed benefits, and the lack of access to confirmatory data, pose major systemic impediments...”

The Yale Applied Science Synthesis Program (YASSP) and Environmental Defense Fund (EDF) convened a workshop in October 2024 to address MMRV needs for soil carbon farming. The workshop brought together a variety of stakeholders, including individuals from climate registries, food and agricultural corporations, government agencies, project developers, academia, carbon markets, growers and non-governmental organizations. The broad goal of the workshop was to address repeated calls to generate data from working landscapes to prove the climate impact of soil solutions. **To answer these calls, here we provide scientific guidelines for MMRV efforts that can build confidence in the extent to which climate-smart agricultural practices accrue and reduce emissions of soil carbon.**

Causal Claims Require Causal Data at Commercial Scales

Emissions from the agriculture sector represent ~30% of global anthropogenic sources and continue to increase, including those from within the "farm-gate" (FAO, 2024). Climate-smart agricultural management (e.g., cover cropping, reduced tillage, cropping system diversification) is considered a technologically ready and scalable climate solution (IPCC, 2022). Soil carbon farming – involving the accrual of carbon in soil organic matter as a temporary biogenic stock, or reduced soil emissions from this stock – is already in use for generating carbon credits and for national-level and other accounting of GHG emissions (Mathers et al., 2023; Ogle et al., 2023b; US-EPA, 2022). Proving these use cases are meeting GHG targets demands an answer to a key question:

How much carbon is really being accrued or retained in soils?

As in all open natural systems, quantification of the amount of soil carbon accrued through an intervention is not possible, but it can likely be estimated with plausible accuracy (Bradford et al., 2024, 2023; Potash et al., 2025). Such an estimate demands a suitable counterfactual (i.e. control) that represents what would have happened in the absence of intervention (for causal definitions, see "Glossary of Key Terms" in the "Additional Insights" at end of the main text). Because the interventions and controls cannot be applied to the same fields, we can never know with 100% certainty the "true" effect of an intervention: it is always an estimate (Holland, 1986). However, this should not preclude demanding high standards of evidence. For example, regulatory impact assessments in healthcare demand the generation of causal data before and after health interventions are introduced – compared to counterfactual conditions where the intervention is not applied – to ensure their effectiveness. High-quality causal data must reliably support the conclusion that the impact is (i) the result of the intervention and (ii) realized at the scale of practice (Rothwell, 2005).

The effectiveness of a vaccine in a mouse model, for example, might satisfy criterion (i) but not (ii) since we would have little confidence that the vaccine would be effective for human populations. Satisfying criterion (ii) requires randomized-control trials and/or observational studies that quantify disease prevalence and morbidity in human subjects that are a representative subsample of the target population for vaccination (Figure 2). The mouse model, just like small-plot experimental agricultural studies, has high evidentiary value for helping to resolve mechanisms of action and for showing the *potential* effectiveness of an intervention (i.e. these data have high "internal validity"). These highly-controlled data can inform mathematical models that predict how vaccine delivery will lower disease spread in a human population; just as internally valid data from small-plot experimental studies in agriculture are then used to inform and validate models for estimating GHG outcomes at larger scales of commercial agriculture. Yet human health is so important that we place the burden-of-proof for vaccine effectiveness on studies that collect data at the scale at which the vaccine is delivered to a human population (Figure 2). Such data have high "external validity". Similarly, to generate externally-valid data that then directly speaks to the real-world effectiveness of soil carbon farming, we need studies that are designed to collect data and estimate treatment effects at the scales at which practices are employed to commercial fields within agricultural growing regions.

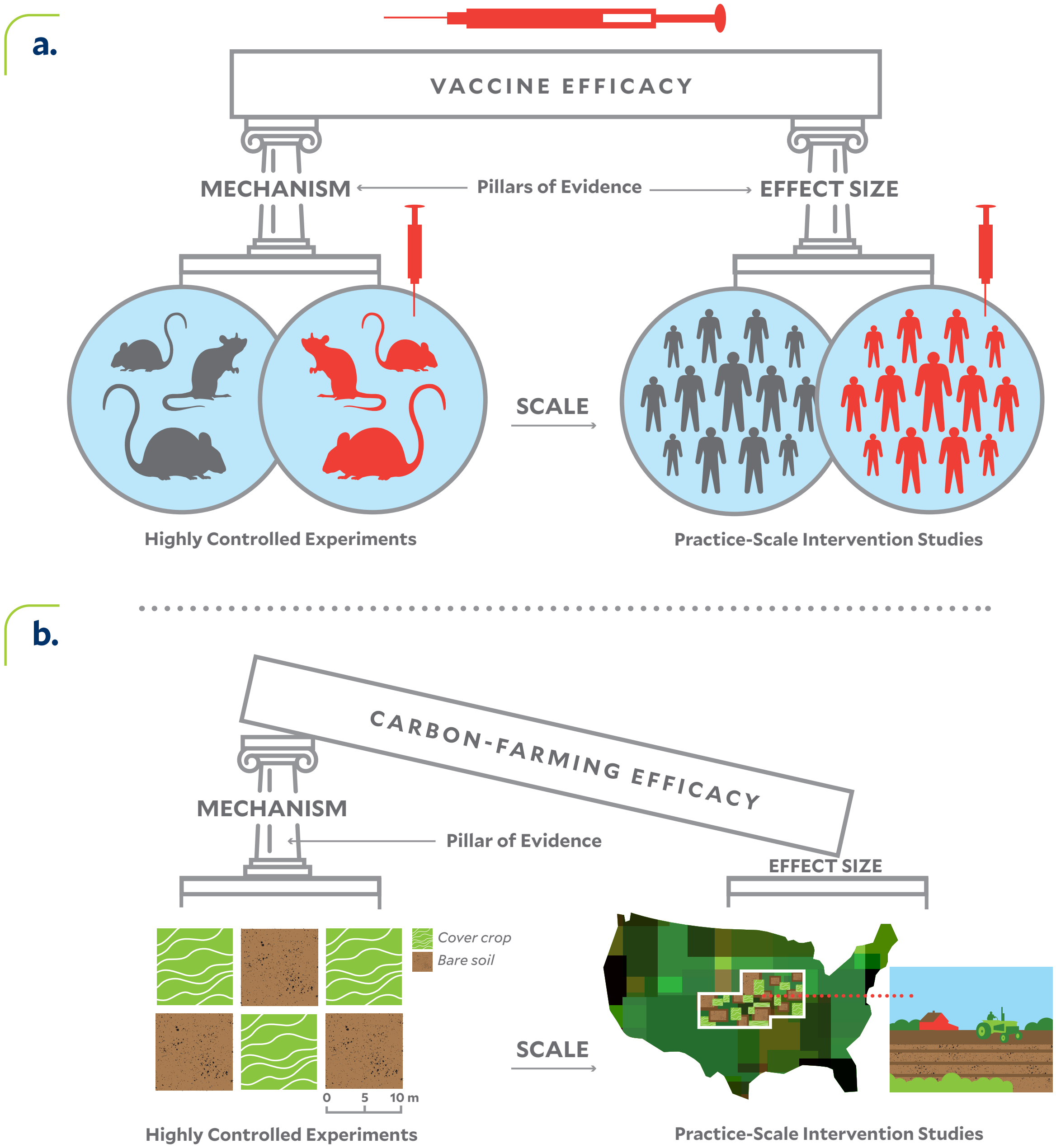

Figure 2 | Necessary pillars of evidence for making robust causal claims about the impact of interventions. *(a) Impact assessments in causal fields, such as epidemiology, rely on i) an understanding of mechanism and potential effectiveness from controlled studies and ii) real-world intervention effects when delivered by practitioners to human populations. These complementary endeavors build confidence that causal claims of real-world quantitative benefit are built upon reliable pillars of evidence. In contrast, (b) causal claims for soil carbon accrual and avoided emissions rely primarily on upstream evidence from only controlled, small-plot empirical studies, that are then scaled through modeling approaches (see Figure 1). To build confidence in causal claims of climate mitigation benefits requires the evidence base be expanded through quantitative impact studies of climate-smart agricultural interventions, conducted by practitioners at the scale and conditions of regional agriculture.*

We believe carbon accounting claims demand causal data that have high external validity.

It is critical that we channel resources towards effective climate action now and avoid investments that have limited climate benefit.

Causal data can help us prioritize the policies and practices that result in meaningful climate mitigation. The burden-of-proof for carbon accounting claims should then fall upon empirical data collected at the scales at which management interventions are delivered to an agricultural region. Exploring the data underlying the quantification of soil carbon farming interventions reveals several short-comings (Oldfield et al., 2024; Poulton et al., 2018): (1) small plots from research trials designed to minimize heterogeneity of soils and landscape; (2) implementation of aspirational management that does not necessarily reflect how interventions are practiced on working farms; (3) lack of business-as-usual or poorly-defined controls that make it difficult to establish causal impact (i.e. treatment effect of intervention); (4) insufficient replication, where treatments may occur over three or four small plots, and sometimes are unreplicated (e.g. see the appendix of Mathers et al., 2023). Collectively and individually these shortcomings prevent the determination of a robust population-level (e.g. VCM project) treatment effect of what we might expect when evaluating many farms. Yet such data are used for soil carbon estimation approaches (e.g. emission factors, carbon intensity scores and process-based models) to make claims of GHG removals for corporate GHG targets, VCM credits, and national-level inventories (Figure 1); all of which are being scrutinized due to a lack of confidence in estimated GHG outcomes.

We see a pressing need for causal, empirical measure-and-remeasure studies at the scale of commercial agriculture to increase standards of evidence and build confidence that we are making progress towards climate goals through carbon farming. Such studies will also enable policies that support practices that are demonstrated to be most effective for climate mitigation. That is, they can help us identify what works best where and what practices are ineffective. The general guidelines we share are translated from principles underpinning causal inference approaches (e.g. Angrist and Pischke, 2010; Holland, 1986; Imbens, 2024; Rubin, 1974) used in fields such as economics and epidemiology to make robust claims about the quantitative effectiveness of interventions. The most important take-home messages we offer, to those involved in carbon accounting, are that they need to know:

(a) The most rigorous causal claim is typically made at the population level; and (b) study design and knowledge of the system are critical to making externally valid, causal claims about mitigation effectiveness.

The most reliable answer to a causal claim is the estimated average, population-level benefit (and potential downsides) of an intervention. For example, in a VCM agricultural project with 10,000 fields, a rigorous and externally valid estimate of intervention effectiveness is then achieved at the project scale, with an estimate of the average carbon accrual rate and uncertainty per unit area of the project (e.g. Mg C ha^{-1} y^{-1}; see Potash et al., 2025). Estimates of carbon accrual for any individual field in the project – including the sample of fields where soil carbon stocks are measured – will likely be highly inaccurate (Bradford et al., 2023). Essentially, variability in soil carbon stocks can overwhelm the ability to accurately measure – or model – change at the individual field scale (Bradford et al., 2024, 2023; Ogle et al., 2010; Potash et al., 2025). Indeed, even the causal claim for a subgroup of the fields, such as those with sandier soils, may not be representative of the intervention's effectiveness (the same can be true if the spatial scale is too large; see "Additional Insights" section). Yet the idea that the causal claim is most reliable at the project-scale should not be a surprise. In a vaccination study, for example, epidemiologists quantify the average benefit for the target population and cannot predict whether a particular vaccinated individual will contract the communicable disease. The focus then is on making an accurate causal claim at the project (i.e. population) scale. Such high quality, externally valid evidence is a strong basis for making policy and practice decisions about the most beneficial and practical interventions to enact.

The focus on making accurate causal claims, through well-designed measure-and-remeasure approaches, can help redirect the considerable efforts that soil carbon accounting projects pore into quantifying and reducing uncertainty in estimates of carbon accrual or avoided emissions. **Currently, soil carbon accounting is overly focused on only one type of uncertainty, namely precision. Efforts need to be redirected to give at least as much attention to the accuracy of the causal claim.** Specifically, the value of striving to achieve greater precision in the estimated GHG impact is questionable if the accuracy of the estimate cannot be rigorously validated (Bradley et al., 2021). To paraphrase, being approximately correct should be preferred to being precisely wrong. Yet when the downstream approaches, such as emissions factors and modeling, rely on data from small-scale experimental studies, we must assume that the experimental results are accurate representations of the mitigation benefits achieved under real-world conditions. That tenuous assumption demands testing to “prove” the suitability of carbon accounting approaches and to validate mitigation claims at real-world, commercial scales.

Testing the assumption requires applying domain knowledge and principles of study design that together generate externally valid, causal estimates of mitigation effectiveness that can reliably represent the true intervention effect. In the table below, we describe major components and the associated principles of study design that are needed for rigorous, practice-scale empirical studies that yield high quality, causal data. **Such study designs will focus on robustly estimating the differences, due to the intervention, in temporal changes in soil carbon stocks between intervention and comparison fields.** We give examples in Table 1 of how to apply the principles to an agricultural intervention project, considering knowledge about how soil carbon is accrued and lost.

PHOTO BY MARK BRADFORD

Table 1 | *Causal study design components and guidelines needed to reliably estimate the mitigation effectiveness of carbon farming practices such as reduced and no-tillage, cover cropping and cropping system diversification. For definitions of causal terms, see "Glossary of Key Terms" in the "Additional Insights" section, beginning after main text.*

DESIGN COMPONENTS	GUIDELINES	EXAMPLES OF GUIDELINE APPLICATION
Individual	**Define** "individual" as the smallest unit that receives the management intervention	An agricultural field is the individual when it is the unit of management.
Population (spatial scale)	**Define** the "population" as all the individuals, under a defined context, that are eligible to receive the intervention within the spatial scale of interest	The population should consist only of fields under similar contexts. If the population of interest is only rainfed fields in the U.S. corn belt, then including irrigated fields in the region would confound understanding of intervention effectiveness.
Measurement	**Decide** how to measure the individual response over time in a manner that is representative for that individual	Agricultural soil studies often measure only a single small area within a larger field. Compared to taking samples across a field, smaller sub-plot designs are inefficient and generally make estimated effects of interventions unreliable. Similarly, many studies sample at only one point in time, making it hard to resolve intervention effects from other causes of spatial and temporal variation in stocks (see final row of table). Lastly, the measured outcome variable for the individual should represent the desired response. For example, soil carbon concentrations are not suitable proxies for assessing changes in soil carbon stocks.
Causes	**Ensure** the cause is the management intervention as applied in the real-world context of the study	The focus should be on quantifying the effects of carbon farming, and not on 'causes' that cannot be manipulated but which might influence the outcome (e.g. soil texture). Such a focus on the cause that is manipulated (i.e. the policy or practice intervention) shifts study goals from explaining variation in measured soil carbon change to quantification of the causal effect of the intervention. Note that a cause does not have to be a single practice (e.g. cover crops) but could be an approach to farming that integrates multiple practices, such as "organic regenerative".

DESIGN COMPONENTS	GUIDELINES	EXAMPLES OF GUIDELINE APPLICATION
Controls / comparators	**Choose** a control/ comparison group of individuals that most plausibly represent the management practices that would have continued absent the intervention	A causal effect is always estimated relative to a suitable comparison group. A suitable control/ comparison group supports dynamic baselines that serve as a counterfactual scenario to estimate what would have happened had fields not received the management intervention. Note that the process of assignment to intervention control (versus treatment) should satisfy – or approximate – the "exchangeability criterion" (i.e. fields in the population should have an equal likelihood of being assigned to the representative 'control' group). Such considerations help to avoid potential bias in determining treatment effectiveness; for example, where interventions occur on less fertile fields but are compared with outcomes for more fertile control fields.
Sampling individuals	**Select** the sample of individuals to measure from the target population in a representative manner	Population-level variation in non-target causal variables which influence the outcome (e.g. soil texture on soil carbon) should ideally be captured across the sampled individuals. For example, higher clay fields might accrue more carbon, so treatment and control fields might be stratified by soil texture. Capturing such variation through study design, such as stratification, makes the sample treatment effect more representative of the likely population treatment effect.
Timescale	**Choose** a timescale that matches with policy needs	Recognizing that soil carbon farming relies on temporary storage, policy designs for incentivizing carbon farming are focusing on shorter timescales (e.g. practices are maintained for 10 years). Such realities require study designs that measure impact within and across the timescales over which interventions are applied at scale.

DESIGN COMPONENTS	GUIDELINES	EXAMPLES OF GUIDELINE APPLICATION
Robust estimation of treatment effect	**Select and measure** treatment and control individuals in a way that means treatment effects are reproducible and ideally transferable	The estimated treatment effect should be approximately equal (i.e. reproducible) when different treated and control fields are sampled from the same population; and when different soil sampling locations are used within the same set of fields. Such robustness to sampling decisions builds confidence that the effectiveness of the intervention is realized at the scale of the population of fields of interest (i.e. the sample is generalizable). Testing the impact of the intervention under new contexts (e.g. in one region of France versus another physiographic province) and being able to explain why treatment effects might vary for the same intervention, builds confidence that the intervention will be applied only to populations where it is effective (i.e. transferability).
Treatment effect	**Understand** that causal approaches are suited to estimating the average (and not individual) treatment effect	Because soil carbon stocks in individual fields are variable and respond to multiple causes, it is unlikely that soil carbon change in a field can be reliably measured or attributed to the intervention. Yet adhering to study best practices overcomes this individual variability when effects are considered at the population scale. The focus should then be on robustly estimating the mean population response, such as the average treatment effect (i.e. the "ATE"). Importantly, the ATE in this instance is the difference in stocks between intervention and comparison (or control) fields – attributable to the intervention – that is realized over time. The most accurate estimates of this effect will involve measuring stocks pre- and post-intervention, to separate changes due to the intervention versus spatial and temporal variation in stocks due to other causes (e.g. clay content, precipitation, non-target management effects).

Putting Guidelines into Practice

We acknowledge that carbon farming demands climate financing or other revenue streams to pay for practice adoption by farmers and to compensate them if their yields initially suffer under the new practices. Use of process-based models, per current protocols, generates this financing because the models are used to estimate emission reductions on an annual basis, permitting issuance of carbon credits before they may be directly measurable. For VCM projects a potential downside to empirical causal approaches may then be that credits – and therefore revenue – are issued only once practices have been in place for the multiple years necessary for a project to make measurable causal claims (Potash et al., 2025). Yet the reality is that without empirical approaches being adopted, whether for market or non-market based carbon accounting, the evidence base for making quantitative causal claims is much weaker than it could be (Figure 2). **Policymakers and practitioners involved in these efforts need to decide and communicate the standards of evidence they deem acceptable for carbon accounting. In the absence of such guidance, we argue higher standards of evidence are required, which then demands causal impact studies conducted at the real-world scales and conditions at which agriculture is practiced.**

Rather than place the burden of funding rigorous empirical causal studies on any one carbon accounting effort (e.g. VCMs vs. national GHG inventories), we suggest that a pre-competitive federal-private-academic cooperative effort could generate measure-and-remeasure datasets at project scales to independently validate both non-market and market-based accounting approaches for soil carbon. All these approaches, from national inventories to Scope 3 to carbon markets, currently rely primarily on measure-and-model approaches for their "most rigorous" accounting. The "measure" in these approaches is not designed to quantify intervention effects, and the explicit purpose of measurement in soil VCMs, beyond establishing initial stocks, remains ambiguous (Lavallee et al., 2024). These approaches then are model-centric and require practice-scale, causal empirical evidence to "prove" that they can be trusted for making causal claims about carbon accrual and avoided emissions at real-world scales (see Le Noë et al., 2023). Such evidence will also inform and update less technical approaches, such as carbon intensities and emissions factors.

Governments, corporations and/or academic research institutions could work together to establish regional-scale agricultural studies that adhere to the causal principles we outline in Table 1, thereby generating rigorous causal datasets at the scale of practice. Transparently and openly sharing such population-scale data, even when there is a need to de-identify them (a similar model to healthcare), would foster innovation in carbon farming and build trust that mitigation claims are accurate. To ensure carbon farming – and soil stewardship more broadly – is grounded in rigorous science we need to go back upstream and redress the deficit of real-world causal evidence (Figure 1). In short, to have confidence that model estimates are realistic, and to promote practices that have the best outcomes, we need causal data from well-designed, empirical studies conducted under real-world scales and conditions.

1 Glossary of Key Terms

Average Treatment Effect (ATE): The average difference in the outcome variable (e.g. change in SOC stock) between the treated and control (or comparison) fields sampled from the population of project fields. In the context of carbon farming, the ATE incorporates variability in response in SOC stock changes at the individual field level to arrive at the average effect on SOC change of adopted policies and practices.

Carbon Intensity Score: An estimation of the GHG emissions associated with the production of a commodity.

Causal Inference: Causal inference and causal study designs rest on estimating the amount of measured change in time (e.g. in SOC stock) which is attributable to a cause, such as a policy or practice intervention. Causal study designs rely on assumptions, applied to experimental or observational data, that build confidence we can reliably estimate measured change in outcomes between a treatment and a control (or comparison) population.

Control/ Comparison Group: The untreated units in an experiment (e.g. agricultural fields that are not receiving the treatment interventions), where data are either observational (i.e. natural or quasi-experimental) or collected under a randomized controlled trial. This group provides the counterfactual scenario.

Counterfactual: In carbon accounting, the counterfactual represents what would have happened on a given field had business-as-usual practices continued or, perhaps more appropriately, what practices would have been used absent climate payments (e.g. carbon credits) or other incentives (including regulations).

Dynamic Baseline: The reference condition (e.g. the counterfactual representing practices used absent climate payments) practiced among the control population that serves as a comparison enabling the estimate of the causal effect on the treatment population. A dynamic baseline accounts for the fact that certain conditions, like drought, can impact outcomes on both the control and treatment populations which might, with only a static (e.g. historical) baseline, be falsely attributed as part of the treatment effect.

Emission Factor: A quantitative tool used in GHG accounting; for soil carbon accounting, emission factors represent values that attempt to relate the amount of GHG flux per unit activity, such as tonnes of CO_2 equivalent emitted per hectare (see https://ghgprotocol.org/agriculture-guidance).

External Validity: The extent to which a study's results (e.g. ATE for the sampled fields) are thought applicable to all the treated fields in a carbon farming project (termed generalizability), or potentially to other projects (termed transferability). The more representative the sample of fields is of the project population, the higher the external validity of the results for that project.

Internal Validity: A study that has high internal validity refers to one that accurately estimates the causal treatment effect (e.g. the ATE). To have high confidence in the climate mitigation achieved in a carbon farming project, one would want the study design to generate both high internal and external validity.

Measurement, Monitoring, Reporting and Verification (MMRV): A system or protocol for developing and tracking specific methods and outcomes, transparently communicating assumptions and data, and validating that the information is accurate and complete.

Random Assignment: An approach to treatment assignment in which all units have an equal probability of receiving the treatment, regardless of underlying characteristics. Randomization is intended to ensure that there are no systematic differences between the treated and control units, allowing for an unbiased estimation of the treatment effect (Bradford et al., 2019). However, note that the idea of "unbiased" is not the same as "accurate". For example, an unbiased estimate – obtained using randomization in the study design to select treatment and control fields – can be highly inaccurate when sample numbers of fields, or sampling densities within fields, are too low to capture variability in conditions and outcomes across a project (see Bradford et al., 2024, 2023).

Static Baseline: The baseline is set, for example, at the first time point (time 0) of measurement and changes in SOC stocks over subsequent years are measured against time 0. See "dynamic baseline" for mention of pitfalls with static baselines.

2 Key Concepts

Carbon Farming vs. Climate-Smart Agriculture
We deliberately use the term "carbon farming" in this report to underscore that the primary policy and market incentives are focused on soil organic carbon (SOC) as a critical mitigation outcome. We recognize the multitude of benefits associated with adoption of climate-smart and regenerative practices that are critical not only to mitigation outcomes but also adaptation. The scientific guidelines we share for generating evidence to increase confidence in mitigation outcomes can also be used to support stewardship of agricultural soils that is essential for human and environmental services such as clean waters and food security.

Causal Measure-and-Remeasure is Not Monitoring
We clearly distinguish monitoring approaches versus causal measure-and-remeasure approaches. Monitoring involves measuring change in soil carbon over time following practice adoption, looking for differences from the static, historical baseline in those fields. Attributing monitored change to the practice therefore makes the seemingly tenuous assumption that carbon accrual or avoided emissions are only due to the practice, as opposed to other factors (see Mitchell et al., 2024). In contrast, the measure-and-remeasure approach we outline here follows a causal framework that compares these changes in time following practice adoption to control fields which continue with "business-as-usual" practices. That is, they use a dynamic baseline. See Table 1 for causal study design components and guidelines needed to reliably estimate the mitigation effectiveness of carbon farming practices.

GHG Inventory Accounting
An inventory, in science, is an estimate of how much of something is in a defined area at a single point in time. For example, tree counts within discrete plots (i.e. the sample) – distributed across a particular forest – can be used to estimate total standing trees in the forest; as opposed to a census that counts the entire tree population. As such, inventories are simply an estimate of a stock at a single point in time and, alone, do not support causal claims. However, the term "inventory" in GHG accounting is often used to support causal claims, creating confusion about the assumptions and approaches necessary for a rigorous GHG inventory. Notably, in GHG accounting, inventories for national GHG emissions typically use a range of modeling approaches, from process-based models to emission factors, which assume a change in time against a reference condition/baseline (see IPCC, 2019, 2006). The assumptions of change in time are primarily based on small plot studies (see Figures 1 and 2). And these small plot studies are used to create such things as emission factors.

For an example of compiling an inventory using emission factors, consider IPCC GHG Inventory Guidelines (see IPCC, 2019, 2006) used to estimate soil carbon stocks for cropland in warm temperate climates on Mollisol soils in the US Midwest under a change in tillage. In this instance, the reference condition is established cropland (>20 years cropping) under full tillage with medium return of organic inputs (i.e. crop residues returned). If this cropland was converted to no-till, the inventory compiler would introduce a multiplier (in this case an emission factor of 1.11) that assumes that the SOC stock in those previously tilled fields would accrue annually – to a new steady state – due to conversion to no-till. In compiling the inventory with an emission factor, a causal claim about SOC stock accrual is therefore made. The guidelines we propose here (e.g. Table 1) could help substantiate that these claims are based on robust evidence of impact under commercial agricultural scales.

3 Study Design Challenges

Defining and Establishing Baselines

Causal study design necessitates a baseline/counterfactual which would be, for example, the control in a Randomized Control Trial, or the comparison group in an observational design like a natural experiment (Angrist and Pischke, 2010; Holland, 1986; Potash et al., 2025; Rubin, 1974; Siegel and Dee, 2025). Ideally, as indicated above and in Table 1, the comparison group would serve as a dynamic baseline to represent a counterfactual scenario to estimate what would have happened had fields not received the management intervention. Selecting a baseline then involves making a series of assumptions about what a suitable baseline is, which will influence confidence in the accuracy of the estimated intervention effect. For example, a static (time 0) baseline may be used as a "starting" condition with changes under new management practices measured over time. Such static baselines can lead to miscalculation of treatment effects, where for example variation in an external factor, and not the intervention, drives change in soil carbon stocks (Mitchell et al., 2024). In contrast, a dynamic baseline estimates the difference in the outcome of interest (e.g. soil carbon) between the "project" scenario and the baseline/counterfactual and, when chosen well, permits disambiguation of the intervention effect from effects of external factors.

Such a dynamic approach is used in protocols within the voluntary carbon market (e.g., Climate Action Reserve's Soil Enrichment Protocol and Verra's VM0042). However, both the protocol – and the project developed under it – shapes whether the dynamic baseline is modeled or empirically measured, and hence the assumptions that go into deciding if a baseline is a suitable approximation of what would have happened absent the intervention. Hence, knowing a baseline is dynamic is not in itself enough information to engender confidence that the chosen baseline will lead to accurate estimation of the average intervention effects.

To facilitate causal study designs where there is confidence in the dynamic baseline, one proposed solution is the establishment of regional baselines. The aim of a regional baseline is to understand background trends in soil carbon which may vary by region due to differences in climate, soil type, terrain, and common agricultural practices (Oldfield et al., 2022b). Regional baselines are intended as a tool to simplify comparisons between the practice intervention and baseline/counterfactual scenarios. In this concept, the regional baseline represents the average effect of the prevailing agricultural practice in a region. This type of baseline is highly dynamic, evolving with practice trends over time, helping to avoid under or over estimation of soil carbon changes attributable to interventions which are instead an outcome of factors such as climate extremes (e.g. droughts) and/or shifting baseline practice trends.

The Jurisdictional REDD+ model for tropical forest protection is an example approach that employs regional baselines that follow jurisdictional/ political boundaries. In this model, the baseline is measured as the average deforestation across the whole jurisdiction (defined by national or subnational political boundaries): interventions to avoid deforestation are measured against that jurisdictional baseline. A key benefit of jurisdictional baselining includes accounting for leakage within a jurisdiction, where carbon/ deforestation displaced within a jurisdiction will still be detected and accounted (Schwartzman et al., 2021). Further, regional baselines could, in theory, be a shared responsibility across projects for baselining, which may increase comparability among projects in estimated intervention effects and reduce costs incurred by any one entity. Such baselines could support carbon accounting across various platforms from national GHG inventories to Scope 3 to VCMs by approximating the counterfactual, building confidence in the effects of interventions.

Challenges remain, however, in moving from concept to implementation. In theory, regional baselines could be defined by soil types, farming practices, landforms, shared cropping systems, counties and/or ecoregions (Oldfield et al., 2022b). Indeed, the development of spatial analysis frameworks for crop yields and resilience – to compare intervention effects in a common way for an agricultural region – are underway (Rattalino Edreira et al., 2018). The unique combinations of biophysical and socio-economic circumstances captured in these frameworks are referred to as "technology extrapolation domains" (TEDs). Yet achieving the appropriate scale that captures heterogeneity also raises questions about how representative intervention effects are for a specific landscape. For instance, the larger the region and/or the more diverse landscapes it encompasses, the less certain one can be that the average treatment effect applies at sub-regional scales. Relatedly, variability within a region could lead to targeting "favorable" conditions for carbon crediting projects, creating a potential issue of "gaming the system" (Badgley et al., 2022; Randazzo et al., 2023). For example, if a region is defined as a county, and there are multiple soil types within a county, project developers may target sites with soil types that have, absent an intervention, rates of soil carbon losses that are lower than the baseline for the county, leading to over-crediting. Appropriate stratification of study designs by factors, other than the target interventions, can help to refine baselines and minimize such issues of poor comparability of treatment and baseline conditions.

Measurement to establish regional baselines for soils presents another challenge: forest cover and aboveground biomass lend themselves better to regional-scale monitoring through remote sensing approaches, such as satellite imagery. In the absence of demonstrated remote technologies to do this for soil carbon change, investment in national-scale monitoring of soil carbon stocks – using demonstrated soil sampling approaches – may increase the feasibility of regional baselines for soil carbon farming. Similarly, development of more rapid and logistically feasible approaches to measure soil carbon change may facilitate adoption of regional baselining for soil carbon (see Section 5 below).

Defining Treatments

Understanding the average treatment effect of a practice intervention requires a clear definition of the treatment/ intervention. The available data from long-term agricultural experiments generally have focused on quantifying the impacts of single practices. At the scale of commercial agriculture, farming and cropping systems can be much more complex and may demand adaptive approaches in response to climate, economic, and other individual motivations. For example, organic agriculture often requires some tillage to kill weeds due to the lack of herbicides; whereas conventional growers may minimize tillage in combination with pesticides to reduce labor and equipment costs. Additionally, some growers employ multiple regenerative practices, known as practice stacking, that can lead to different outcomes than using any practice in isolation while not necessarily having an additive effect. These complexities make tracking impacts of individual practices to inform management decisions and investment in soil carbon incentive programs challenging.

An emergent approach is the potential to use a continuous point scale system to encapsulate the spectrum of practices farmers employ (Fenster et al., 2021; Prairie et al., 2024). An alternative approach is to consider the treatment as a policy incentive – such as payments – to encourage use of more sustainable practices. Consensus or standardization of such systems among researchers and institutions will be needed to enable data sharing and external validity of measured treatment effects.

Measurement and Data: Field and Laboratory Standardization
There is a pressing need for concerted efforts that improve and develop methods and approaches that support high confidence in approaches for field sampling for soil carbon change, and the accuracy and reliability of analyses across laboratories of variables such as soil carbon and bulk density. Such development should occur together with cross-project data harmonization. Numerous calls and research papers emphasize these needs for accurate, repeatable field and lab measures, together with data harmonization. We refer the interested reader to further reading on these matters (see Even et al., 2025). We do not seek to replicate these calls but do recognize the importance of these non-trivial needs for rigorous soil carbon accounting. Yet, given the relative inattention to the need for causal measure-and-remeasure study designs, we focus our white paper on the need for soil carbon science to adopt causal, empirical study approaches. We note that rigorous measurement and data harmonization alone cannot be used to substantiate causal GHG accounting claims.

4 Accounting Context

Assessing Overall Climate Impact
Our focus in this report has been on causal designs to arrive at agricultural intervention impacts on SOC stocks. As mentioned early in the report, this is why we use the term "carbon farming" throughout the report, as it is the outcome of focus for many policies and investments. We note, however, that it is critically important to account for potential increases in other GHGs such as N_2O and CH_4 that might accompany practice interventions (Eagle et al., 2017; Guenet et al., 2021; Lugato et al., 2018). Just as with soil carbon, there is a dearth of downstream data and evidence to understand the net GHG impact of practice interventions. Similarly, other important processes such as soil erosion and lateral transfer also need to be considered within the full accounting context.

Reversals and Benefits Beyond Carbon
A primary reason for lack of confidence in soil carbon farming as a climate mitigation strategy is that gains in SOC stocks can be reversed. That is, SOC sequestration through carbon accrual is not permanent (or durable, another term often used in this context) – carbon continuously cycles through the soil and so durability rests on continuing (and monitoring) the management practices that build and maintain SOC stocks over time. Farmers may be wary about committing to requirements to maintain practices for decades into the future, making the design of effective incentives that meet durability standards difficult.

EU Carbon Farming certificates address this directly through recognition that soil carbon farming involves accrual or avoided emissions for a dynamic biogenic pool of carbon (EU, 2024). Beyond climate mitigation, the EU program also stresses the benefits for climate adaptation, reduced soil erosion, water quality, etc. These are important considerations, where a sole focus on permanent carbon removal may reduce investment for protecting and restoring lands (and waters), with consequent environmental degradation (e.g. biodiversity loss) and human impacts around such things as health, livelihoods and climate adaptation of food production (Bradford et al., 2019).

We note that concerns about reversals are important and exist whether measure-and-model or measure-and-remeasure approaches are used. Reversal and other concerns, such as additionality, are wrapped up in important considerations about whether investments in GHG reductions perform only this purpose or whether – as with agricultural interventions – they also have broader human and nature benefits, such as improved water quality, reduced soil erosion, and yield resilience (Huang et al., 2021; Kane et al., 2021; Oldfield et al., 2022a, 2019). Such issues reflect broader societal demands and whether addressable or not, they do not reduce the importance of credible MMRV for soil carbon accounting.

5 Proving Alternative Solutions: Spectroscopy and Remote Sensing

In response to the challenges of quantifying SOC change due to adoption of new practices, public and private sectors are investing in various efforts to optimize tools and approaches that balance cost and accuracy for MMRV. Further discussion of these approaches and other MMRV challenges and potential solutions can be found in Oldfield et al. (2024). Briefly summarizing that discussion, soil spectroscopy is an alternative to dry combustion methods – with potential in-field application, thereby avoiding shipping and laboratory costs – intended to address the need for repeated soil sampling campaigns at a reduced cost (Dangal et al., 2019; Wijewardane et al., 2020). Already a well-established technology in the research domain, private sector startups are capitalizing on these advancements to present the business case for spectroscopy as a scalable SOC MMRV solution.

Remote sensing technologies are also under development as scalable solutions to MMRV but are still in early phases of development. For instance, some remote sensing products can track agricultural yields and adoption of conservation agricultural practices such as no-till and winter cover cropping (Deines et al., 2023; Hagen et al., 2020; Surdoval et al., 2024). This information could be useful for monitoring practices as well as for determining additionality and leakage for SOC farming projects. With the proliferation of higher-resolution, satellite-based sensors, there is growing research linking these remotely sensed spectral signatures over bare ground to measured SOC data (Angelopoulou et al., 2019). This work generally shows promise for mapping the spatial distribution of surface SOC concentrations under ideal conditions (Castaldi et al., 2019). However, vegetation, crop residues and variable soil moisture conditions all confound the direct use of remote sensing to estimate SOC. So, while there is limited but growing success in mapping surface SOC concentrations over bare fields, there has been no demonstrated proof that remote sensing – alone or combined with advancing computing – can account for changes in time of SOC stocks to at least 30 cm.

While all these technologies may reduce MMRV costs, real-world scale causal data are needed – as for process-based models – to develop and independently validate the ability of these emerging approaches to reliably quantify SOC change due to agricultural interventions.

REFERENCES CITED

Amundson, R., Biardeau, L., 2018. Soil carbon sequestration is an elusive climate mitigation tool. Proceedings of the National Academy of Sciences 115, 11652–11656. *https://doi.org/10.1073/pnas.1815901115*

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

Angrist, J.D., Pischke, J.-S., 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. Journal of Economic Perspectives 24, 3–30. *https://doi.org/10.1257/jep.24.2.3*

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, 1433–1445. *https://doi.org/10.1111/gcb.15943*

Bradford, M.A., Carey, C.J., Atwood, L., Bossio, D., Fenichel, E.P., Gennet, S., Fargione, J., Fisher, J.R.B., Fuller, E., Kane, D.A., Lehmann, J., Oldfield, E.E., Ordway, E.M., Rudek, J., Sanderman, J., Wood, S.A., 2019. Soil carbon science for policy and practice. Nature Sustainability 2, 1070–1072. *https://doi.org/10.1038/s41893-019-0431-y*

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, 116719. *https://doi.org/10.1016/j.geoderma.2023.116719*

Bradford, M.A., Polussa, A., Eash, L., Jevon, F.V., Kuebbing, S.E., Rosenzweig, S., Oldfield, E.E., 2024. Research design, statistics and soil carbon: Responses to von Haden et al. and Spertus. Geoderma 448, 116961. *https://doi.org/10.1016/j.geoderma.2024.116961*

Bradley, V.C., Kuriwaki, S., Isakov, M., Sejdinovic, D., Meng, X.-L., Flaxman, S., 2021. Unrepresentative big surveys significantly overestimated US vaccine uptake. Nature 600, 695–700. *https://doi.org/10.1038/s41586-021-04198-4*

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*

Dangal, S.R.S., Sanderman, J., Wills, S., Ramirez-Lopez, L., 2019. Accurate and precise prediction of soil properties from a large mid-infrared spectral library. Soil Systems 3, 11. *https://doi.org/10.3390/soilsystems3010011*

Deines, J.M., Guan, K., Lopez, B., Zhou, Q., White, C.S., Wang, S., Lobell, D.B., 2023. Recent cover crop adoption is associated with small maize and soybean yield losses in the United States. Global Change Biology 29, 794–807. *https://doi.org/10.1111/gcb.16489*

Eagle, A.J., Olander, L.P., Locklier, K.L., Heffernan, J.B., Bernhardt, E.S., 2017. Fertilizer management and environmental factors drive N_2O and NO_3 losses in corn: A meta-analysis. Soil Science Society of America Journal 81, 1191–1202. *https://doi.org/10.2136/sssaj2016.09.0281*

EU, 2024. Regulation (EU) 2024/3012 of the European Parliament and of the Council. Official Journal of the European Union 3012, 1–29.

Even, R.J., Machmuller, M.B., Lavallee, J.M., Zelikova, T.J., Cotrufo, M.F., 2025. Large errors in soil carbon measurements attributed to inconsistent sample processing. SOIL 11, 17–34. *https://doi.org/10.5194/soil-11-17-2025*

FAO, 2024. Greenhouse gas emissions from agrifood systems - Global, regional and country trends, 2000–2022 (Analytical Brief No. 94), FAOSTAT Analytical Brief. Food and Agriculture Organization of the United Nations, Rome, Italy.

Fenster, T.L.D., LaCanne, C.E., Pecenka, J.R., Schmid, R.B., Bredeson, M.M., Busenitz, K.M., Michels, A.M., Welch, K.D., Lundgren, J.G., 2021. Defining and validating regenerative farm systems using a composite of ranked agricultural practices. F1000Research 10, 115. *https://doi.org/10.12688/f1000research.28450.1*

Garsia, A., Moinet, A., Vazquez, C., Creamer, R.E., Moinet, G.Y.K., 2023. The challenge of selecting an appropriate soil organic carbon simulation model: A comprehensive global review and validation assessment. Global Change Biology 29, 5760–5774. *https://doi.org/10.1111/gcb.16896*

Guenet, B., Gabrielle, B., Chenu, C., Arrouays, D., Balesdent, J., Bernoux, M., Bruni, E., Caliman, J., Cardinael, R., Chen, S., Ciais, P., Desbois, D., Fouche, J., Frank, S., Henault, C., Lugato, E., Naipal, V., Nesme, T., Obersteiner, M., Pellerin, S., Powlson, D.S., Rasse, D.P., Rees, F., Soussana, J., Su, Y., Tian, H., Valin, H., Zhou, F., 2021. Can N_2O emissions offset the benefits from soil organic carbon storage? Global Change Biology 27, 237–256. https://doi.org/10.1111/gcb.15342

REFERENCES CITED

Hagen, S.C., Delgado, G., Ingraham, P., Cooke, I., Emery, R., P. Fisk, J., Melendy, L., Olson, T., Patti, S., Rubin, N., Ziniti, B., Chen, H., Salas, W., Elias, P., Gustafson, D., 2020. Mapping conservation management practices and outcomes in the Corn Belt using the Operational Tillage Information System (OpTIS) and the Denitrification–Decomposition (DNDC) model. Land 9, 408. *https://doi.org/10.3390/land9110408*

Hayes, D.J., Ferruolo, S., Gajardo, D., Lu, L., McEvoy, K., Moore, K., Mui, K., Sachdeva, S., Tsao, A., Zehr, B., 2023. Data progress needed for climate-smart agriculture. Stanford Law School, Stanford, CA, USA.

Holland, P.W., 1986. Statistics and causal inference. Journal of the American Statistical Association 81, 945–960. *https://doi.org/10.1080/01621459.1986.10478354*

Huang, J., Hartemink, A.E., Kucharik, C.J., 2021. Soil-dependent responses of US crop yields to climate variability and depth to groundwater. Agricultural Systems 190, 103085. *https://doi.org/10.1016/j.agsy.2021.103085*

Imbens, G.W., 2024. Causal inference in the social sciences. Annual Review of Statistics and its Application 11, 123–152. *https://doi.org/10.1146/annurev-statistics-033121-114601*

IPCC, 2022. Summary for Policymakers, in: Shukla, P.R., Skea, J., Reisinger, A., Slade, R., Fradera, R., Pathak, M., Al Khourdajie, A., Belkacemi, M., van Diemen, R., Hasija, A., Lisboa, G., Luz, S., Malley, J., McCollum, D., Some, S., Vyas, P. (Eds.), Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Intergovernmental Panel on Climate Change, Cambridge, UK and New York, NY, USA, p. 48.

IPCC, 2019. Chapter 5: Cropland. Volume 4, in: 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories. p. *https://www.ipcc-nggip.iges.or.jp/public/2019rf/pdf/4_Volume4/19R_V4_Ch05_Cropland.pdf*.

IPCC, 2006. Chapter 5: Cropland. Volume 4, in: 2006 IPCC Guidelines for National Greenhouse Gas Inventories.

Janzen, H.H., Van Groenigen, K.J., Powlson, D.S., Schwinghamer, T., Van Groenigen, J.W., 2022. Photosynthetic limits on carbon sequestration in croplands. Geoderma 416, 115810. *https://doi.org/10.1016/j.geoderma.2022.115810*

Kane, D.A., Bradford, M.A., Fuller, E., Oldfield, E.E., Wood, S.A., 2021. Soil organic matter protects US maize yields and lowers crop insurance payouts under drought. Environmental Research Letters 16, 044018. *https://doi.org/10.1088/1748-9326/abe492*

Lavallee, J.M., McConkey, B.G., Kyker-Snowman, E., Dietze, M.C., Harrison, M.T., Smith, W., Grant, B., Karunaratne, S., Liu, K., Marklein, A., 2024. Modeling soil carbon and greenhouse gas emissions: Identifying challenges and advancing guidance for using process-based models in soil emission reduction and removal projects. Environmental Defense Fund, New York, NY, USA.

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., Schiedung, M., Schimel, J., Wieder, W., Abiven, S., Barré, P., Cécillon, L., Guenet, B., 2023. Soil organic carbon models need independent time-series validation for reliable prediction. Communications Earth & Environment 4, 158. *https://doi.org/10.1038/s43247-023-00830-5*

Lugato, E., Leip, A., Jones, A., 2018. Mitigation potential of soil carbon management overestimated by neglecting N_2O emissions. Nature Climate Change 8, 219–223. *https://doi.org/10.1038/s41558-018-0087-z*

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, 116647. *https://doi.org/10.1016/j.geoderma.2023.116647*

Minasny, B., Malone, B.P., McBratney, A.B., Angers, D.A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z.-S., Cheng, K., Das, B.S., Field, D.J., Gimona, A., Hedley, C.B., Hong, S.Y., Mandal, B., Marchant, B.P., Martin, M., McConkey, B.G., Mulder, V.L., O'Rourke, S., Richer-de-Forges, A.C., Odeh, I., Padarian, J., Paustian, K., Pan, G., Poggio, L., Savin, I., Stolbovoy, V., Stockmann, U., Sulaeman, Y., Tsui, C.-C., Vågen, T.-G., Van Wesemael, B., Winowiecki, L., 2017. Soil carbon 4 per mille. Geoderma 292, 59–86. *https://doi.org/10.1016/j.geoderma.2017.01.002*

Mitchell, E., Takeda, N., Grace, L., Grace, P., Day, K., Ahmadi, S., Badgery, W., Cowie, A., Simmons, A., Eckard, R., Harrison, M.T., Parton, W., Wilson, B., Orgill, S., Viscarra Rossel, R.A., Pannell, D., Stanley, P., Deane, F., Rowlings, D., 2024. Making soil carbon credits work for climate change mitigation. Carbon Management 15, 2430780. *https://doi.org/10.1080/17583004.2024.2430780*

Ogle, S.M., Breidt, F.J., Del Grosso, S., Gurung, R., Marx, E., Spencer, S., Williams, S., Manning, D., 2023a. Counterfactual scenarios reveal historical impact of cropland management on soil organic carbon stocks in the United States. Scientific Reports 13, 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, 810–822. *https://doi.org/10.1111/j.1365-2486.2009.01951.x*

Ogle, S.M., Conant, R.T., Fischer, B., Haya, B.K., Manning, D.T., McCarl, B.A., Zelikova, T.J., 2023b. Policy challenges to enhance soil carbon sinks: the dirty part of making contributions to the Paris agreement by the United States. Carbon Management 14, 2268071. *https://doi.org/10.1080/17583004.2023.2268071*

Oldfield, E.E., Bradford, M.A., Augarten, A.J., Cooley, E.T., Radatz, A.M., Radatz, T., Ruark, M.D., 2022a. Positive associations of soil organic matter and crop yields across a regional network of working farms. Soil Science Society of America Journal 86, 384–397. *https://doi.org/10.1002/saj2.20349*

Oldfield, E.E., Bradford, M.A., Wood, S.A., 2019. Global meta-analysis of the relationship between soil organic matter and crop yields. SOIL 5, 15–32. *https://doi.org/10.5194/soil-5-15-2019*

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

Oldfield, E.E., Lavallee, J.M., Blesh, J., Bradford, M.A., Cameron-Harp, M., Cotrufo, M.F., Eagle, A.J., Eash, L., Even, R.J., Kuebbing, S.E., Kort, E.A., Lark, T.J., Latka, C., Lin, Y., Machmuller, M.B., O'Neill, B., Raffeld, A.M., RoyChowdhury, T., Rudek, J., Sanderman, J., Sprunger, C.D., Toombs, T.P., Uludere Aragon, N., Vidal, M., Woolf, D., Zelikova, T.J., Gordon, D.R., 2024. Greenhouse gas mitigation on croplands: clarifying the debate on knowns, unknowns and risks to move forward with effective management interventions. Carbon Management 15, 2365896. *https://doi.org/10.1080/17583004.2024.2365896*

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

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, 024025. *https://doi.org/10.1088/1748-9326/ada16c*

Poulton, P., Johnston, J., Macdonald, A., White, R., Powlson, D., 2018. Major limitations to achieving "4 per 1000" increases in soil organic carbon stock in temperate regions: Evidence from long-term experiments at Rothamsted Research, United Kingdom. Global Change Biology 24, 2563–2584. *https://doi.org/10.1111/gcb.14066*

Prairie, A., Rosenzweig, S., Lundgren, J., Bredeson, M., Welch, K.D., Kane, D.A., Cotrufo, M.F., 2024. Integrating regenerative agriculture: Effects on soil arthropod biodiversity and soil organic matter dynamics. Preprint. *https://doi.org/10.2139/ssrn.4992411*

Randazzo, N.A., Gordon, D.R., Hamburg, S.P., 2023. Improved assessment of baseline and additionality for forest carbon crediting. Ecological Applications 33, e2817. *https://doi.org/10.1002/eap.2817*

Rattalino Edreira, J.I., Cassman, K.G., Hochman, Z., Van Ittersum, M.K., Van Bussel, L., Claessens, L., Grassini, P., 2018. Beyond the plot: technology extrapolation domains for scaling out agronomic science. Environmental Research Letters 13, 054027. *https://doi.org/10.1088/1748-9326/aac092*

Rothwell, P.M., 2005. External validity of randomised controlled trials: "To whom do the results of this trial apply?" The Lancet 365, 82–93. *https://doi.org/10.1016/S0140-6736(04)17670-8*

Rubin, D.B., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 66, 688–701. *https://doi.org/10.1037/h0037350*

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

REFERENCES CITED

Schwartzman, S., Lubowski, R.N., Pacala, S.W., Keohane, N.O., Kerr, S., Oppenheimer, M., Hamburg, S.P., 2021. Environmental integrity of emissions reductions depends on scale and systemic changes, not sector of origin. Environmental Research Letters 16, 091001. *https://doi.org/10.1088/1748-9326/ac18e8*

Siegel, K., Dee, L.E., 2025. Foundations and future directions for causal inference in ecological research. Ecology Letters 28, e70053. *https://doi.org/10.1111/ele.70053*

Surdoval, A., Jain, M., Blair, E., Wang, H., Blesh, J., 2024. Financial incentive programs and farm diversification with cover crops: assessing opportunities and challenges. Environmental Research Letters 19, 044063. *https://doi.org/10.1088/1748-9326/ad35d8*

US-EPA, 2022. Inventory of U.S. greenhouse gas emissions and sinks: 1990-2020 (No. EPA 430-R-22-003). US Environmental Protection Agency, Washington D.C., USA.

Wijewardane, N.K., Hetrick, S., Ackerson, J., Morgan, C.L.S., Ge, Y., 2020. VisNIR integrated multi-sensing penetrometer for in situ high-resolution vertical soil sensing. Soil and Tillage Research 199, 104604. *https://doi.org/10.1016/j.still.2020.104604*

Zomer, R.J., Bossio, D.A., Sommer, R., Verchot, L.V., 2017. Global sequestration potential of increased organic carbon in cropland soils. Scientific Reports 7, 15554. *https://doi.org/10.1038/s41598-017-15794-8*

Yale Applied Science Synthesis Program