IOP Publishing Environ. Res. Lett. 20 (2025) 024025 https://doi.org/10.1088/1748-9326/ada16c

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
15 July 2024

REVISED
9 October 2024

ACCEPTED FOR PUBLICATION
19 December 2024

PUBLISHED
17 January 2025

LETTER

Measure-and-remeasure as an economically feasible approach to crediting soil organic carbon at scale

Eric Potash[1,2,*], **Mark A Bradford**[3,4], **Emily E Oldfield**[3,5] **and Kaiyu Guan**[1,2]

[1] Agroecosystem Sustainability Center, Institute for Sustainability, Energy, and Environment, University of Illinois Urbana-Champaign, Urbana, IL 61801, United States of America
[2] Department of Natural Resources and Environmental Sciences, College of Agricultural, Consumer and Environmental Sciences, University of Illinois Urbana Champaign, Urbana, IL 61801, United States of America
[3] The Forest School, Yale School of the Environment, Yale University, New Haven, CT 06511, United States of America
[4] Yale Center for Natural Carbon Capture, Yale University, New Haven, CT 06511, United States of America
[5] Environmental Defense Fund, New York, NY, United States of America
[*] Author to whom any correspondence should be addressed.

E-mail: epotash@illinois.edu

Keywords: carbon markets, natural climate solutions, causal inference, climate smart agriculture, soil organic carbon

Supplementary material for this article is available online

Abstract

Agricultural carbon crediting predominantly relies on process-based biogeochemical models to estimate accrual of soil organic carbon stock (SOC). We investigate the conditions under which it may be economical to estimate SOC accrual by measuring and remeasuring SOC, which relies on fewer assumptions than modeling. We analyze multi-field measure-and-remeasure SOC projects with two key features: first, practice assignment is randomized to compare the effect of a treatment (e.g. no tillage) to a control (e.g. conventional tillage); second, a random subset of fields is sampled (two stage cluster sampling) to cost-effectively measure SOC changes. We use statistical modeling to characterize the estimated treatment effect, accounting for within-field and between-field variability in SOC change, as well as measurement error. We then use these statistics to evaluate how prices for measurement, treatment, and carbon credits influence the economics of measure-and-remeasure projects. We specifically investigate the potential advantages of larger spatial scale (number of fields) and temporal scale (years before remeasurement). We find economies of both spatial and temporal scale so that projects with thousands of fields, with only about 10% of fields measured for SOC change, are likely to yield a competitive return on investment in five years if the treatment effects found in the research literature can be achieved commercially. Our analysis suggests that measure-and-remeasure can be cost effective in both market and non-market SOC projects at scale. Moreover, measure-and-remeasure projects provide valuable data for independent validation on commercial farms of the accrual rates estimated by biogeochemical models using field trials. We provide next steps and software for researchers, credit registries, and project developers to move forward with measure-and-remeasure SOC projects.

1. Introduction

Carbon crediting aims to incentivize practices that reduce greenhouse gas emissions. Climate-smart agricultural practices (e.g. no-tillage and cover cropping) that can accrue soil organic carbon stock (SOC) have received particular attention (IPCC 2022). Measurement, reporting, and verification (MRV) protocols set standards for how projects can receive credits (Oldfield *et al* 2022). In general, a project must estimate the total change in SOC under the project, as well as the counterfactual total change that would have occurred without the project. The difference between these changes is the SOC treatment effect and is attributed to the project. Historically, soil surveys and research have measured SOC change using resource-intensive soil sampling and lab analysis. Many soil cores must be sampled since annual changes in SOC are small relative to their spatial variability and measurement error. The high demands of measurement have raised questions about the economic feasibility of a 'measure-and-remeasure' approach to SOC crediting (Bradford *et al* 2023).

To reduce costs, alternatives to traditional SOC measurement are being developed including biogeochemical modeling (Mathers *et al* 2023) as well as *in-situ* (Wijewardane *et al* 2020) and remote (Wang *et al* 2022) sensing. In particular, as evidenced by its primacy in MRV protocols (Oldfield *et al* 2022), an approach called 'measure-and-model' currently dominates. In measure-and-model, traditional measurements of SOC are taken before a project starts. These measurements are used to initialize a process-based biogeochemical model. Then, each year, the model estimates changes in SOC at each measurement location under the implemented practice, as well as under a control (also known as baseline, counterfactual, or business-as-usual) practice.

Besides the lower cost of modeling compared to measurement, there are other reasons why SOC project developers may prefer modeling. First, there is a perception that measurement alone is not economical for generating carbon credits (Bradford *et al* 2023). Some major MRV protocols do not even accommodate measure-and-remeasure (e.g. CAR SEP). Those that do allow it, may be unnecessarily onerous (Verra VM0042; see discussion). Second, because developers can model their project under a variety of scenarios prior to implementation, the carbon credits that the project will generate are well understood, making earnings more reliable.

We investigate whether measure-and-remeasure can be an economically feasible approach to SOC crediting. Our motivation is that biogeochemical models are developed using field trial data and we lack evidence about their predictive performance for SOC change across commercial farms (Ellis and Paustian 2024). Measurement at commercial scale is critical for building confidence that the effects of climate-smart agricultural practices documented in the academic literature for small experimental plots, and predicted using soil biogeochemical models, can be realized on commercial farms (Oldfield *et al* 2024). Compared to measure-and-model, measure-and-remeasure approaches at commercial scales are more direct and make fewer, more transparent assumptions. As a result, there can be greater confidence that SOC credits generated from well-designed measure-and-remeasure projects reflect a real reduction (relative to the counterfactual) in atmospheric CO_2 concentration. While our focus is market-based projects that sell carbon credits, we also discuss implications for non-market projects (scope 3 emissions 'insetting').

However, the feasibility of measure-and-remeasure for estimation of small changes in SOC has been questioned given the challenges of spatial variability and measurement error (Bradford *et al* 2023). Indeed, on small spatial scales such as an experimental plot or agricultural field, these challenges have been shown to result in sample sizes that are prohibitively large for both inventorying and monitoring SOC (Smith 2004). However, there is some cause for optimism on longer time scales and larger spatial scales (Saby *et al* 2008, Bradford *et al* 2023) where we only need to estimate the *average* treatment effect rather than the effect in any individual field. But there are several gaps in applying these studies to evaluating the economic feasibility of agricultural SOC crediting projects. Saby *et al* (2008), for instance, estimate sampling demands for a national monitoring network including woodland and grassland which, due to the higher SOC variability that they document, require more sampling resources than farmland. Moreover, their objective is SOC monitoring rather than crediting so they do not estimate the treatment effect due to a change in management. While Bradford *et al* (2023) study SOC crediting, they only consider project sizes up to 60 fields, whereas commercial projects can be on the order of 10 000 fields (typically 10–50 ha per field). Finally, to evaluate the economic feasibility of an SOC crediting project, all relevant costs and revenues should be quantified. Costs include sampling as well as practice implementation, while revenues are due to SOC credits. But neither Saby *et al* (2008) nor Bradford *et al* (2023) considers economic feasibility, instead defining feasibility in terms of sample size. While de Gruijter *et al* (2016) perform an economic optimization, they do not include the cost of practice implementation and their project design does not estimate an SOC treatment effect nor do they quantify the revenue from SOC credits, so they do not evaluate economic feasibility.

Our work addresses these limitations of the SOC measurement literature by simultaneously considering (1) economic feasibility, (2) large spatial scales and (3) SOC treatment effects (as opposed to inventory or monitoring). Our specific contributions are as follows. First, we develop a multilevel statistical model of measure-and-remeasure SOC projects that accounts for the challenges of within- and between-field variability in SOC changes and measurement error. Second, we design a measure-and-remeasure project that supports efficient and rigorous inference of the treatment effect (figure 1). To rigorously estimate average SOC changes under both treatment and control, the design includes randomized management practice assignment (among non-randomly enrolled fields). For efficiency, the design also includes a two stage cluster sampling strategy, in which a subset of fields are randomly selected for sampling. A key parameter of the project is its scale, both spatial and temporal. We investigate project spatial scales from 100 to 100 000 fields and temporal

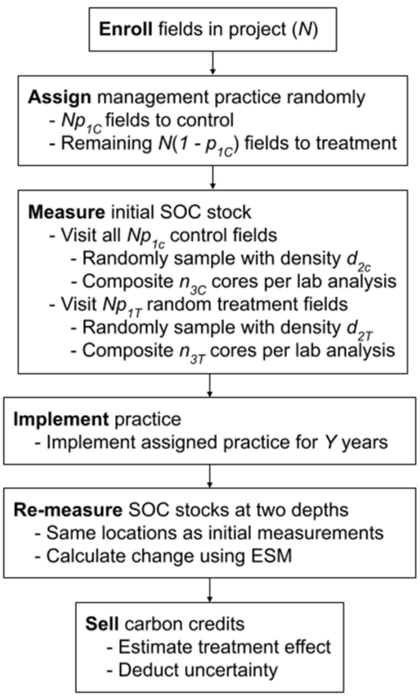

Figure 1. Flowchart of proposed SOC crediting project using randomized treatment assignment, two stage sampling, and compositing. Abbreviations: SOC (soil organic carbon stock), ESM (equivalent soil mass).

scales from 1 to 10 years. Finally we evaluate the economic feasibility of these projects as a function of their scale and economic parameters such as prices for sampling, treatment, and carbon.

2. Materials and methods

First we parameterize the dynamics and economics of a measure-and-remeasure SOC project with randomized practice assignment and two stage random sampling in which a random subset of fields are selected for measurement. Next we discuss economic feasibility. Finally we describe how to answer questions about economic feasibility using numerical optimization.

2.1. Project parameters

We statistically model the SOC project using four categories of parameters (table 1): scale (number of fields N and years to remeasurement Y), design (e.g. proportion of fields assigned to control), carbon dynamics (e.g. average treatment effect), and economics (e.g. price for carbon credits). We choose default values for the SOC and economic parameters in a way that is optimistic but conservative (see SI) and also assess the sensitivity of our results to these choices.

We directly model the SOC changes (table 1(b)) on which credits are issued rather than the absolute SOC. We decompose the changes into a population average treatment effect ($\tilde{\tau}$) with between-field variability (σ_b) determining field-level average changes and within-field variability (σ_w) determining location-level changes (figure 2). Note that between-field variability $\sigma_b = 0.5$ was estimated (see SI) from a no-till study 'representing different soil associations and precipitation distributions across Iowa' (Al-Kaisi and Kwaw-Mensah 2020). The SOC changes are not observed directly but with error due to relocation (σ_n) since we cannot sample the same core twice (Poeplau *et al* 2022, Lark 2009), as well as lab error (σ_l). Note that our choice of $\sigma_l = 2$ Mg ha^{-1} corresponds to a relative error of 4% in SOC concentration and 2% in bulk density for 0–30 cm samples with average SOC concentration 2% and bulk density 1.5 g cm^{-3}.

The project design includes a randomized experiment and two stage random sampling. A proportion p_{1C} of the N fields in the project are assigned to the control practice. All of these fields are sampled. Similarly a proportion p_{1T} of the N fields are treated and sampled. The remaining fields are treated and unsampled. The sampling density (cores ha^{-1}) in control and treatment groups are d_{2C} and d_{2T}, respectively. We assume uniformly random selection of both practice assignment as well as fields for monitoring and locations within each monitored field for sampling. This design is conservative given the potential for techniques such as stratification and pairing to reduce estimation uncertainty (Potash *et al* 2023). Finally, within each field, random sets of n_{3C} and n_{3T} cores in control and treatment fields, respectively, are composited for lab analysis. Following standard formulas (Brus 2022), the squared standard error of the estimated mean change in the control group after Y years is

$$\mathrm{SE}^2\left(\widehat{\Delta\mathrm{SOC}_{\mathrm{C},Y}}\right) = \frac{Y{\sigma_b}^2}{Np_{1\mathrm{C}}} + \frac{Y{\sigma_w}^2 + 2{\sigma_n}^2}{Np_{1\mathrm{C}}Ad_{2\mathrm{C}}} + \frac{2{\sigma_l}^2}{Np_{1\mathrm{C}}Ad_{2\mathrm{C}}/n_{3\mathrm{C}}} \quad (1)$$

with a similar squared standard error $\mathrm{SE}^2(\widehat{\Delta\mathrm{SOC}_{\mathrm{T},Y}})$ in the treatment group (see SI). Since these sampling errors are independent, the squared standard error of the estimated treatment effect is the sum of the squared standard errors in each group:

$$\mathrm{SE}^2\left(\widehat{\tilde{\tau}_Y}\right) = \mathrm{SE}^2\left(\widehat{\Delta\mathrm{SOC}_{\mathrm{C},Y}}\right) + \mathrm{SE}^2\left(\widehat{\Delta\mathrm{SOC}_{\mathrm{T},Y}}\right). \quad (2)$$

These design parameters are selected to maximize economic feasibility (see below).

The project incurs two costs: measurement and payments to farmers. For measurement, there is a cost per field visited, per core sampled and processed, and per (composite) sample analyzed, with default values based on labor and transportation costs and a survey of commercial services currently available in the US (see SI). Note that we assume that at remeasurement two depths are sampled to enable equivalent soil mass (ESM) comparison (von Haden *et al* 2020). Farmers

Table 1. Parameters describing (a) scale, (b) SOC, (c) design, and (d) economics. Abbreviations: SOC (soil organic carbon stock), SD (standard deviation).

(a) Project scale parameters		
	Description	Comments
N	Number of fields in project	Larger N increases potential SOC stored at project level
Y	Time until remeasurement (years)	Larger Y increases SOC accrued at cost of waiting for credits

(b) SOC dynamics parameters			
	Description	Default	Comments
$\tilde{\tau}$	Average treatment effect (Mg C ha^{-1} y^{-1})	0.3	Larger $\tilde{\tau}$ makes projects more feasible
σ_b	Between-field SD in SOC change (Mg ha^{-1} y^{-1})	0.5	Larger σ_b incentivizes sampling more fields
σ_w	Within-field SD in SOC change (Mg ha^{-1} y^{-1})	1	Larger σ_w incentivizes more samples within field
σ_n	Relocation SD (Mg ha^{-1})	5	Larger σ_n incentivizes more samples within each field
σ_l	Lab measurement SD (Mg ha^{-1})	2	Larger σ_l incentivizes more assays
A	Field size (ha)	25	Larger A increases potential SOC stored per field

(c) Experimental and sampling design parameters		
	Description	Comments
p_{1C}	Proportion of fields assigned to control	Larger p_{1C} reduces uncertainty in control estimates at the cost of storing carbon at project level
d_{2C}	Control sampling density (samples ha^{-1})	Larger d_{2C} reduces uncertainty in the effect of control practice
n_{3C}	Control cores composited per assay	Larger n_{3C} increases lab errors in control estimates
p_{1T}	Proportion of fields treated and sampled	Larger p_{1T} reduces uncertainty in treatment estimates
d_{2T}	Treatment sampling density (samples ha^{-1})	Larger d_{2T} reduces uncertainty treatment estimates
n_{3T}	Treatment cores composited per assay	Larger n_{3T} increases lab errors in treatment estimates

(d) Economic parameters			
	Description	Default	Comments
P_{CO2}	Sale price of carbon credit (\$/Mg CO_2)	40	Larger P_{CO2} increases carbon credit revenues
z_{deduct}	Uncertainty deduction (unitless)	0.43	Larger z_{deduct} incentivizes more sampling
c_C	Control practice payment (\$ ha^{-1} y^{-1})	25	Larger c_C increases costs
c_T	Treatment practice payment (\$ ha^{-1} y^{-1})	25	Larger c_T increases costs
c_{1s}	Cost to visit a field for sampling (\$)	400	Larger c_{1s} incentivizes sampling fewer fields, higher density
c_{2s}	Cost to sample and process a location in a field (\$)	15	Larger c_{2s} incentivizes sampling more fields, lower density
c_l	Cost to analyze a sample (\$)	20	Larger c_l incentivizes more compositing
r_b	Alternative investment real annual return (%)	5	Larger r_b incentivizes shorter remeasurement Y

are paid per hectare per year for both treatment and control practice. The cost of no-till varies (Havens *et al* 2023) and may actually reduce costs compared to conventional tillage (Che *et al* 2023). We set a default treatment payment of $c_T = \$25\,\text{ha}^{-1}\,\text{y}^{-1}$ and an equal control payment. Project earnings are from selling carbon credits, with a default price of $P_{CO2} = \$40$ per tonne CO_2 which is conservative compared to recent SOC credit sales of \$100 per tonne. The number of credits sold is equal to the estimated treatment effect $\widehat{\tilde{\tau}}$ minus an uncertainty deduction, which is assumed (following Verra VM0042 version 2.0) to be 0.43 standard errors SE($\widehat{\tilde{\tau}}$) (SI equation (21)). We note that farmers do not assume any financial risk

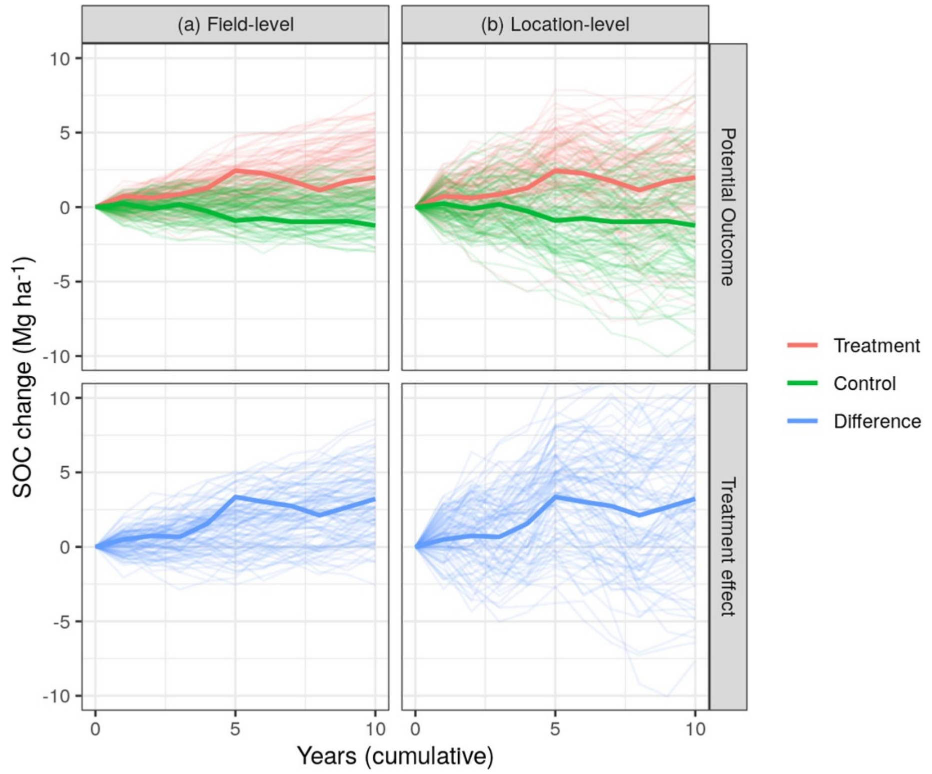

Figure 2. Simulated SOC changes over $Y = 10$ years with mean effect $\tilde{\tau} = 0.3$ Mg C ha^{-1} y^{-1} at (a) the field level and (b) the location level. In the top panel of (a) each line is the average SOC change across a field (100 fields shown) under treatment (red) and control (green). The between-field variability is governed by $\sigma_b = 0.5$. The thick lines highlight a single field that is explored in (b), where each line is a location (100 locations shown) within the highlighted field. The within-field variability is controlled by $\sigma_w = 1$. The bottom panels show the corresponding treatment effects, i.e. the differences between treatment and control changes on the field or location level. Since each field receives either treatment or control, treatment effects are unobserved. Moreover, treatment and control changes are measured with relocation error (σ_n) and lab error (σ_l). Note these simulations use normal distributions and zero trend in the control group, though this is not an assumption in our analysis. Abbreviations: SOC (soil organic carbon stock).

for treatment effectiveness: farmer payments here are made annually for practice change (treatment group) and practice maintenance (control group) and do not depend on SOC accrual. However, after SOC credits are sold, any profits may be shared with farmers.

2.2. Economic performance

To summarize the economic performance of the SOC project we use the Sharpe ratio (Sharpe 1966). In business and finance, the Sharpe ratio is a common measure of the performance of an investment defined as

$$\frac{E[\mathrm{ROI}] - \mathrm{ROI}_b}{\sqrt{\mathrm{Var}[\mathrm{ROI}]}} \tag{3}$$

where $E[\mathrm{ROI}]$ is the expected return on the investment (SOC project), $\sqrt{\mathrm{Var}[\mathrm{ROI}]}$ is its standard deviation, and ROI_b is the return of an alternative investment. Here we set ROI_b to be the minimum return for the SOC project to be attractive. The numerator is thus the 'excess return', in expectation, of the carbon project over the minimum. We calculate ROI_b by assuming that the costs of the project (measurements and farmer practice payments; see SI) are instead invested with a real annual return r_b. Assuming a private equity 'hurdle' of 7% and a 2% inflation rate, we set $r_b = 5\%$. Note that we have not adjusted the prices (table 1(d)) in the SOC project for inflation. If these prices keep pace with inflation, the project ROI is then also in real terms. However we note that there are many factors influencing these choices that are beyond the scope of this study.

The Sharpe ratio was selected to assess economic feasibility because it captures the following important features: the expected return including an uncertainty deduction due to measurement variability (equation (2)); the time value of money, as captured by the alternative investment; uncertainty, here due to measurement variability as well as variation in the treatment effect (SI, equation (23)). We note that de Gruijter et al (2016) also analyzed the economics of SOC crediting (albeit for SOC change rather than a treatment effect) though their objective only included the first of these three features.

A large Sharpe ratio can come about in two ways: a large expected return on investment or a small variability in return on investment (or both). Maximizing the Sharpe ratio is akin to maximizing the probability that the investment outperforms the alternative investment. If returns follow any of

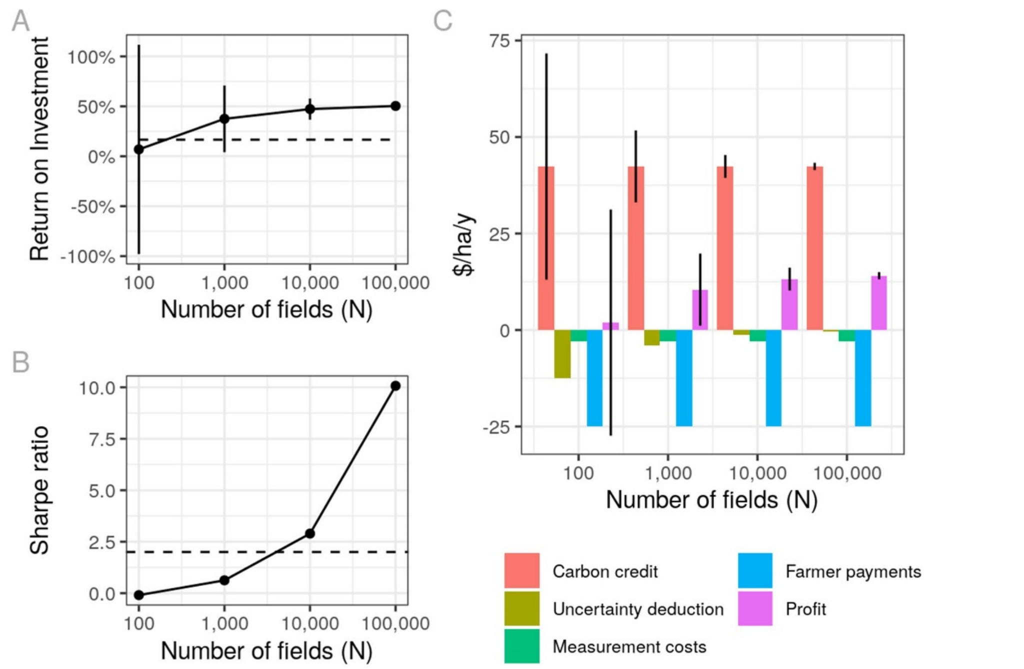

Figure 3. Economies of spatial scale in a project with 5 years to re-measurement. (A) The return on investment in expectation (dots) increases with the number of fields while the standard deviation (error bars) decreases. The dashed line shows the return on an alternative investment for comparison. (B) The Sharpe ratio, a summary measure of the returns taking into account their expected value, standard deviation, and the alternative. An investment with a Sharpe ratio of 2 (dashed line) or above is generally considered attractive. (C) Distribution of costs and earnings. As the size of the project increases, the variability in the carbon credit earnings decreases, as does the uncertainty deduction. The measurement costs and farmer payments are constant. Thus the profit increases and its uncertainty (error bar) decreases.

a variety of probability distributions then a Sharpe ratio exceeding 2.0 implies that the investment outperforms the alternative with approximately 95% probability. Note this is equivalent to the net present value of the investment having a 95% probability of being greater than zero. Although we recognize that investment decisions are influenced by multiple factors, in our analysis we define an SOC project to be economically feasible if its Sharpe ratio exceeds 2.0.

2.3. Optimization

Using the equations above, we can calculate the expected return on investment as well as its standard deviation as a function of the parameters (table 1). Then we can calculate the Sharpe ratio to summarize economic performance (equation (3)). We then perform two kinds of optimization. First, viewing the carbon and economic parameters as well as the number of fields (N) and years to remeasurement (Y) as fixed, we can maximize the Sharpe ratio, as a function of the remaining design parameters (table 1(c)). If this maximized Sharpe ratio exceeds 2.0 we say that this project is economically feasible. Second, we can target a particular level of an outcome, e.g. economic feasibility defined as a Sharpe ratio equal to 2.0, and then optimize a different parameter using constrained optimization. For example, we can find the minimum carbon price P_{CO2} that is economically feasible, i.e. the sensitivity of economic feasibility to the carbon price. Optimization, analysis and visualization were performed in R version 4.0.3 (R Core Team 2020).

3. Results

We investigate the effects of spatial and temporal scale on the economic feasibility of measure-and-remeasure SOC projects. The parameters in our model are initially chosen to reflect the SOC characteristics and economics of no-till agriculture in the U.S. Midwest. We then assess the sensitivity of our analysis to these choices, and in the discussion we comment specifically on practices such as cover cropping as well as combinations of multiple practices. These results and their sensitivity to assumptions can be interactively explored at https://asc.illinois.edu/soc-econ (R source code provided).

3.1. Economy of spatial scale

We start by investigating, under the default SOC and economic parameters, projects with Y =5 years to remeasurement and increasing numbers of fields N from 100 to 100 000. Note that SOC crediting projects on this scale are currently being developed using measure-and-model approaches (Indigo 2024). We find that larger projects are able to achieve higher expected returns as well as reduce the uncertainty in their returns (figure 3). Under our default SOC and

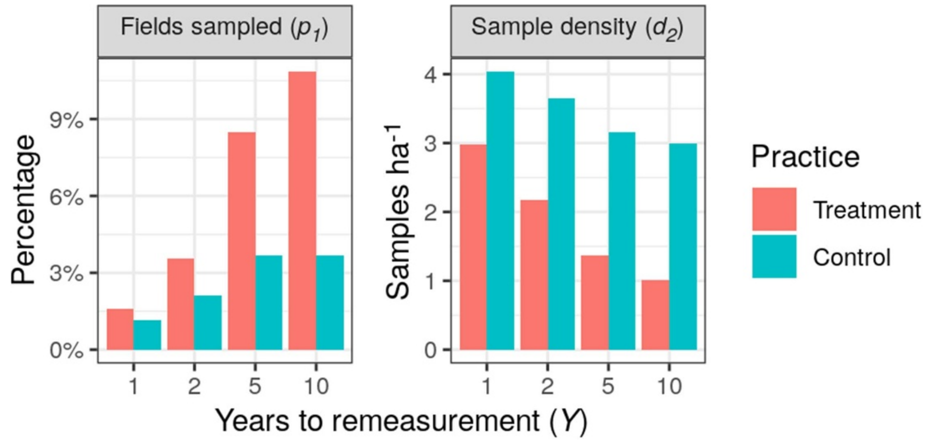

Figure 4. Allocation of samples optimized to maximize Sharpe ratio as a function of the number of years to remeasurement (Y). These values are approximately independent of the project size (N). As Y increases, between-field variation also increases while within-field variation increases relatively slower since most within-field variation is due to relocation error (equation (1)). As a result, it is advantageous to increase the proportion of fields sampled $p_{1\cdot}$ while decreasing sample density $d_{2\cdot}$ within those fields. We also note that, since control fields do not generate carbon credits, it is (for any Y) more profitable to allocate fewer control fields (blue bars shorter than red on left) but sample them more densely (blue bars taller than red on right). Finally, the optimal number of cores per composite lab analysis is 4 for all years in both treatment and control.

economic parameters, projects over about $N = 5000$ fields meet our economic feasibility criterion of a Sharpe ratio exceeding 2.0.

We find that the optimal project design depends on the time to remeasurement and is approximately independent of the number of fields (figure 3). For the 5 years to remeasurement considered here, measurements cost about \$3 per hectare per year, much less than the farmer payments which are \$25 per hectare per year. The optimal proportion of fields allocated to the control group (all of which are sampled) is $p_{1\mathrm{C}} = 4\%$, with $p_{1\mathrm{T}} = 8\%$ of all fields being treated and sampled, and the remaining 88% of fields being treated and unsampled. In the control fields, a density of $d_{2\mathrm{C}} = 3$ samples ha^{-1} is optimal, whereas within the treatment fields the optimal density is $d_{2\mathrm{T}} = 1$ samples ha^{-1}. We note that the greater number of treated than control fields sampled reflects the fact that every additional control field is one fewer treatment field from which to earn SOC credit revenue. The greater sampling density in control compared to treatment fields helps to equalize the estimation error in the control group compared to the treatment group. Finally, in both treatment and control groups, each lab analysis is performed on a composite of $n_{3\mathrm{C}} = n_{3\mathrm{T}} = 4$ random cores per field, resulting in about 1 lab analysis ha^{-1}.

Since the optimal sampling parameters are independent of the spatial scale, the project costs (measurement and treatment) per hectare per year are also independent of the spatial scale. The economies of scale are then a consequence of reduced uncertainty in estimated SOC accrual, which leads to both greater expected revenue (through a smaller uncertainty deduction), as well as reduced variability in revenue (figure 3).

The underlying cause of declining uncertainty with increasing spatial scale is that the sources of variability (equations (1) and (2)) do not increase proportionally with increasing N. An assumption within our model was that between-field variability $\sigma_b = 0.5$, is independent of the number of fields N. We note that in practice σ_b may increase with N. For example, a small project, if it occurs in a single county, may have smaller σ_b than a large, multi-state project. However, to negate the economy of spatial scale in between-field variability, we would need σ_b in a 10 000 field project to be about 10 times larger than in a 100 field project, which is extremely unlikely due to physical constraints on the range of SOC changes. Moreover, the other sources of variability (within-field, relocation, and lab) are unlikely to increase with spatial scale. In conclusion, we have high confidence in the economy of spatial scale in measure-and-remeasure SOC projects.

3.2. Economy of temporal scale

Next we investigate how the economics of a carbon credit project of a fixed size $N = 10\,000$ fields vary as a function of the number of years to remeasurement Y. Unlike with spatial scale above, the optimal design varies strongly with temporal scale (figure 4). As Y increases, the proportion of fields sampled increases while the sampling density decreases. This is driven by the fact that, as Y increases, between-field variability of SOC changes increases faster than within-field variability, which is dominated by a fixed relocation error.

This optimized design produces an economy of temporal scale. A 10 000 field project with 10 years to remeasurement has a Sharpe ratio of 3.6 (figure S1), suggesting a less than 0.1%

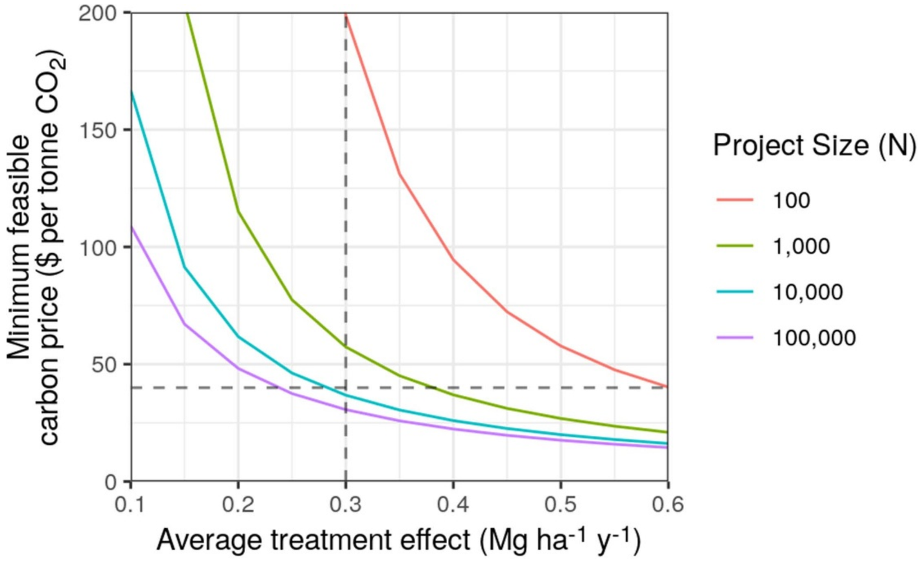

Figure 5. Sensitivity of minimum feasible carbon price (P_{CO2}, vertical axis) of $Y = 5$ year projects as a function of average treatment effect ($\tilde{\tau}$, horizontal axis) and project size (N, color), with other parameters at their default values (table 1). For example, with $\tilde{\tau} = 0.3$ Mg C ha^{-1} y^{-1}, a project of $N = 100$ fields (red curve) is economically feasible at a carbon price of at least $200 while a project of $N = 100\,000$ fields (purple curve) is feasible down to $31. Dashed lines indicate default values $\tilde{\tau} = 0.3$ and $P_{C02} = \$40$.

probability of performing worse than an alternative investment returning 5% annually. This is despite the fact that, as Y increases, the absolute return of the alternative investment increases to over 25%.

Similar to the economy of spatial scale, we emphasize that this economy of temporal scale is a consequence of the nature of SOC change and measurement variability. While two sources of uncertainty (within- and between-field variability in SOC changes) do increase with time, the other two sources of uncertainty (relocation and lab variability) do not increase with time. Thus our confidence is high that increasing (to a point) the number of years to remeasurement increases the economic feasibility of a measure-and-remeasure SOC project.

3.3. Sensitivity analysis

If a project is economically feasible at our default carbon price of $P_{CO2} = \$40$ per tonne, then it is feasible at higher prices as well. But we may ask, how low can the price be before the project is infeasible? Conversely, for projects that were infeasible at $40, we may ask, how high must the price be before they become feasible? We address these questions using constrained optimization. We find that larger projects are feasible down to carbon prices of about $30 per tonne and that projects as small as 250 fields are feasible at a price of $100 per tonne on a 5 year timeframe (figure 5; box 1).

We performed similar sensitivity analyses for all of the major parameters in our model (figure S2). In summary, in addition to the carbon price and average treatment effect above, another parameter that was an important constraint on economic feasibility was the treatment payment. With other parameters at their default values, the highest treatment payment that was economically feasible was $c_T = \$30$ ha^{-1} y^{-1}. However this is not surprising since under the default assumptions of $P_{CO2} = \$40$ per tonne CO_2 and $\tilde{\tau} = 0.3$ Mg ha^{-1} y^{-1}, the implied value of treatment is $44 ha^{-1} y^{-1}.

In contrast, the variability and measurement cost parameters did not pose a major constraint on economic feasibility for larger projects on longer timeframes ($\geqslant$5 years). We also considered the possibility of autocorrelated changes in field- and location-level SOC and found our economic feasibility results to be robust to this modification (figure S3).

Box 1. Example project of 250 fields over 5 years and $100 per tonne CO_2.

First, $N = 250$ fields spanning 6250 ha are enrolled. Next, 34 fields ($p_{1C} = 14\%$) are randomly selected for control practice, with the remaining fields being treated. All p_{1C} fields, along with 95 treatment fields ($p_{1T} = 38\%$), are sampled to 30 cm at densities of 3.8 and 1.4 samples ha^{-1}, respectively. Within each sampled field, random groups of 4 cores each are composited for lab analysis (about 1.0 and 0.4 analyses ha^{-1} in control and treatment, respectively). We note that because only 52% of fields are sampled, averaged across the project area this amounts to 1.1 samples ha^{-1} and 0.3 lab analyses ha^{-1}. Following this baseline sampling, the treatment group receives treatment for 5 years. Follow-up sampling occurs at the same fields and locations, with two depths per core to facilitate ESM comparison. These measurements are used to issue and sell carbon credits. The total cost of measurement is about \$400 000 (\$13 ha^{-1} y^{-1}) while the total farmer payments are about \$800 000 (\$25 ha^{-1} y^{-1}). The project is expected to accrue about 8000 Mg C (0.26 Mg C ha^{-1} y^{-1}). The standard error on the accrual is about 2000 Mg C (0.05 Mg C ha^{-1} y^{-1}), so that after uncertainty deduction the project is credited for about 7000 Mg C (0.24 Mg ha^{-1} y^{-1}). At a price of \$100 per tonne CO_2 the project is thus expected to earn \$3 $\pm$ 0.7 million (\$95 $\pm$ 21 ha^{-1} y^{-1}). This gives a return on the \$1.2 million investment of 127% $\pm$ 56%. Had the sampling and measurement costs over these five years been put toward an investment with 5% annual return, the total return would have been 18%. Thus this project has a Sharpe ratio of 2.

4. Discussion

There is a general perception that measuring SOC treatment effects in commercial projects is economically infeasible due to the small signal of change relative to spatial and measurement variability (Smith 2004, Bradford *et al* 2023). To investigate this perception, we developed a statistical model of multi-field agricultural SOC dynamics that represented the challenges of between-field, within-field, and measurement variability in SOC. Coupling the statistical model with prices for measurement, practice adoption, and carbon credits, we confirmed that measure-and-remeasure is not economical for small projects on short timeframes. However, we found that larger projects and/or those that wait longer for remeasurement are economical under our default assumptions. Only a small proportion of fields are selected for sampling so that, while selected fields are sampled intensively, sampling demands across the project are relatively low. In such projects, measurement makes up a small fraction of the total costs, which are dominated by the treatment itself (i.e. annual payments to farmers for practice change, such as no-tillage). Our results, based on a sensitivity analysis, appear robust and even optimistic when some of the default parameters are perturbed. For example, raising the price of SOC credits from \$40 to \$100 per tonne CO_2 allowed much smaller projects (250 fields, with a subset remeasured after 5 years: box 1) to be attractive investments. We note that SOC credits have recently been sold at this price. This evidence adds to recent work on the feasibility of estimating SOC changes and treatment effects (Bradford *et al* 2023).

Compared to past studies of SOC measurement, our study is more directly relevant to agricultural SOC crediting projects (see introduction) and our findings are more optimistic for two reasons. First, we analyzed a project that used a two stage cluster sampling design which first randomly selects a subset of fields and second randomly selects locations within those fields to sample. While this design is common in MRV protocols, we are not aware of any SOC measurement research analyzing it. Compared to the more common research designs of simple random (e.g. Saby *et al* 2008) or stratified sampling (e.g. Bradford *et al* 2023), two stage cluster sampling is much more efficient, especially for larger projects (figure S4) since it only samples a subset of fields. Second, past research has largely assessed feasibility in terms of sample densities, which requires a subjective judgment about which sample densities are feasible. However, we posit that a sample density per se is neither feasible nor infeasible. Instead, we used the criterion of economic feasibility in which a sample density is feasible if its costs are outweighed by the revenues from SOC credits that the measurements generate and we developed a framework to estimate these costs and revenues. We note that in two stage cluster sampling only a small subset of fields are sampled so that while sampled fields may have relatively high sampling densities, total sample size across the project is much lower (box 1).

Two major drivers of economic feasibility in measure-and-remeasure projects are spatial and temporal economies of scale. We have high confidence in these benefits because they are consequences of basic properties of SOC variability (see results). Note that, for lack of information, we assumed a fixed between-field variability (σ_b) and cost per field visit (c_{1s}). As a result, the economy of spatial scale is best conceived as increasing the number of fields within a fixed geographic extent. However, even with data on how these parameters vary with the geographic extent of a project, we expect our qualitative results to continue to hold because of physical limits on the range of between-field variability (σ_b) and the fact that per core (c_{2s}) sampling costs dominated per field (c_{1s}) costs.

Our project used a randomized controlled trial (RCT) design to address the so-called *fundamental problem of causal inference* (Holland 1986): although we can directly measure SOC changes in treated fields, we cannot also measure the changes that would have occurred in those same fields under the counterfactual control practice. The RCT addresses this problem by assigning treatment or control practice randomly among enrolled fields, relying on randomization to allocate fields across confounding predictors of SOC change and hence control for them through study design. The result is that well-designed RCTs have high internal validity, i.e. the ability to robustly quantify the effect of treatment in the study population, which is the population of interest in a carbon project. An alternative to this RCT design, now represented in some of the major agricultural SOC protocols (e.g. VM0042), is to treat all fields in a project and pair them with control fields outside of the project. These control fields can then be re-used across projects (Oldfield *et al* 2022), which would be more economical. However, pairing has disadvantages compared to randomization: we can only pair on observable characteristics, which precludes controlling for unobserved differences, e.g. a farmer's skill in sowing cover crops. Moreover, given the large number of variables influencing SOC dynamics (VM0042 matches on at least 14 variables), finding appropriate matches may be difficult for some fields, and may require more control fields than our RCT. We recommend broader consideration of our RCT design as a simpler and more rigorous way to measure treatment effects and issue carbon credits.

The measure-and-remeasure RCT project design is not a panacea, and will have to deal with some challenges. For example, if there is non-compliance with practice assignment, estimation of the treatment effect should be adjusted to account for it (Hernán and Robins 2020). Randomization does not address additionality or permanence, two general concerns for SOC projects. Regarding additionality, it is not possible to ensure that the management practice at control sites is representative of the counterfactual practice that would have occurred absent the project and therefore whether the SOC accrual measured in the RCT is additional. Regarding permanence, SOC credits are typically based on climate benefits on a 100 year timescale. But uncertainty about future management and climate is not addressed by shorter term project design, regardless of the quantification approach (measure-and-remeasure or measure-and-model).

Our economic feasibility results would have been different had we considered practices such as cover cropping which can cost about \$100 per hectare per year (Bowman *et al* 2022), compared to the \$25 we assumed for no-till. At current carbon credit prices (about \$40 per tonne CO_2), cover cropping costs likely exceed the market value of the carbon accrued, regardless of the quantification approach. It is possible that cover cropping is economical when implemented in combination with other practices such as no-till. But we note that earnings generated from selling carbon credits are not the sole motivation for climate-smart agriculture. Indeed, additionality criteria for agricultural SOC protocols do not depend on a project generating profit. Instead, carbon financing simply has to be a necessary factor for a farmer to make the practice change. Adoption of cover crops, for example, can improve soil health, reduce fertilizer application, and support adaptation to extreme weather (Bergtold *et al* 2019). In a measure-and-remeasure project, the measurements themselves provide valuable knowledge. Thus SOC credits are one of many benefits to consider in planning a climate-smart agriculture project.

We suggest that measure-and-remeasure projects can help build confidence that SOC credits represent real climate benefits (Badgley *et al* 2022, West *et al* 2023), ensuring continued supply of carbon financing. For example, measure-and-model approaches are currently informed and validated using published small plot field trials (Mathers *et al* 2023). There are concerns that these trials do not have the external validity necessary to independently verify measure-and-model issued credits because, for example, the small plot size and experimenter-controlled management is not representative of the size nor management of commercial fields (Ellis and Paustian 2024, Kravchenko *et al* 2017). Measure-and-remeasure projects at commercial scale then can provide data which can be used to independently verify measure-and-model credits. Such an objective could also be used, outside of carbon crediting, to test and develop the ability of soil biogeochemical models to estimate SOC change at sub-regional and regional scales, such as are currently used to develop national-level budgets for agricultural emissions (Ogle *et al* 2023). There is also an opportunity for a hybrid approach where modeling is used to issue annual payments to farmers while carbon credits are ultimately issued based on remeasurement on a longer timescale (e.g. 5 years). To permit project developers, carbon registries, researchers, and government agencies alike to explore the costs of conducting such measure-and-remeasure studies across commercial fields, we developed an open-source web application based on this study where results may be obtained for project-specific parameters (https://asc.illinois.edu/soc-econ).

Recognizing that not all SOC projects are intended to sell carbon credits, we highlight that our approach also permits entities such as agrifood producers to estimate the cost of 'inset' projects that reduce greenhouse gas emissions in their supply chains. A measure-and-remeasure approach allows such non-market projects to rigorously quantify the SOC benefits of climate-smart agriculture (though other greenhouse gas fluxes such as N_2O will likely need to be modeled). Note that measure-and-model projects leave undetermined who bears responsibility for model errors leading to flawed credits or insets: the project developer, carbon market, credit buyer, and/or society. Measure-and-remeasure shifts more of this risk onto the project itself, building confidence that the impacts of climate-smart agriculture are real. This responsibility suggests that measure-and-remeasure is, at minimum, a necessary complement to verify that measure-and-model approaches can estimate SOC accrual rates at commercial scales.

Lastly, our work reveals shortcomings in our fundamental understanding of SOC dynamics, in particular the validity of extrapolating small plot research to landscape and regional scales. For example, the most influential parameter in our sensitivity analysis was the average treatment effect $\tilde{\tau}$. This parameter has been estimated in scores of studies of different treatments in different locations (Nicoloso and Rice 2021). Yet it is crucial to recognize that it is unknown to what extent these small plot experimental estimates can be extrapolated to large-scale, commercial farms. Thus we do not know the utility of these estimates for SOC markets as well as governmental and academic efforts to understand the mitigation potential of climate smart agriculture (Buma *et al* 2024). Our statistical framework highlights other important and under-studied characteristics of SOC including finescale spatial variability of SOC (σ_n) and between-field and within-field variability in SOC change (σ_b and σ_w). A Bayesian approach to project design (Chaloner and Verdinelli 1995) could be used to account for uncertainty in SOC dynamics. In the SI we take a first step in this direction by allowing for uncertainty in the average treatment effect $\tilde{\tau}$ with a parameter $\sigma_{\tilde{\tau}}$.

In conclusion, our work shows that measure-and-remeasure SOC projects have the potential to rigorously quantify climate benefits and can provide a competitive return on investment if these benefits are realized. Moreover, such projects would serve to validate the predominant measure-and-model approach. We hope that these findings will spur the development of measure-and-remeasure SOC projects.

Data availability statement

All data that support the findings of this study are included in the figshare repository at https://doi.org/10.6084/m9.figshare.28083182.

Acknowledgments

E P and K G were supported by the University of Illinois Agroecosystem Sustainability Center, DOE ARPA-E SMARTFARM program, NSF Environmental Sustainability Program, and FFAR Seeding Solutions. MAB was involved in this work through the Yale Applied Science Synthesis Program, which is an initiative of The Forest School (at the Yale School of the Environment) and the Yale Center for Natural Carbon Capture. EEO was supported by gifts to EDF, and MAB a grant from EDF, arising from King Philanthropies and Arcadia, a charitable fund of Lisbet Rausing and Peter Baldwin. We acknowledge Viridiana Alcantara-Shivapatham, Helaina Black, John Cobb, Ben Dube, Doria Gordon, Jocelyn Lavallee, Andrew Margenot, Dave McLaughlin, Steven Rosenzweig, Jonathan Sanderman, and Asa Watten for their feedback on this work.

Author contributions

EP conceived, designed and carried out the research and drafted the manuscript. MAB, EEO, and KG shaped the design, and reviewed and edited the manuscript.

ORCID iDs

Eric Potash https://orcid.org/0000-0003-4658-9802
Mark A Bradford https://orcid.org/0000-0002-2022-8331
Emily E Oldfield https://orcid.org/0000-0002-6181-1267
Kaiyu Guan https://orcid.org/0000-0002-3499-6382

References

Al-Kaisi M M and Kwaw-Mensah D 2020 Quantifying soil carbon change in a long-term tillage and crop rotation study across Iowa landscapes *Soil Sci. Soc. Am. J.* **84** 182–202

Badgley G, Freeman J, Hamman J J, Haya B, Trugman A T, Anderegg W R L and Cullenward D 2022 Systematic over-crediting in California's forest carbon offsets program *Glob. Change Biol.* **28** 1433–45

Bergtold J S, Ramsey S, Maddy L and Williams J R 2019 A review of economic considerations for cover crops as a conservation practice *Renew. Agric. Food Syst.* **34** 62–76

Bowman M, Poley K and McFarland E 2022 Farmers employ diverse cover crop management strategies to meet soil health goals *Agric. Environ. Lett.* **7** 1–6

Bradford M A, Eash L, Polussa A, Jevon F V, Kuebbing S E, Hammac W A, Rosenzweig S and Oldfield E E 2023 Testing the feasibility of quantifying change in agricultural soil carbon stocks through empirical sampling *Geoderma* **440** 116719

Brus D J 2022 *Spatial Sampling with R* (Chapman and Hall/CRC)

Buma B *et al* 2024 Expert review of the science underlying nature-based climate solutions *Nat. Clim. Change* **14** 402–6

Chaloner K and Verdinelli I 1995 Bayesian experimental design: a review *Stat. Sci.* **10** 273–304

Che Y, Rejesus R M, Cavigelli M A, White K E, Aglasan S, Knight L G, Dell C, Hollinger D and Lane E D 2023 Long-term economic impacts of no-till adoption *Soil Secur.* **13** 100103

de Gruijter J J, McBratney A B, Minasny B, Wheeler I, Malone B P and Stockmann U 2016 Farm-scale soil carbon auditing *Geoderma* **265** 120–30

Ellis E and Paustian K 2024 Importance of on-farm research for validating process-based models of climate-smart agriculture *Carbon Bal. Manage.* **19** 16

Havens D *et al* 2023 Carbon farming: a preliminary economic analysis of carbon credits for no-till and cover crops (Cornhusker Economics) vol 1206 (available at: https://digitalcommons.unl.edu/agecon_cornhusker/1206)

Hernán M A and Robins J M 2020 *Causal Inference: What If* (Chapman & Hall/CRC)

Holland P W 1986 Statistics and causal inference *J. Am. Stat. Assoc.* **81** 945–60

Indigo 2024 Project monitoring report 3 (Indigo U.S. Project No. 1) (available at: https://thereserve2.apx.com/mymodule/reg/TabDocuments.asp?r=111&ad=Prpt&act=update&type=PRO&aProj=pub&tablename=doc&id1=1459) (Accessed 5–9 January–April 2024)

IPCC 2022 *Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* ed Shukla P R *et al* (Cambridge University Press) (https://doi.org/10.1017/9781009157926)

Kravchenko A N, Snapp S S and Robertson G P 2017 Field-scale experiments reveal persistent yield gaps in low-input and organic cropping systems *Proc. Natl. Acad. Sci. U.S.A.* **114** 926–31

Lark R M 2009 Estimating the regional mean status and change of soil properties: two distinct objectives for soil survey *Eur. J. Soil Sci.* **60** 748–56

Mathers C *et al* 2023 Validating DayCent-CR for cropland soil carbon offset reporting at a national scale *Geoderma* **438** 116647

Nicoloso R S and Rice C W 2021 Intensification of no-till agricultural systems: an opportunity for carbon sequestration *Soil Sci. Soc. Am. J.* **85** 1395–409

Ogle M, Breidt F J, Del Grosso S, Gurung R, Marx E, Spencer S, Williams S and Manning D 2023 Counterfactual scenarios reveal historical impact of cropland management on soil organic carbon stocks in the United States *Sci. Rep.* **13** 14564

Oldfield E E *et al* 2024 Greenhouse gas mitigation on croplands: clarifying the debate on knowns, unknowns and risks to move forward with effective management interventions *Carbon Manage.* **15** 2365896

Oldfield E E, Eagle A J, Rubin R L, Rudek J, Sanderman J and Gordon D R 2022 Crediting agricultural soil carbon sequestration *Science* **375** 1222–5

Poeplau C, Prietz R and Don A 2022 Plot-scale variability of organic carbon in temperate agricultural soils—Implications for soil monitoring *J. Plant Nutrition Soil Sci.* **185** 403–16

Potash E *et al* 2023 Multi-site evaluation of stratified and balanced sampling of soil organic carbon stocks in agricultural fields *Geoderma* **438** 116587

R Core Team 2020 *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing) (available at: www.R-project.org/)

Saby N P A *et al* 2008 Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content? *Glob. Change Biol.* **14** 2432–42

Sharpe W F 1966 Mutual fund performance *J. Bus.* **39** 119–38

Smith P 2004 How long before a change in soil organic carbon can be detected? *Glob. Change Biol.* **10** 1878–83

von Haden A C, Yang W H and 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 *Glob. Change Biol.* **26** 3759–70

Wang S, Guan K, Zhang C, Lee D, Margenot A, J, Ge Y, Peng J, Zhou W, Zhou Q and Huang Y 2022 Using soil library hyperspectral reflectance and machine learning to predict soil organic carbon: assessing potential of airborne and spaceborne optical soil sensing *Remote Sens. Environ.* **271** 112914

West T A P, Wunder S, Sills E O, Börner J, Rifai S W, Neidermeier A N, Frey G P and Kontoleon A 2023 Action needed to make carbon offsets from forest conservation work for climate change mitigation *Science* **381** 873–7

Wijewardane N K, Hetrick S, Ackerson J, Morgan C L S and Ge Y 2020 VisNIR integrated multi-sensing penetrometer for *in situ* high-resolution vertical soil sensing *Soil Till. Res.* **199** 104604