Sampling decisions for quantifying soil carbon change: resampling the same locations increases efficiency and precision

Alexander Polussa[a,b*], Emily E. Oldfield[c], Tiffany Runge[d], Elizabeth Forbes[d], Matthew Sheffer[d], Jonathan Sanderman[e], Mark A Bradford[a,b]

[a]*The Forest School, Yale School of the Environment*, *360 Prospect St., Yale University, New Haven, CT 06511, USA*

[b]*Yale Center for Natural Carbon Capture, Yale University, New Haven, CT 06511, USA*

[c]*Environmental Defense Fund, Washington, DC 20004, USA*

[d]*Hudson Carbon Research Foundation Inc, Hudson, NY 12534, USA*

[e]*Woodwell Climate Research Center, 149 Woods Hole Rd., Falmouth, 02540 MA, USA*

**Author for correspondence: alexander.polussa@yale.edu; The Forest School, Yale School of the Environment, 360 Prospect St., Yale University, New Haven, 06511 CT, USA*

Abstract

Quantifying changes in soil carbon commonly relies on sampling designs that either return to new or the same locations in the project area. When the intention is to quantify change caused by adoption of different management practices, these same approaches are used but complemented with control areas that function as counterfactual, business-as-usual conditions. For both monitoring and causal estimation, a key design decision underpins the efficiency and precision of a sampling campaign: whether to 1) resample an independent set of samples or 2) resample the

same point locations (paired resampling). Using repeated measurements of soil carbon concentrations and stocks from 796 locations across 80 agricultural fields distributed across 20 strata, we first evaluate the sampling efficiency and precision benefits of paired resampling within a monitoring context. In monitoring, we find that resampling the same point locations resulted in an ~2.5-times improvement in precision, which translated to gains in sampling efficiency of 6.7- and 5.5-times fewer samples required to detect change in soil carbon concentrations and stocks, respectively. For causal estimation, where more factors generate uncertainty in the change estimate, paired resampling improved precision by 1.75- and 1.95-times and sampling efficiencies by 3.1- and 3.8-times for concentrations and stocks. Our results are consistent with directional expectations from sampling theory that revisiting sampling locations provides gains in precision and efficiency for quantifying change. However, by testing this theory with a large empirical dataset, we reveal that theoretical benefits translate to real-world, practical benefits of paired resampling designs for overcoming noise generated by the large size and high variability of soil carbon contents to detect the signal of soil carbon change. Our results are relevant for informing the efficiency and precision of designs for regional and national programs focused on tracking and improving soil conditions for soil health, food security and climate mitigation.

Keywords: soil security, food security, mitigation, soil health, carbon accounting, soil carbon, equivalent soil mass, measure-and-remeasure.

1. Background

Small changes in carbon stored in soil can have a large impact on the global carbon balance (1). This idea has spurred efforts to manage agricultural soils for carbon storage and led to a variety of programs and protocols that aim to quantify and verify the extent to which adoption of

47 sustainable management practices increase soil carbon and improve soil health . These protocols
48 often rely on inventory and monitoring sampling designs that were developed to sample soils for
49 the purpose of assessing current conditions (i.e. inventory) and how these conditions are
50 changing (i.e. monitoring), but not for quantifying management effects on those changes (i.e.
51 causal estimation). The first two of these purposes – current conditions and change – are
52 descriptive. Their measurement has motivated the science of natural resource monitoring,
53 resulting in an extensive body of sampling and statistical theory. This theory has been used to
54 design national and continental-scale monitoring networks to track changes in soil condition
55 given its importance for sustaining ecosystem services and increasingly because of soil carbon's
56 role as a potential source or sink of atmospheric CO_2 (5,6). For the latter purpose – quantifying
57 management effects – the same theory and practice that informs soil monitoring has been applied
58 to applications ranging from IPCC national-level greenhouse gas accounting guidelines (7,8) to
59 those used in the voluntary carbon market (3,4). However, monitoring change versus attributing
60 change to management are different goals, demanding assessment of how natural resource
61 monitoring theory translates into practical guidelines for quantifying soil carbon change versus
62 the component of that change attributable to management decisions.
63 Within descriptive sampling, an efficient approach is one with the fewest samples for a
64 target threshold of uncertainty (9). For instance, two approaches to monitoring soil carbon
65 change can be described as (1) taking two inventories at two time points ('independent
66 resampling') and (2) taking one inventory and then revisiting the same locations for resampling
67 ('paired resampling'). An inventory estimates the current stock of a region by randomly placing
68 points across it and estimating the mean (9). Repeating this process (as in 1) to provide a
69 spatially representative estimate at each time point is optimal if the goal is to capture additional

spatial variation when resampling, thereby improving confidence in the inventory estimate. However, designs that improve inventories are commonly inefficient for estimating change. Revisiting the same point locations (as in 2) should more efficiently measure change because the influence of the local-scale spatial variation in soil carbon contents is reduced when the same location is remeasured, isolating temporal change more precisely (10,11). These two approaches to monitoring have different benefits: the first, 'independent resampling', brings in additional spatial information at the subsequent time points at the cost of lower efficiency for estimating change; the second, 'paired resampling', estimates change across the system more precisely by reducing spatial variation at the cost of not updating spatial coverage at the subsequent time points. Both approaches are included within and among accounting guidelines for sampling soil carbon, but there have been few empirical evaluations of how much more efficient the latter approach, 'paired resampling', is over 'independent resampling' in terms of real-world practical benefits for sampling efficiencies and precision.

The practical benefits of paired resampling for causal estimation of management effects on soil carbon change has received far less attention than monitoring, despite the huge demand for quantification and independent validation of how adoption of sustainable practices benefit soil health and result in carbon removals. For causal estimation, the focal outcome is whether change differs between management units that did and did not receive an intervention. The relevant spatial unit in this causal context is the area at which management interventions are assigned, changing the unit from the resampling location to, for example, an agricultural field where cover crops have been adopted. The causal effect of change in soil carbon is then the difference between outcomes across many units which have and have not received the management intervention. With this shift in relevant spatial unit, sampling within each

93 management unit is used to better, or more precisely, estimate its change. The shift in sampling
94 unit raises an important question about what the tradeoff is between sampling more point
95 locations within a management unit (e.g. an agricultural field) versus fewer within-field locations
96 and more management units. This tradeoff has not been thoroughly evaluated with empirical data
97 but theoretically (see 12) should have important cost and logistical consequences for project
98 design to achieve sampling efficiency and precision goals when quantifying the effectiveness of
99 management interventions for building soil carbon stocks.

100 Here, we make available a recent dataset on soil carbon concentrations and stocks that
101 comprises 796 sampling locations across 20 soil strata and 80 agricultural fields, representing a
102 combined area of 498 ha within a region of ~600 km^2. We use these data to assess precision and
103 sampling efficiency benefits of 'paired resampling' relative to 'independent resampling' designs
104 in the context of monitoring soil carbon change. We next query the extent to which these benefits
105 translate to a causal estimation framework, where sampling locations are nested within field
106 boundaries (i.e. management units). We focus on two soil carbon metrics: carbon concentration
107 in the topsoil (0–15 cm), which is relevant to soil health and crop yield resilience (13,14), and
108 carbon stocks to 30 cm on an equivalent soil mass basis, which are relevant to climate mitigation
109 and carbon accounting (8). Our overarching goal is to provide empirical data and estimates that
110 can be used to inform and advance the practical design of guidelines and protocols that translate
111 to efficient and precise estimation of both soil carbon change and the extent to which change is
112 due to management.

2. Methods

2.1. Study Data

The dataset is from the Hudson Valley in New York State, USA (42°N 73°W) and consists of 80 agricultural fields within an area of ~600 km^2. Fields were selected from a population of 394 fields identified within the study region, restricted to 223 fields of intermediate size (3–14 ha), and stratified into 20 strata based on soil organic carbon (SOC) and sand content estimated from gSSURGO map units (15). Within each of the 20 strata, up to four fields were selected under each of two management systems, regenerative and conventional, giving 40 fields per system. Within a causal framework, we use the comparison between these two management groups as an example for evaluating the benefit of revisiting sampling locations, however, interpreting the difference between groups as an effect of management here would require additional assumptions (16). Stratified field selection was intended to improve comparability between groups, but other assumptions, such as parallel trends, could not be tested with these data and were not the focus of this study. Fields were sampled at an average density of ~1.6 locations ha^{-1}. Further information on the farms and subregion are provided in Polussa et al. (16).

Baseline samples were collected during winter 2023–2024 and follow-up samples during winter 2024–2025. Sampling locations were relocated using GPS coordinates, allowing repeated measurements at the same locations, within 1-2 m of the baseline soil core (GNSS in single-receiver mode; Reach RS2, Emlid Ltd., Budapest, Hungary). The final dataset consisted of 796 sampling locations. Carbon concentration was evaluated for 0–15 cm, the depth most commonly used in soil health assessment, at all 796 locations. Carbon stocks were evaluated for 0–30 cm, the standard depth for soil carbon stock accounting, at the 756 locations with complete profiles, using the equivalent soil mass (ESM) approach (17, 18).

2.2. Soil Sampling and processing

Soil sampling was performed by collecting a single intact soil core to a target depth of 60 cm using a truck-mounted coring system, or a slide hammer where vehicular access was constrained. Cores were 4 cm dia., except for five baseline samples that were collected using a 7-cm dia. corer. Samples were capped to minimize moisture loss and stored under refrigerated conditions prior to laboratory processing. Each core was divided into three depth increments (0–15, 15–30, and 30–60 cm), air-dried, and passed through a 2 mm sieve. The mass of the <2 mm fine-earth fraction was recorded and analyzed for total carbon concentration (%C) by dry combustion (vario MAX cube, Elementar Analysensysteme GmbH, Langenselbold, Germany). We report total carbon concentration (%) from the 0–15 cm depth interval for relevance in soil health assessment (13). Soil carbon stocks for the 0-30 cm profile were calculated using ESM to account for re-sampling differences in soil mass and coarse fragment content.

To calculate stocks using ESM, for each sampling location, the cumulative soil mass of the 0-30 cm section at the initial sampling (time 0) was used as the reference soil mass. At remeasurement (time 1), cumulative soil mass and cumulative soil carbon stocks were calculated through the profile. Soil carbon stocks corresponding to the reference soil mass were then estimated using a cubic spline (Hyman) interpolation of the cumulative %C and soil mass relationship. The value represents the amount of SOC contained within the same mass of mineral soil from the initial sampling event and avoids bias from differences in bulk density or soil volume collected over time (17–19).

158 **2.3. Temporal correlation**

159 Using the entire dataset we first calculated the temporal correlation between the baseline
160 measurements and follow-up for %C and soil carbon stocks. Temporal correlation provides a
161 measure of the covariance between repeated measurements and is directly related to the expected
162 gains in precision from revisiting sampling locations. Specifically, the variance of estimated
163 change is:

164 $$Var(T1 - T0) = Var(T1) + Var(T0) - 2 \times Cov(T0, T1) \quad (1)$$

165 Or equivalently:

166 $$Var(T1 - T0) = \sigma_{T1}^2 + \sigma_{T0}^2 - 2 \times \rho \times \sigma_{T1} \times \sigma_{T0} \quad (1a)$$

167 The variance of change equals the sum of the variances at each time point minus twice their
168 covariance where T0 and T1 are baseline and follow-up soil carbon measurements (%C or
169 stocks), σ is the standard deviation for each time point, and ρ is the correlation between the two
170 time points. As covariance between repeated measurements increases, the variance of estimated
171 change decreases, improving the precision of change estimates (9). Consequently, the benefits of
172 revisiting sampling locations are expected to increase with temporal correlation. Our analyses
173 treat sampling locations within a stratum or field as independent. We tested for spatial
174 autocorrelation among locations within fields, which were at least 49 m apart (Supplementary
175 Section S2.3). For both stocks and concentration, baseline values showed weak autocorrelation,
176 with a correlation of about 0.1 between the closest locations and change showed no spatial
177 autocorrelation (Table S4, Figure S2).

2.4. Monte-Carlo resampling simulation for monitoring

To evaluate the influence of resampling design on the precision of soil carbon change estimates, we conducted a Monte-Carlo simulation using the repeat measurements from the same locations. Analyses were performed separately for soil %C and carbon stocks.

Within each stratum, a baseline sample (T0) was randomly drawn. The number of samples drawn was approximately half of the total number of available locations within each stratum to allow for direct comparison between a fully independent resampling to a fully paired resampling. The follow-up sample, T1, varies the number of point locations revisited depending on the design (Fig. 1). The fraction of revisited locations (p) ranged from 0 to 1 in increments of 0.1. When (p=0), all T1 observations were drawn from locations not sampled at T0, representing complete independent resampling. When (p=1), all locations were revisited, representing complete paired remeasurement. Intermediate values of (p) represent rotational sampling designs in which a fraction of locations were remeasured and the remainder were replaced with newly sampled locations. Within a single iteration, one set of follow-up (T1) draws was generated and the proportion of revisited locations was then increased incrementally from that same set, so that designs at different values of p are directly comparable within the iteration.

For each resampling fraction, the mean change between sampling years was estimated using linear mixed-effects models:

$$C_{ps} \sim \beta_0 + \beta_1 \times year_{ps} + b_s + c_{ps} + \varepsilon_{ps} \quad (2)$$

Where C_{ps} is soil carbon (concentration or stocks) at point p within stratum s, β_0 is the intercept, β_1 is the coefficient for the year change from T0 to T1, b_s is the stratum specific random effects, c_{ps} is the point level random effect within each stratum. Stratum was included as a random effect in all models. For designs containing remeasured locations (p>0), point identity

was included as a nested random effect to account for repeated measurements at the same location. The fixed effect of year represented the estimated change between sampling campaigns, and the associated standard error (SE) and variance (SE^2) was used as the primary measure of statistical precision. The mixed-effects model produces comparable results to conventional t-tests at the endpoints for paired (p=1) and independent (p=0) resampling and allows for intermediate rotational sampling designs to be evaluated. A model-based approach for assessing the uncertainty of an estimate was primarily used as a flexible approach to use variance estimates across remeasured points and strata. We complement this using design-based estimators for the two endpoint scenarios of 100% independent sampling and 100% paired location resampling (Supplementary Section S2.1, Table S2). The design-based estimates provide similar estimates for variance and subsequent conclusions.

Monte-Carlo simulations were repeated 200 times to capture the range of estimates across our dataset. For every resampling fraction, we calculated the mean estimated change, standard error and variance (SE and SE^2), relative efficiency, and minimum detectable difference (MDD). Relative efficiency was calculated as the ratio of the variance under complete independent resampling (p=0) to the variance of the evaluated design ($SE_{p=0}^2$/ SE_p^2). Minimum detectable differences were calculated using the equation:

$$MDD = \left(z_{1-\alpha/2} + z_{power}\right) \times SE\left(\Delta_{(\%C\ or\ stock)}\right) \quad (3)$$

Where $z_{1-\alpha/2}$ is the critical value of a two-sided test at significance level α, z_{power} is the desired power (here, 0.80), and SE is the standard error of estimated mean change (%C or stocks) estimated from the linear mixed-effects model.

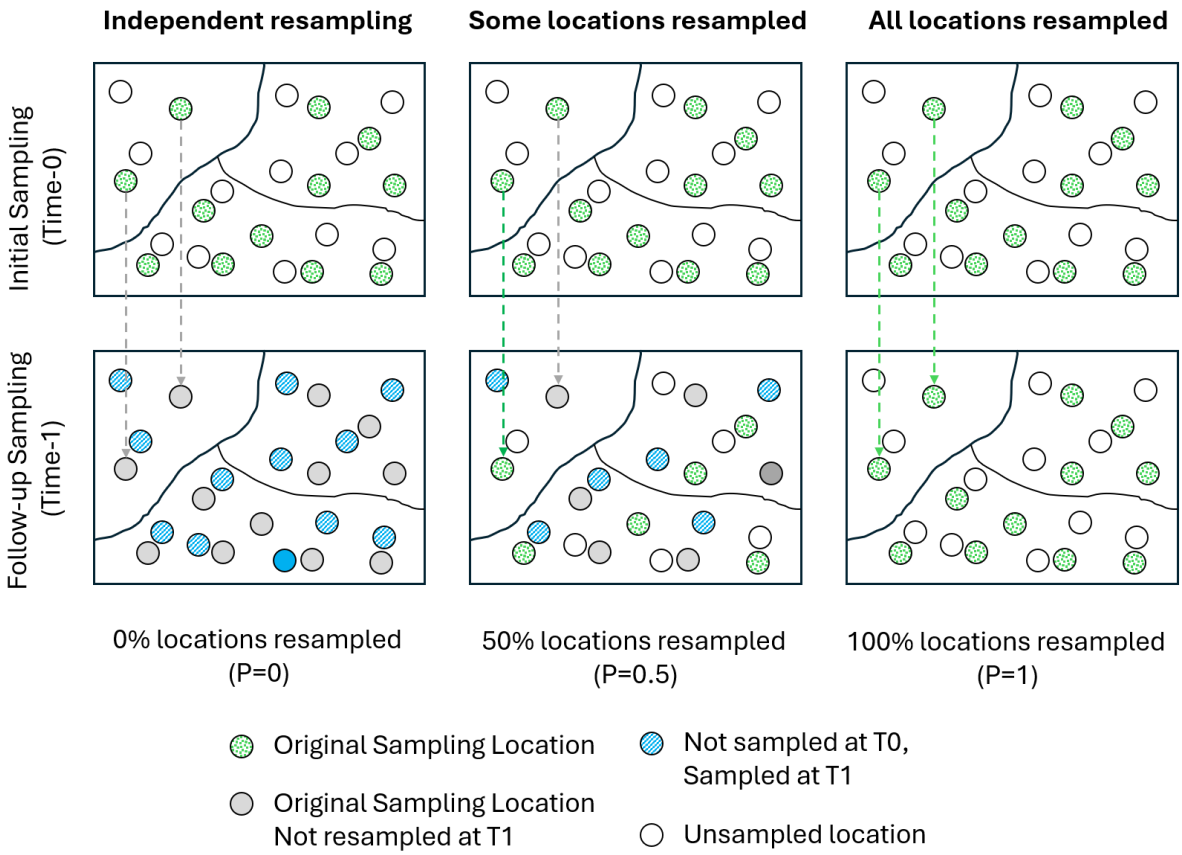


Figure 1. *The continuum of resampling strategies evaluated in this paper from complete independent resampling (0% of initial locations resampled, left panels) to fully paired resampling (100% of initial locations resampled, right panels), with intermediate rotational designs representing mixtures of revisited and newly sampled locations (center panels) within strata. Strata are represented by curved lines that represent the broader system. The original sample locations are highlighted in green (dot fill) in the top row and stay green if resampled in the follow-up sampling (bottom panels). Locations that are sampled only in the follow-up are highlighted in blue (diagonal fill) in the bottom panels. The proportion of revisited locations determines the extent to which temporal covariance between measurements can be exploited when estimating soil carbon change.*

2.5. Causal analysis with paired and unpaired estimators

Within a causal context, to estimate the benefit of returning to the same locations, we use the two populations of individual fields from this study which practice regenerative or conventional management (16). The estimate of interest within a causal framework is the difference in soil carbon change between these two management systems. Individual fields contained too few sampling locations (5-23) to apply the Monte-Carlo approach described above, which required leaving half of the locations unsampled at baseline. We instead evaluate the pairing benefit using two linear mixed-effects models: one that uses a point-level random intercept to capture revisiting locations (i.e. paired re-sampling) and one that does not include this term, essentially permitting us to evaluate independent resampling by omitting sampling location-level correlation between time points. We highlight that in this evaluation we use the entire dataset and thus do not recreate a fully independent design.

The paired model is:

$$C_{pfs} \sim \beta_0 + \beta_1 \times year_{pfs} + \beta_2 \times Management_{fs} + \beta_3 \times (year \times management)_{pfs} + b_s + c_{fs} \left(+ p_{pfs}\right) + \varepsilon_{pfs} \quad (4)$$

Where C_{pfs} is soil carbon at point p in field f within stratum s, β_3 is the management effect on change, b_s and c_{fs} are stratum and field random intercepts, and p_{pfs} is the point-level random intercept only included in the paired model. We use the standard error of the management-by-year term, β_3, to evaluate precision between the paired and unpaired models. We complement this model-based result with an analytic variance decomposition that estimates change for each field and retains variation in change among fields (20; Supplementary Section S2.2).

3. Results

3.1. Temporal correlation

Temporal correlations between baseline and follow-up measurements for %C and soil carbon stocks were high, at 0.88 and 0.85, respectively (Figure 2). Such high correlations arise, at least in part, through persistence of spatial patterns across the sampling area. Within each stratum, the correlations ranged from 0.56 to 0.98 for %C and 0.39-0.95 for stocks (Table S1). These correlations suggest that resampling locations will improve precision of estimated change, and we evaluate below the extent of this improvement for monitoring and causal estimation.

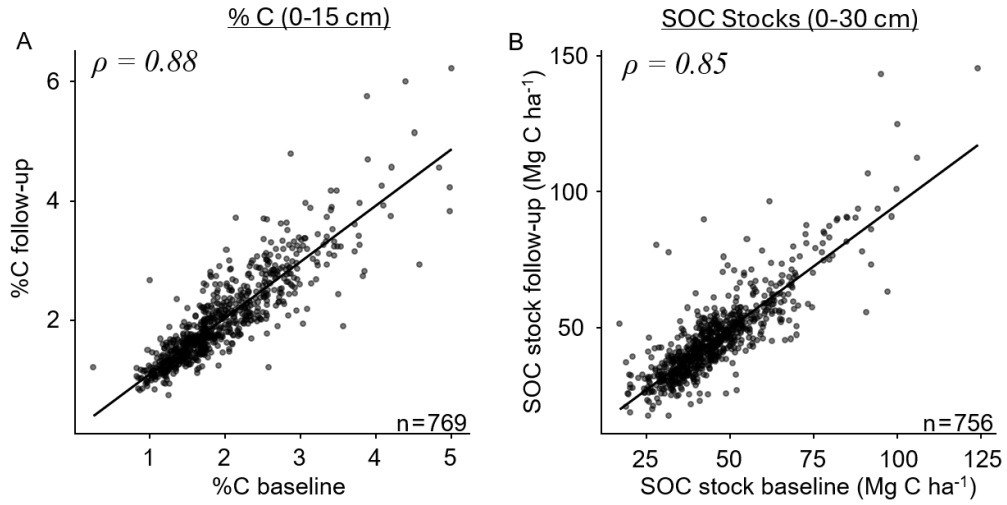


Figure 2. *Correlation of baseline and follow-up for the soil carbon concentrations (0-15cm) and stocks (0-30 cm). Baseline and follow-up measurements for (A) soil carbon concentration (%C) and (B) equivalent soil mass soil carbon stocks (0-30 cm). Pearson correlation coefficients (ρ) quantify temporal dependence between observations and provide the empirical basis for evaluating the efficiency gains associated with paired remeasurement relative to independent resampling designs; n is number of paired locations.*

3.2. Monte-Carlo resampling simulation for monitoring

For each Monte-Carlo simulation, approx. half of the locations were used (i.e. 394 for %C and 374 for stocks) for estimating uncertainty and change across the designs. Increasing the

proportion of resampled locations substantially reduced the standard error and overall uncertainty of the estimated change. The average standard error in %C when independently sampled between campaigns (proportion revisited=0) was 0.047 compared to an average mean change of 0.048, meaning that sampling noise obscured our temporal signal. When resampling the same locations (proportion revisited=1), standard error decreased to 0.018, representing a 2.6-times improvement in precision (Figure 3) that suggests that relatively small changes in carbon concentrations are detectable with a resampling design.

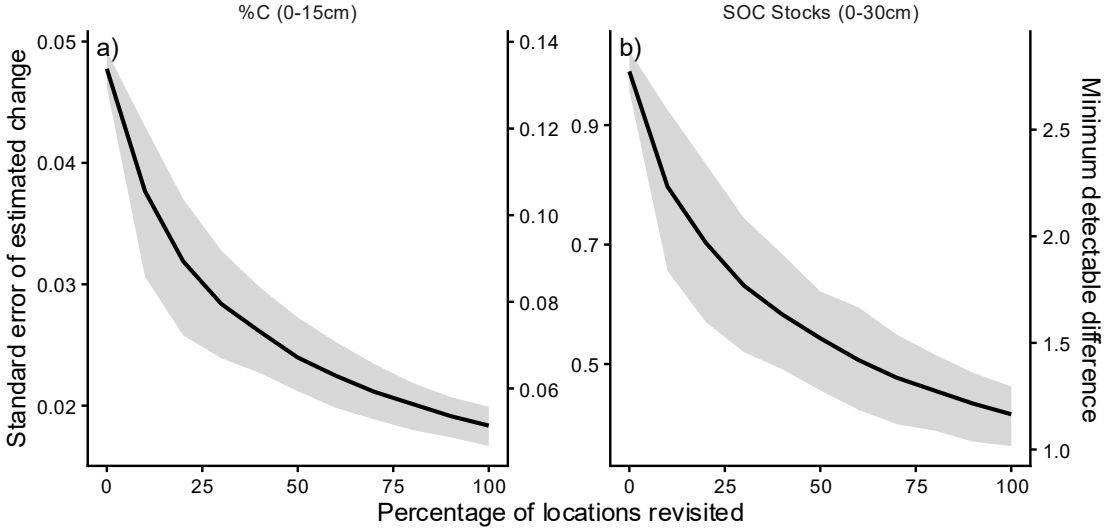


Figure 3. *Simulated independent to paired resampling on uncertainty in estimated soil carbon change. The black line shows the average standard error across 200 Monte-Carlo simulations and the grey band indicates the central 95% of simulation results. Simulations ranged from complete independent resampling (0% revisited locations) to complete paired remeasurement (100% revisited locations). Increasing the proportion of revisited locations consistently reduced uncertainty for both soil carbon concentration (%C; panel a) and equivalent soil mass SOC stocks (panel b), resulting in lower minimum detectable differences (secondary y-axis).*

Similarly, for soil carbon stocks, standard error decreased from 0.98 to 0.42 Mg C ha^{-1}, corresponding to a 2.3-times reduction against a mean estimated change of ~0.4 Mg C ha^{-1}. Although precision of the temporal signal matched sampling noise (i.e. standard error) with the

294 paired design, we note that the mean estimated change is for a single year, whereas gains in
295 precision from paired resampling would be expected to hold across subsequent years as
296 estimated SOC stock change, if it holds directionally, increases in magnitude.
297 Gains in precision were also observed for rotational designs. For example, by resampling
298 50% of same locations the standard error was 2-times smaller for %C and 1.8-times smaller for
299 soil carbon stocks relative to complete independent resampling (Figure 3). The largest gains in
300 precision came from the first few locations revisited. That is, precision gains diminished as a
301 greater proportion of locations were revisited: revisiting 0% versus 10% of initial locations
302 decreased standard error by ~21%, whereas revisiting 90% versus 100% of locations decreased
303 standard error by 4% (Figure 3).
304 To consider how reductions in standard error due to pairing relate to the ability to detect
305 signal from noise, and how that relates to sampling effort, we evaluated how the reductions in
306 standard error affected both minimum detectable difference (MDD) and relative sampling
307 efficiencies. For independent versus fully paired resampling, the MDD decreased from 0.13 to
308 0.051 for %C and from 2.8 to 1.2 Mg C ha^{-1} for stocks (Figure 3). Relatedly, the number of
309 samples required to detect a given change decreased by 6.7-times for %C and 5.5-times for SOC
310 stocks (Figure 4). To illustrate the effect on sampling requirements, we estimated the number of
311 locations needed to detect a change of 0.2 %C or 3 Mg C ha^{-1}. For stocks, this corresponds to a
312 rate of ~0.3 Mg C ha^{-1} yr^{-1} over 10 years, typical of practices such as cover cropping in arable
313 systems, or ~0.6 Mg C ha^{-1} yr^{-1} over 5 years, typical of conversion from arable cropping to
314 pasture (21,22). Smaller changes require more locations, but the relative benefit of revisiting
315 locations is fixed. Detecting a 0.2 %C change required 25 locations under a fully paired design,
316 compared with 169 under independent resampling (i.e. 0% revisited locations). Reductions in

sampling requirements were also observed for rotational designs, where moving from 0% to 50% revisited locations reduced the number of locations by ~4-times, and from 50% to 100% revisited by ~1.8-times. The benefit of reduced sampling requirements was larger for %C than stocks (Figure 4), likely because of a higher correlation between the baseline and follow-up samples for %C, whereas stocks are more variable because they include variance from not just %C measurements, but also measurements of bulk density and coarse fraction (i.e. rocks and roots >2 mm).

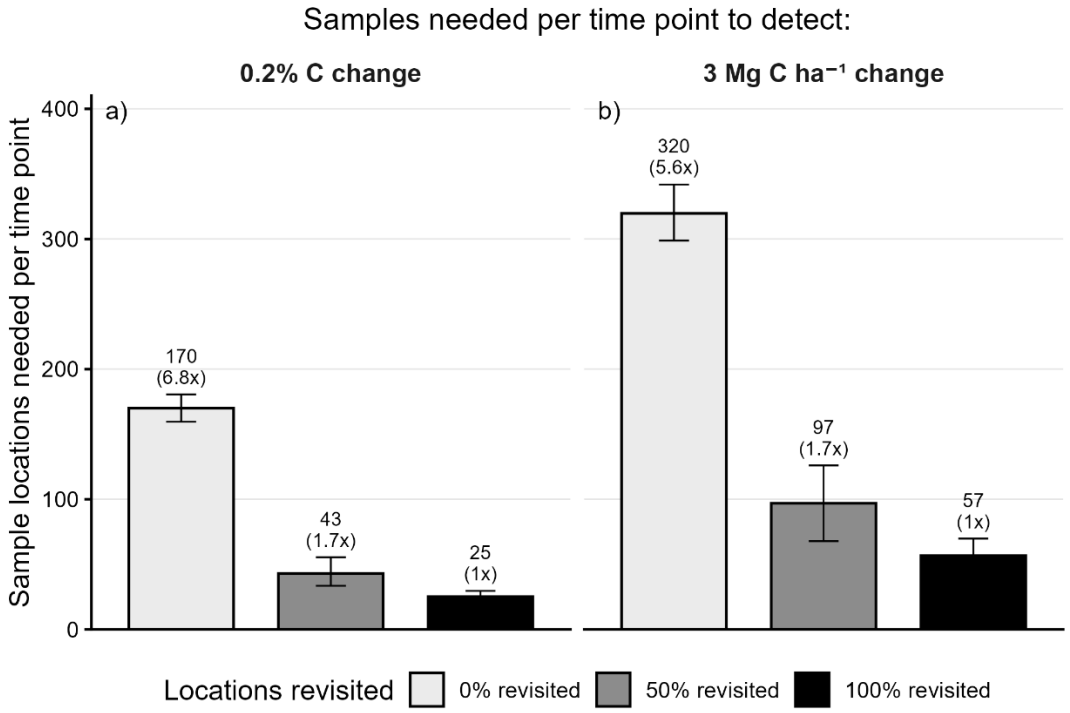


Figure 4. The benefits of revisiting sampling locations. *The relative benefit of revisiting the same locations is fixed across any target effect size. Here, we illustrate the number of sampling locations required per sampling time point to detect a 0.2 %C change or a 3 Mg C ha^{-1} change in soil carbon stocks under independent resampling (0% revisited), rotational sampling (50% revisited), and complete paired remeasurement (100% revisited). The average standard error from the Monte-Carlo simulations was used to calculate the number of sample locations for each time point required with 80% power at α = 0.05. Numbers in parentheses indicate the additional sampling effort required relative to complete paired remeasurement. Across both metrics, independent resampling required more than five times more sampling*

locations than revisiting existing locations, illustrating the large efficiency gains obtained when temporal covariance is incorporated into soil carbon monitoring designs.

3.3. Causal analysis with paired and unpaired estimators

Within the causal framework, we used all revisited locations to estimate the management effect and its uncertainty under a paired versus an unpaired estimator fit to the same physical locations. These locations fall within 80 fields, split between the two management systems (40 regenerative, 40 conventional). The causal framework shifts the objective of estimation from monitoring, which relies on estimating change for the individual point locations across the landscape, to estimating change for each field across the landscape. Specifically, causal estimation uses the multiple samples within a field to estimate more precisely the field-level change. As a result, the effective sample size is now 80 fields and hence the value of pairing versus unpaired relates to estimates of the 80 field-level changes.

The paired estimator reduced the standard error of the management effect by 1.75-times for %C and 1.95-times for stocks which results in relative efficiency gains of 3.1 and 3.8-times, respectively (Figure 5a). An analytic variance decomposition gave smaller but similar gains of 2.6-times for %C and 2.9-times for stocks (Table S3). Because the number of fields required to reach a given precision scales with this variance, an unpaired estimator would need 3.1- or 3.8-times as many fields to detect the same management effect for concentrations and stocks (Figure 5b,c). Similarly, if the number of fields is kept constant, pairing reduces the samples needed per field to reach a given precision on a field's change by about three times for %C and four times for stocks.

We note that our causal comparison retains all resampled locations, as opposed to spatially independent ones as in the monitoring comparison, and it is only through the statistical estimator

that we isolate the value of the pairing. By differencing within each field over time, this approach removes the stable, persistent differences among fields, so its precision depends on the within-field temporal correlation (ρ = 0.70 for %C, 0.77 for stocks).

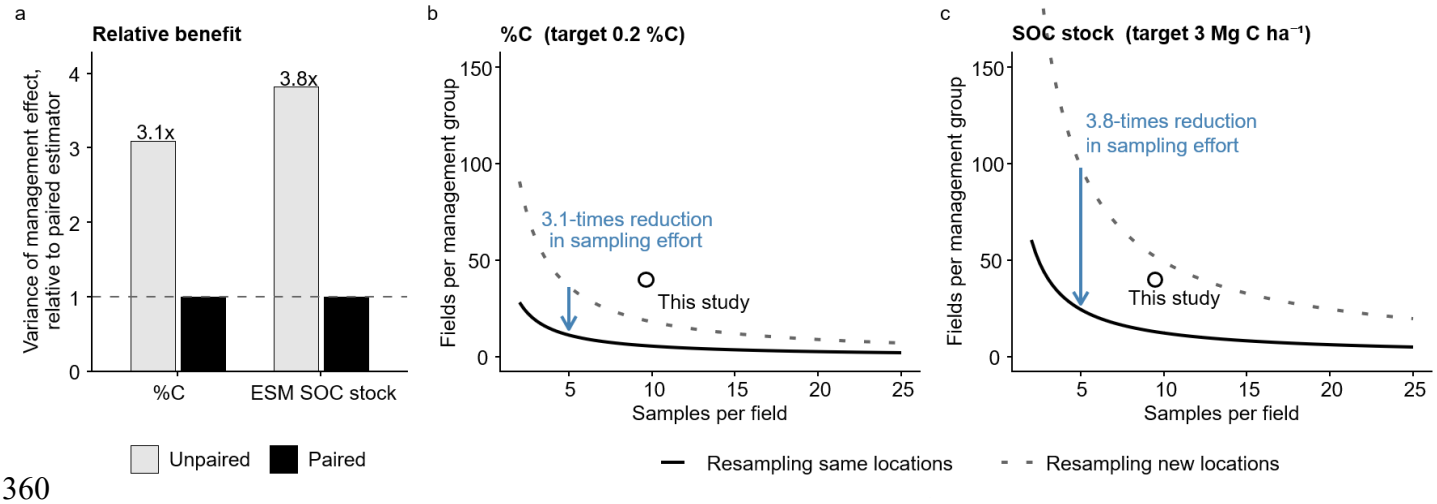


Figure 5. *Efficiency of paired versus unpaired estimation of the management effect. Panel a illustrates the variance of the management effect for an unpaired estimator relative to the paired estimator. Panels b and c show the fields per management group required to detect a 0.2 %C (b) or 3 Mg C ha⁻¹ (c) management difference at 80% power (α = 0.05), as a function of samples per field, under paired (solid) and unpaired (dotted) designs. Curves are power projections from the within-field variance components, assuming between-field variation in field-level change is negligible. At ~10 samples per field this implies roughly 6 vs 18 fields per group for %C and 12 vs 49 for stocks (paired vs unpaired).*

4. Discussion

The value of resampling point locations for estimating soil carbon change is well established in theory. When baseline and follow-up measurements are correlated, revisiting the same locations reduces the influence of spatial variation that could overwhelm the change estimate, and decreases the uncertainty in the estimate of change (9–11,23,24). Lark (11) describes the estimation of soil status (i.e. inventory) and the estimation of change (i.e. monitoring) as distinct objectives that call for different sampling designs. We add a third and equally distinct objective,

the estimation of a management effect, which also calls for distinct sampling designs. For both monitoring and causal estimation objectives, theoretical benefits of revisiting the same locations are seldom evaluated with empirical datasets suited to assessing the practical benefits of these designs under large scale conditions of land use and management change. We make available an empirical dataset at this scale and find that revisiting locations reduced the standard error of mean change by 2.3-2.6 times, translating to about 6-times reduction in the samples needed in a monitoring framework to detect change in soil carbon. Under the causal design comparing the soil carbon change between two groups of fields, we find that revisiting locations better estimates individual field-level changes which contributes to reducing the number of samples per field or number of fields needed by 3.1 for %C or 3.8-times for stocks.

Whereas the effect of returning to the same or nearby location in the context of soil carbon monitoring has been evaluated a few times, it has rarely been evaluated using spatially explicit sampling locations within a population of fields in a growing region. This scale is valuable to build evidence for the effect of land management practices on soil carbon within specific geographies. Brus and de Gruijter (10) showed by simulation that the efficiency of revisiting locations relative to independent resampling increases with the persistence of spatial patterns, and that under high temporal correlation, independent resampling is the least efficient design for estimating change. More recently, von Haden et al. (25) showed that relocating follow-up samples within 10 m of the originals reduced sampling error in estimated change, but their simulations used an interpolated surface that omits short-range variation and so called for direct empirical evaluations. In one of the few empirical evaluations, Yanai et al. (26) reassessed a study that measured forest floor organic matter change across 30 stands remeasured after 10 years. They used paired and unpaired statistical estimators rather than truly independent

399 resampling, but similarly found that pairing at the stand level reduced the number of stands
400 needed to detect change by two to three times.
401 We build on such work by revisiting sampling locations within 1 to 2 m from the original
402 sampling point, allowing us to test benefits of resampling approximately the same point
403 locations. Our findings reveal a precision increase of 2.6-times for estimating %C change and
404 2.3-times for stock-change which translates into needing 6.7 and 5.5-times fewer samples. This
405 reduction illustrates the large benefit of revisiting locations rather than sampling new ones for
406 change estimation. Precision gains diminished as the number of first revisits increased, so
407 revisiting about half of the locations captured the bulk of the benefit (Figures 3, 4). Our approach
408 facilitates evaluation of trade-offs among independent, rotational, and paired sampling for
409 different sampling objectives. However, our two campaigns were separated by one year and
410 designed to evaluate precision and sampling efficiency rather than to establish sustained change
411 in soil carbon. As such, empirical data for repeated measurements across more than two
412 sampling dates are needed to evaluate the efficiency gains of different sampling designs, such as
413 rotational versus paired, when temporal changes in soil carbon follow a sustained trajectory or
414 simply sampling variation (20).
415 The benefit of pairing remeasurement locations under monitoring designs does not
416 necessarily translate to causal estimation because the spatial unit of inference changes from point
417 locations to the individual management units. In our study, these management units were
418 agricultural fields, which then reduced our units of inference from 796 point (756 for stocks)
419 locations to 80 management units, introducing an additional question of whether to allocate more
420 sampling points within versus among fields to reap the greatest benefits in sampling efficiency
421 and precision. From a within-field perspective, benefits should accrue because pairing should

improve the precision of change estimates for individual management units, whereas from a between-field perspective benefits should primarily accrue from greater replication of management units. We find that pairing improves precision from both: within fields, pairing reduced the samples needed to estimate a single field's change by about three times for %C and four times for stocks, regardless of how many samples are taken per field. Between fields, pairing reduced the standard error of the management effect by 1.75-times for %C and 1.95-times for stocks, so that for a given number of samples per field, an unpaired design would need three to four times as many fields to detect the same effect (Figure 5). At about 10 samples per field, detecting a difference of 0.2 %C required roughly 6 fields per management group with pairing compared to 18 without, and detecting a difference of 3 Mg C ha^{-1} required 12 compared to 49.

These reduced sampling efforts arise because soil carbon differs among locations within fields, and these differences are relatively stable over time (within-field correlations of 0.70 and 0.77). These reduced sampling efforts for paired design arise because soil carbon differs substantially among fields, and these differences are relatively stable over time. Because management is assigned at the field-level, replication across fields is necessary to attribute changes in soil carbon to management. Since precision increases both with within-field sampling and more fields, allocating sampling effort to more fields would be a more effective use of resources because more fields selected can better represent the broader population of fields in the growing region and provide management effects that are more generalizable to that population. Unlike the monitoring comparison, our dataset was not large enough to 'hold out' locations, so the causal comparison applies paired and unpaired estimators to the same revisited locations, as in Yanai et al. (26). Our cores were relocated within 1 to 2 m; less precise relocation would be

445 expected to reduce the benefits of paired sampling (25), whereas compositing several cores
446 around each location could increase the benefit by reducing short-range variation (27). In this
447 case, empirical datasets at a similar scale can further reveal how compositing samples will
448 quantitatively influence sampling requirements.

449 Decisions about the number of fields, samples per field, and whether to revisit sampling
450 locations affect both the cost and precision of soil carbon accounting and soil health assessment.
451 Large empirical datasets, like the one we make available here, provide estimates of spatial and
452 temporal soil variability that can be used to determine the number of fields and samples needed
453 for monitoring programs and causal estimation to achieve a desired level of precision, rather than
454 relying on conservative assumptions (20, 28). Because these sampling design decisions directly
455 affect sampling costs, such estimates can help allocate available resources most effectively (12,
456 29, 30). Our results show that sampling new locations, or analyzing revisited locations as if they
457 were independent, can overstate the sampling effort required to achieve a given precision. We
458 highlight that the benefits we quantify are specific to the fields and soils we evaluated. To
459 establish whether the benefits we quantify are generalizable, datasets for other regions and
460 farming practices must also be collected and made available to increase the standards of evidence
461 used for carbon accounting and soil health assessment. We further highlight that gains in
462 precision from paired resampling should not be conflated with confidence in the accuracy of
463 estimated changes or management effects. In causal estimation, identifying comparable fields for
464 counterfactual comparisons is critical (16) for accuracy. As such, studies of management
465 interventions could use efficiency gains from pairing to extend coverage across environmental
466 and management conditions and to include suitable counterfactual fields across those conditions.

Conclusion

Both soil carbon monitoring and the evaluation of management effects rest on robustly quantifying change through time. Our findings suggest that both endeavors stand to gain from paired resampling in terms of sampling efficiencies and precision. To date, these benefits have been largely argued from theory and we extend such work by applying it to estimate the size of the benefits empirically. We could do so because we collected soil carbon data resampled at the same point locations across many working agricultural fields and farms within one region. Using these data, we show that the benefits of pairing are large for monitoring and smaller but still substantial for estimating management effects. Generalizing these benefits of resampling to other growing conditions demands that more resampling datasets are collected at operational scales and made available. As we show, such datasets permit sampling theory to be tested, quantified, and translated into design guidance for more efficient and precise soil health and carbon accounting.

Author contributions

AP and MAB conceived the study. TR, EF, and MS contributed to data collection. AP conducted the analyses and wrote the first draft of the manuscript. All authors contributed to the interpretation of results and to reviewing and editing the manuscript, and read and approved the final manuscript.

Funding

This work was conducted as a project of 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. Funding support was from the Environmental Defense Fund, through an

award from King Philanthropies, M.A.N. Baker and G.L. Baker, Jr., and the Yale Center for Natural Carbon Capture to the Yale Applied Science Synthesis Program.

Acknowledgements

The authors thank the 13 farmers and their families who permitted their fields to be sampled as part of this study.

Availability of data and materials

The dataset supporting the conclusions of this article is available in the Yale Dataverse repository (https://dataverse.yale.edu/) and can be accessed at https://doi.org/10.60600/YU/BWCFEL .

References

1. Minasny B, Malone BP, McBratney AB, Angers DA, Arrouays D, Chambers A, et al. Soil carbon 4 per mille. Geoderma. 2017 Apr;292:59–86. doi:10.1016/j.geoderma.2017.01.002

2. Smith P. Monitoring and verification of soil carbon changes under Article 3.4 of the Kyoto Protocol. Soil Use Manag. 2004 Jun;20(2):264–70. doi:10.1111/j.1475-2743.2004.tb00367.x

3. Verra. VM0042 Improved Agricultural Land Management [Methodology] [Internet]. Verra; 2025 Oct. Available from: https://verra.org/methodologies/vm0042-improved-agricultural-land-management-v2-2/

4. Climate Action Reserve. U.S. Soil Enrichment Protocol [Protocol] [Internet]. Climate Action Reserve; 2022 May. Available from: https://climateactionreserve.org/how/protocols/ncs/soil-enrichment/

5. Smith P, Soussana J, Angers D, Schipper L, Chenu C, Rasse DP, et al. How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. Glob Change Biol. 2020 Jan;26(1):219–41. doi:10.1111/gcb.14815

6. Spencer S, Ogle SM, Breidt FJ, Goebel JJ, Paustian K. Designing a national soil carbon monitoring network to support climate change policy: a case example for US agricultural

lands. Greenh Gas Meas Manag. 2011 Dec;1(3–4):167–78. doi:10.1080/20430779.2011.637696

7. IPCC. 2006 IPCC Guidelines for National Greenhouse Gas Inventories [Internet]. Vol. 4. Institute for Global Environmental Strategies; 2006. (Agriculture, Forestry and Other Land Use). Available from: https://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html

8. IPCC. 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories [Internet]. Vol. 4. IPCC; 2019. (Agriculture, Forestry and Other Land Use). Available from: https://www.ipcc-nggip.iges.or.jp/public/2019rf/vol4.html

9. de Gruijter JJ, Brus DJ, Bierkens MF, Knotters M. Sampling for natural resource monitoring. Springer; 2006.

10. Brus DJ, De Gruijter JJ. Effects of spatial pattern persistence on the performance of sampling designs for regional trend monitoring analyzed by simulation of space–time fields. Comput Geosci. 2013 Dec;61:175–83. doi:10.1016/j.cageo.2013.09.001

11. Lark RM. Estimating the regional mean status and change of soil properties: two distinct objectives for soil survey. Eur J Soil Sci. 2009 Oct;60(5):748–56. doi:10.1111/j.1365-2389.2009.01156.x

12. Potash E, Bradford MA, Oldfield EE, Guan K. Measure-and-remeasure as an economically feasible approach to crediting soil organic carbon at scale. Environ Res Lett. 2025 Feb 1;20(2):024025. doi:10.1088/1748-9326/ada16c

13. Nunes MR, Veum KS, Parker PA, Holan SH, Karlen DL, Amsili JP, et al. The soil health assessment protocol and evaluation applied to soil organic carbon. Soil Sci Soc Am J. 2021;85(4):1196–213. doi:10.1002/saj2.20244

14. Oldfield EE, Bradford MA, Wood SA. Global meta-analysis of the relationship between soil organic matter and crop yields. SOIL. 2019 Jan 15;5(1):15–32. doi:10.5194/soil-5-15-2019

15. Soil Survey Staff. Gridded Soil Survey Geographic (gSSURGO) Database for New York [Internet]. [cited 2023 Nov 7]. Available from: https://gdg.sc.egov.usda.gov/

16. Polussa A, Oldfield EE, Runge T, Evans E, Liu S, Forbes E, et al. Counterfactual designs for quantifying management effects on soil carbon change at multi-field scales. Geoderma. 2026 Sep;473:118007. doi:10.1016/j.geoderma.2026.118007

17. Ellert BH, Bettany JR. Calculation of organic matter and nutrients stored in soils under contrasting management regimes. Can J Soil Sci. 1995 Nov 1;75(4):529–38. doi:10.4141/cjss95-075

18. Wendt JW, Hauser S. An equivalent soil mass procedure for monitoring soil organic carbon in multiple soil layers. Eur J Soil Sci. 2013 Feb;64(1):58–65. doi:10.1111/ejss.12002

19. Von Haden AC, Yang WH, DeLucia EH. 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. 2020 Jul;26(7):3759–70. doi:10.1111/gcb.15124

20. Buchkowski RW, Polussa A, Bradford MA. Designing National Forest Inventories for Accurate Estimation of Soil Carbon Change. Glob Change Biol. 2026 Apr;32(4):e70868. doi:10.1111/gcb.70868

21. Poeplau C, Don A. Carbon sequestration in agricultural soils via cultivation of cover crops – A meta-analysis. Agric Ecosyst Environ. 2015 Feb;200:33–41. doi:10.1016/j.agee.2014.10.024

22. Jensen JL, Eriksen J, Christensen BT. Pre-conversion soil organic carbon level did not affect accumulation rate following conversion from arable land to semi-natural grassland. Geoderma. 2024 Jul;447:116930. doi:10.1016/j.geoderma.2024.116930

23. Papritz A, Webster R. Estimating temporal change in soil monitoring: I. Statistical theory. Eur J Soil Sci. 1995 Mar;46(1):1–12. doi:10.1111/j.1365-2389.1995.tb01808.x

24. Papritz A, Webster R. Estimating temporal change in soil monitoring: II. Sampling from simulated fields. Eur J Soil Sci. 1995 Mar;46(1):13–27. doi:10.1111/j.1365-2389.1995.tb01809.x

25. Von Haden AC, Sanford GR, Cates AM, Jackson RD. Paired resampling to detect field-level soil organic carbon stock change. Comment on "Testing the feasibility of quantifying change in agricultural soil carbon stocks through empirical sampling" by Bradford et al. Geoderma. 2024 Aug;448:116959. doi:10.1016/j.geoderma.2024.116959

26. Yanai RD, Stehman SV, Arthur MA, Prescott CE, Friedland AJ, Siccama TG, et al. Detecting Change in Forest Floor Carbon. Soil Sci Soc Am J. 2003 Sep;67(5):1583–93. doi:10.2136/sssaj2003.1583

27. Poeplau C, Prietz R, Don A. Plot-scale variability of organic carbon in temperate agricultural soils—Implications for soil monitoring#. J Plant Nutr Soil Sci. 2022 Jun;185(3):403–16. doi:10.1002/jpln.202100393

28. Conant RT, Paustian K. Spatial variability of soil organic carbon in grasslands: implications for detecting change at different scales. Environ Pollut. 2002 Mar;116:S127–35. doi:10.1016/S0269-7491(01)00265-2

29. Alexander P, Paustian K, Smith P, Moran D. The economics of soil C sequestration and agricultural emissions abatement. SOIL. 2015 Apr 10;1(1):331–9. doi:10.5194/soil-1-331-2015

30. Antle J, Capalbo S, Mooney S, Elliott E, Paustian K. Spatial heterogeneity, contract design, and the efficiency of carbon sequestration policies for agriculture. J Environ Econ Manag. 2003 Sep;46(2):231–50. doi:10.1016/S0095-0696(02)00038-4