Geoderma 438 (2023) 116587

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Multi-site evaluation of stratified and balanced sampling of soil organic carbon stocks in agricultural fields

Eric Potash [a,b,*], Kaiyu Guan [a,b,c], Andrew J. Margenot [a,d], DoKyoung Lee [a,d], Arvid Boe [f], Michael Douglass [a,d], Emily Heaton [d,e], Chunhwa Jang [a,d], Virginia Jin [g], Nan Li [a,d], Rob Mitchell [h], Nictor Namoi [a,d], Marty Schmer [g], Sheng Wang [a,b], Colleen Zumpf [i]

[a] *Agroecosystem Sustainability Center, Institute for Sustainability, Energy, and Environment, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA*
[b] *Department of Natural Resources and Environmental Sciences, College of Agricultural, Consumer and Environmental Sciences, University of Illinois Urbana Champaign, Urbana, IL 61801, USA*
[c] *National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[d] *Department of Crop Sciences, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA*
[e] *Department of Plant Biology, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA*
[f] *Department of Agronomy, Horticulture & Plant Science, South Dakota State University, Brookings, SD 57007, USA*
[g] *USDA-ARS Agroecosystem Management Research Unit, University of Nebraska, Lincoln, NE 68583-0937, USA*
[h] *USDA-ARS Wheat, Sorghum and Forage Research Unit, University of Nebraska, Lincoln, NE, 68583-0937, USA*
[i] *Argonne National Laboratory, Lemont, IL 60439, USA*

A R T I C L E I N F O

Handling Editor: Budiman Minasny

Keywords:
Soil organic carbon
Sampling
Evaluation
Validation
Geostatistics
Bayesian

A B S T R A C T

Estimating soil organic carbon (SOC) stocks in agricultural fields is essential for environmental and agronomic research, management, and policy. Stratified sampling is a classic strategy for estimating mean soil properties, and has recently been codified in SOC monitoring protocols. However, for the specific task of estimating the SOC stock of an agricultural field, concrete guidance is needed for which covariates to stratify on and how much stratification can improve estimation efficiency. It is also unknown how stratified sampling of SOC stocks compares to modern alternatives, notably doubly balanced sampling. To address these gaps, we collected high-density (average of 7 samples ha^{-1}) and deep (average of 75 cm) measurements of SOC stocks at eight commercial fields under maize-soybean production in two US Midwestern states. We combined these measurements with a Bayesian geostatistical model to evaluate stratified and balanced sampling strategies that use a set of readily-available geographic, topographic, spectroscopic, and soil survey data. We examined the number of samples needed to achieve a given level of SOC stock estimation accuracy. While stratified sampling using these variables enables an average sample size reduction of 17% (95% CI, 11% to 23%) compared to simple random sampling, doubly balanced sampling is consistently more efficient, reducing sample sizes by 32% (95% CI, 25% to 37%). The data most important to these efficiency gains are a remotely-sensed SOC index, SSURGO estimates of SOC stocks, and the topographic wetness index. We conclude that in order to meet the urgent challenge of climate change, SOC stocks in agricultural fields could be more efficiently estimated by taking advantage of this readily-available data, especially with doubly balanced sampling.

1. Introduction

Soil organic carbon (SOC) has been greatly diminished by agricultural activities, contributing to climate forcing (Sanderman et al., 2017). In order to monitor SOC stocks under agricultural land use and manage stock increases to offset further anthropogenic climate forcing, accurate measurement of SOC stocks at the field scale is essential. In particular, estimating SOC stocks at the scale of an agricultural field, the scalar unit at which management practices are conducted, is needed to develop sustainable management practices and support carbon credits.

However, accurately estimating the total SOC stock of an agricultural field is at present resource and cost intensive (Smith et al., 2020;

* Corresponding author at: Agroecosystem Sustainability Center, Institute for Sustainability, Energy, and Environment, University of Illinois Urbana-Champaign, Urbana, IL 61801, USA.

E-mail address: epotash@illinois.edu (E. Potash).

https://doi.org/10.1016/j.geoderma.2023.116587
Received 5 January 2023; Received in revised form 8 May 2023; Accepted 27 June 2023
Available online 28 July 2023

Oldfield, 2022), hindering or even precluding scalability of SOC stock estimation. The relatively high spatial variability of SOC stocks necessitates a large number of soil samples across the field to precisely estimate the total stock (Garten and Wullschleger, 1999; McBratney and Pringle, 1999). Moreover, there is evidence that subsurface depths (e.g., $>$ 30 cm) contain a major portion of total SOC stocks in soils under agricultural use (Kravchenko and Robertson, 2011; Olson and Al-Kaisi, 2015), requiring deep soil sampling that decreases throughput and is labor-intensive (e.g., Tautges et al., 2019). Bulk density from intact cores and lab analyses of the resulting soil samples by dry combustion for SOC concentration add further cost and labor to the SOC stock measurement process. Improvements in SOC stock estimation can be realized by increasing the accuracy of the estimate and/or by reducing the cost of the estimate. This is because SOC stock estimation error is a decreasing function of the sample size, i.e. the number of cores or pits at which soils are sampled and measurements are made.

To improve SOC stock estimation, there are broadly two potential avenues: sampling and modeling. Both of these avenues can take advantage of auxiliary information, but incorporate it at different stages of the estimation process. For example, remote sensing imagery may be used to inform SOC stock measurement locations, such as by stratified sampling (Potash et al., 2022), or the imagery could be used to predict SOC stocks themselves, such as modeling its relationship to measurements taken at similar fields (Hbirkou, 2012). Some approaches, such as *in situ* spectroscopy, can make use of both modeling and sampling (Smith et al., 2020).

In this article, we are concerned with the sampling strategy avenue for improving SOC stock estimation. Stratified sampling is a classical approach which uses auxiliary information to partition a field into strata and apply simple random sampling to each one. Stratified sampling has recently been codified in carbon crediting protocols (Oldfield et al., 2021), but for the specific task of estimating SOC stocks in agricultural fields, concrete quantitative guidance is lacking on which auxiliary information to use and what benefits can be expected. Moreover, it is unknown how stratification performs for SOC stock estimation compared to modern alternatives, notably balanced sampling, which also leverages auxiliary information to potentially improve estimation efficiency but does not require constructing a stratification (Deville and Tillé, 2004).

A major reason for this gap in knowledge about the performance of sampling strategies for SOC stock estimation is that, similar to conducting SOC stock assessments, evaluations of sampling strategies are themselves also resource intensive. Evaluating sampling strategies in a field traditionally requires implementing the strategies to estimate the SOC stock of the field. As a result, there are only a handful of studies evaluating SOC stock sampling strategies. Even then, these studies do not quantify the uncertainty of the evaluation, which may be large enough to qualitatively affect the interpretation of the results (Potash et al., 2022).

The present study addresses challenges to evaluating SOC stock sampling strategies in two ways: a rich evaluation dataset and a Bayesian evaluation methodology. Our dataset consists of high-density (an average of 7 samples ha^{-1}) and deep (to an average depth of 75 cm) measurements of SOC stocks at eight commercial fields across four states of the US Midwest, a globally important region of maize (*Zea mays* L.) and soybean (*Glycine* max L.) production and a major industry focus area for SOC sequestration. The Bayesian methodology of Potash et al. (2022) allows us to combine this unique dataset with a model to map SOC stocks to evaluate a variety of sampling strategies without having to implement each one in the field. Moreover, unlike a traditional standard error estimate, the Bayesian methodology provides quantification of uncertainty to better enable comparison of sampling strategies. Leveraging this dataset and methodology, we provide a comprehensive evaluation of stratified and balanced sampling strategies for estimating SOC stocks at the field scale, assessing whether these strategies can reliably improve estimation efficiency compared to simple random sampling (SRS). In doing so, we contribute much-needed quantitative evidence to support SOC stock sampling strategy selection. Moreover, we share our dataset to maximize the scientific impact of our work. Finally, we discuss the implications for sampling strategy selection for SOC research and monitoring programs.

2. Materials and methods

2.1. Data

We used field-scale data on SOC stocks from eight commercial fields across two US Midwestern states: Illinois and Nebraska (Fig. 1). SOC stock measurement locations (Fig. 3) were gridded (sites IL-DG, IL-PT, IL-RD, IL-RS, IL-RT), gridded within subplots (IL-BR, IL-MC), or model-based (NE). Across the fields, sample density averaged 7 samples per hectare and sample depth averaged 75 cm (Table 1). Vertical core samples were taken using a Giddings probe (Giddings machine company: Windsor, CO) mounted on an all-terrain vehicle or a tractor. The intact cores were removed from plastic liners and sectioned into depthwise segments, varying by site from 5 to 30 cm in length, and homogenized by gentle hand crumbling. Gravimetric water content was measured by drying 5–7 g of subsample at 100 °C for 24 h, at which point the mass of soil was found to be constant. Bulk density (BD) for each depth interval was obtained by dividing the oven-dry mass of soil from each section (g) by the volume of the segmented portion of the core (cm^3). At the sampled depths, there were no rock fragments (less than 2 mm diameter), meaning that measured bulk density is that of the fine earth fraction. Soils from each segment were air-dried at ambient (25 °C), ground to pass a 2 mm sieve, and measured for total carbon concentration by dry combustion. For soil samples with pH $>$ 7.2 (Soil Science Division Staff, 2017), inorganic carbon was removed with the addition of 1% HCl (Walthert et al., 2010) so that the total carbon measured by dry combustion can be interpreted as organic carbon. Among the six sites at which pH was recorded, less than 2% of samples had pH $>$ 7.2 (Table S1). Inorganic carbon in samples with pH $>$ 7.2 was measured at one site (NE) by measuring total carbon on separate soil subsamples with and without acidification. These measurements supported the assumption that inorganic carbon was negligible in samples with pH $<$ 7.2 (Fig. S1).

We collected the following auxiliary information (covariates) proposed by Potash et al. (2022) because of their (1) potential to predict SOC, (2) recommendation in SOC monitoring protocol guidance (Oldfield et al., 2021), and (3) availability in public databases for every point of the field, a requirement for stratified and balanced sampling. From SSURGO (Soil Survey Staff, 2022), we obtained the soil series map units and the estimated SOC stock to the sampled depth for each map unit. From the National Elevation Dataset (U.S. Geological Survey, 2018), we collected elevation information, from which we derived three topographic covariates: slope, aspect, and topographic wetness index (TWI). We used northing and easting geographic coordinates, measured in meters from the SW corner of each site. Finally, we used an SOC Index (SOCI) defined as *blue / (green × red)* (Thaler et al., 2019). We computed the index from the most recent Landsat ARD image available prior to planting and sampling that was free from clouds, snow, and crop residues. An image was considered residue-free if the average NDTI was less than 0.35 (Beeson et al., 2020). The Landsat ARD product was chosen to provide uniform imagery across the range of sampling years (2012–2021) in our dataset. All of these covariates were processed to a 10 m × 10 m UTM raster grid (100 pixels ha^{-1}) using bilinear interpolation.

We also included an additional variable: predictions of SOC stock from the POLARIS soil mapping product (Chaney et al., 2019). POLARIS combines national pedon databases and SSURGO summaries with a machine learning model to predict various soil properties including SOC concentration and bulk density across the contiguous United States.

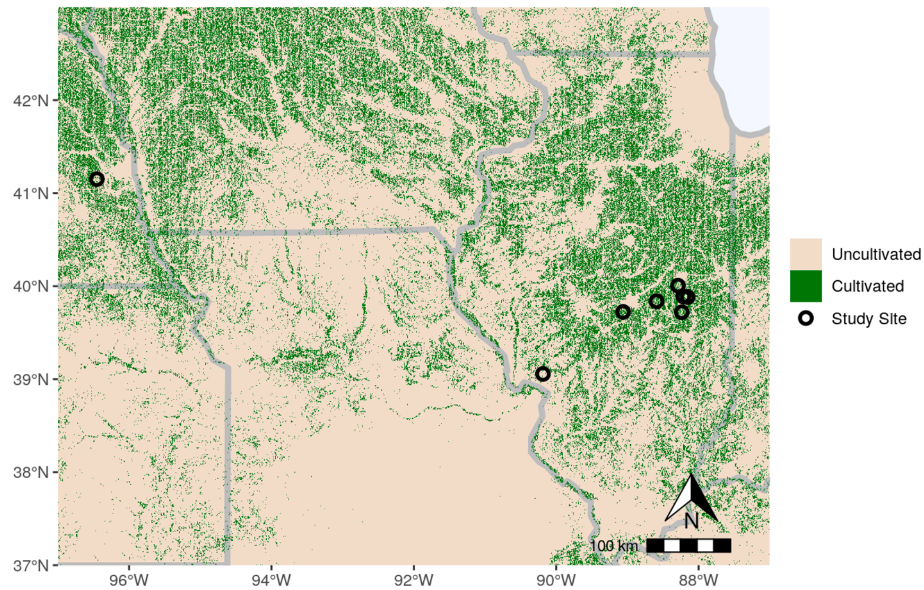

Fig. 1. Map of United States Midwest showing cropland (USDA 2020 National Cultivated Layer) and the eight Midwest field sites included in the study.

Table 1
Site characteristics. Soil texture was measured at IL-DG, IL-MC, and IL-PT and derived from SSURGO at the other sites. Abbreviations: mean annual temperature (MAT), mean annual precipitation (MAP).

Site	IL-BR	IL-DG	IL-MC	IL-PT	IL-RD	IL-RS	IL-RT	NE
Location (°)	−90.19, 39.06	−88.24, 39.72	−89.06, 39.72	−88.59, 39.84	−88.21, 39.89	−88.29, 40.01	−88.15, 39.88	−96.45, 41.15
Area (ha)	10	19	23	20	31	31	31	12
MAT (°C)	11	11	12	11	11	11	11	10
MAP (cm)	94	94	93	94	94	94	94	80
Tillage	No-till	Strip and conventional	Conventional (chisel)	Conventional (chisel)	No-till	Conservation	Conventional (chisel)	No-till
Soil Texture	Silt loam	Silt loam	Silty clay loam	Silty clay loam	Silt loam	Silt loam	Silty clay loam	Silt loam
Sample Depth (cm)	60	90	75	90	60	60	60	120
Sample date	June 2019	Oct 2021	June 2021	June 2021	Oct 2021	April 2020	Oct 2021	June 2012
Sample size	59	86	50	89	247	223	229	144
Sample density (samples ha^{-1})	6	4	2	4	8	7	7	12
Core depth segments (cm)	0–10, 10–20, 20–30, 30–60	0–15, 15–30, 30–45, 45–60, 60–75, 75–90	0–15, 15–30, 30–45, 45–60, 60–75	0–15, 15–30, 30–45, 45–60, 60–75, 75–90	0–30, 30–60	0–15, 15–30, 30–60	0–30, 30–60	0–5, 5–15, 15–30, 30–60, 60–90, 90–120

2.2. Sampling strategies

We evaluated five sampling strategies: SRS, univariate stratified, multivariate stratified, balanced, and doubly balanced sampling. Each strategy was evaluated in each of the eight fields at a range sampling densities to capture performance at varying levels of estimation accuracy and resources: 0.25, 0.5, 1.0, and 2.0 samples ha^{-1}. All five strategies use probability sampling, which supports robust and unbiased estimation of the population mean (i.e., mean SOC stock of a given field).

Stratified and balanced sampling have the potential to improve on the SRS baseline by incorporating auxiliary information (covariates) such as remote sensing into the sampling strategy. In addition to choosing which auxiliary information to include, stratified sampling requires several further choices: rescaling these covariates to make them comparable, an allocation of samples among the strata, and the number of strata (de Gruijter et al., 2006). While the standard k-means method for stratifying only supports continuous covariates, there are other clustering algorithms that accommodate categorical covariates (Huang, 1998).

Balanced sampling (Deville and Tillé, 2004; Brus, 2016) selects samples that are representative in the sense that the (inverse probability weighted) mean value of a covariate (e.g. slope) at the sample locations is equal to the mean value across the field. Balanced sampling has several advantages over stratified sampling. First, it can naturally incorporate categorical covariates, including the SSURGO soil series. Since SSURGO SOC stocks are constant within each soil series, balancing on soil series automatically balances on SSURGO SOC stocks. Thus it was not necessary to include the SSURGO stocks as a covariate in (doubly) balanced sampling. Second, users do not have to make the somewhat arbitrary choices listed above for constructing a stratification (Grafström and Schelin, 2014). One disadvantage of balanced sampling is that its uncertainty quantification is less rigorous than simple or stratified sampling because it is not possible to have a design-based unbiased variance estimate (Grafström and Schelin, 2014).

Doubly balanced sampling (Grafström and Tillé, 2013) builds on balanced sampling by ensuring that samples are not only balanced, i.e. the average covariate value in the sample equals the average in the population, but that the covariate values of the sample are well-spread in the population covariate distribution. Compared to balanced sampling, doubly balanced sampling may provide more efficient estimation as well as better support for mapping (Brus, 2015).

The POLARIS SOC stock predictions are unique among our covariates in that they are intended to provide predictions of SOC stocks at each location in the field. The *cumrootf* rule (Dalenius and Hodges Jr. 1959) is a stratification that takes advantage of this additional information. The recently developed stratification *ospats* (de Gruijter, 2016) is another such strategy applied to SOC stock estimation. However, *ospats* requires estimates of the prediction error (R^2 and autocorrelation range). Since neither of these is provided by POLARIS, we applied the *cumrootf* strategy to stratify the fields using the POLARIS predictions of SOC stock.

2.3. Evaluation of sampling strategies

We sought to evaluate how these sampling strategies (section 2.2) would perform at our eight study sites. This was accomplished by combining the high-density SOC stock measurements with a Bayesian geostatistical model to simulate implementations of the sampling strategies. Also known as ex-ante evaluation (Potash et al., 2022, see also Hofman and Brus, 2021), this evaluation method has several advantages over traditional ex-post evaluation, which evaluates a strategy by implementing it. Instead, we can repurpose existing SOC stock measurements to evaluate alternative sampling strategies, without having to implement those strategies. We can also evaluate multiple strategies using a single common set of measurements. Specifically, a Bayesian Kriging with external drift (KED) model of SOC stock (see Supplementary Methods) was used to simulate 200 SOC stock maps for each field. This model includes all covariates (Table 2) except the SSURGO map unit (to avoid a singular regression design matrix). For each sampling strategy and sample density, we generated 200 sample designs. Each of the 200 × 200 combinations of an SOC stock map and sampling strategy led to a point estimate and confidence interval (CI) for mean SOC stock.

The relative error was calculated for each of these estimates relative to the mean SOC stock of the corresponding map. For each map, sampling strategy (SRS, univariate stratified, multivariate stratified, balanced, and doubly balanced), and sample density there were thus 200 relative errors, one for each sample. The relative error bound for this map was then calculated as the 95th percentile of these 200 values. There is thus a relative error bound for each of the 200 posterior maps. We then compared the sample size needed by each strategy to achieve a given relative error by assuming, based on the statistical theory of probability sampling, that the relative error bound was linear in the inverse square root of the sample size.

We also considered a non-parametric evaluation approach in which the grid of sample locations was treated as a coarse representation of the field and these points were sub-sampled using each of the strategies under evaluation. On the one hand, this approach avoids the use of the KED model and its assumptions employed in our primary evaluation methodology. On the other hand, this approach does not account for the spatial distribution of SOC stocks between the locations sampled in the data. Thus, we do not consider the quantitative results of the non-parametric evaluation to be realistic but we consider its qualitative results as a useful check on the qualitative results of the parametric evaluation.

For multivariate stratification, we used the standard k-means clustering algorithm (de Gruijter et al., 2006). Since k-means does not naturally accommodate categorical covariates, the SSURGO map unit was not included. To put the covariates on a common scale, we used a percent rank transformation (Potash et al., 2022). We also considered the Mahalanobis distance function to account for the correlation structure of the covariates. In the absence of prior information on the variability of SOC stocks within each stratum, we allocated samples in proportion to the size (area) of the strata (de Gruijter et al., 2006). Since uncertainty quantification is essential, each stratum must have at least 2 samples. Thus the number of strata was set such that, under proportional allocation, each stratum received at least two samples. For balanced sampling, we included all covariates except POLARIS predictions and generated samples in R using the `BalancedSampling` package (Grafström and Lisic, 2019).

3. Results

Across the eight sites, Landsat SOCI, SSURGO SOC stock, and TWI covariates had the strongest and positive correlations with measured SOC stocks (Figs. 2-3). Notably, no single covariate was consistently predictive across all sites. Positive correlations are expected given the putative relationships of these covariates to SOC stocks. When all the covariates are incorporated into the KED model, the strength and sign of their coefficients (Fig. S2) are less clear, which is also expected due to correlations among the covariates. The nugget-to-sill ratio, which quantifies the degree of spatial autocorrelation in SOC stock variability that is not explained by the covariates, varies significantly across sites (Fig. S3).

We generated posterior simulations of the SOC stock maps from the fitted KED model for each site. A map of the posterior median of these simulations is plotted for each site in Fig. 4. There was a large variation in SOC stock magnitudes across sites, as well as significant variation in the posterior distributions of SOC stock, both at each pixel (Fig. S5) and across each field (Fig. S6).

The simulated SOC stock maps, which are informed by high density SOC stock measurements, were used to simulate the four sampling strategies. The relative error performance of each sampling strategy was compared to SRS at each site for each sampling density (Fig. S7), across

Table 2
Covariates used by the sampling strategies. Abbreviations: Global Positioning System (GPS), National Elevation Dataset (NED), Soil Organic Carbon Index (SOCI), Soil Survey Geographic Database (SSURGO), Topographic Wetness Index (TWI), Kriging with external drift (KED).

Category	Covariate	Source	Included in Sampling Strategy?				Included in the KED model?
			Simple Random	Univariate Stratified	Multivariate Stratified	(Doubly) Balanced	
Geography	Northing	GPS			✓	✓	✓
	Easting	GPS			✓	✓	✓
Topography	TWI	NED			✓	✓	✓
	Slope	NED			✓	✓	✓
	Aspect	NED			✓	✓	✓
Spectroscopy	SOCI	Landsat ARD			✓	✓	✓
Survey	Soil Series	SSURGO				✓	
	Estimated map unit SOC stock	SSURGO			✓		✓
Multiple	Predicted SOC stock	POLARIS		✓			✓

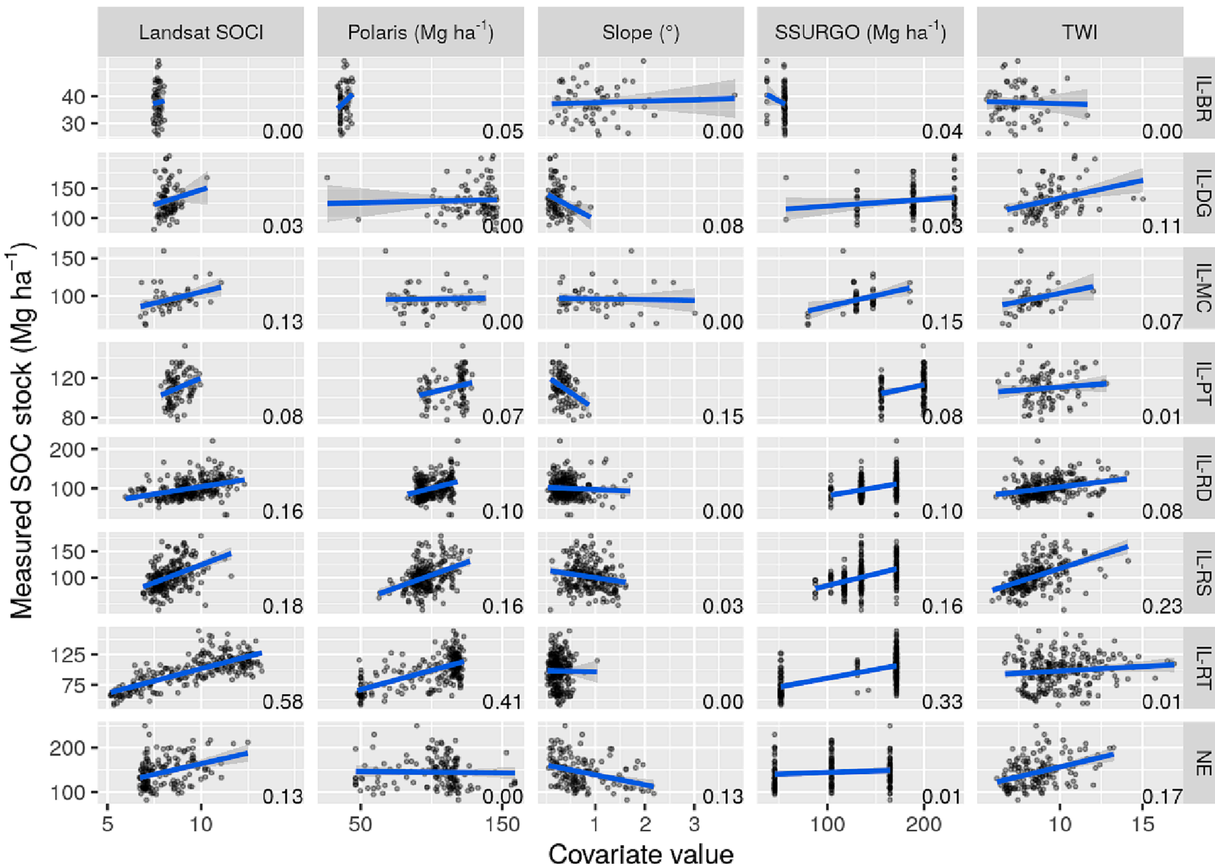

Fig. 2. Univariate relationships between covariates (columns) and measured SOC stock at each of the eight sites (rows). Explained variance (R^2) is displayed in the corner of each panel. Figure excludes three covariates with consistently low R^2 (aspect, Northing, and Easting), which can be found in Fig. S4.

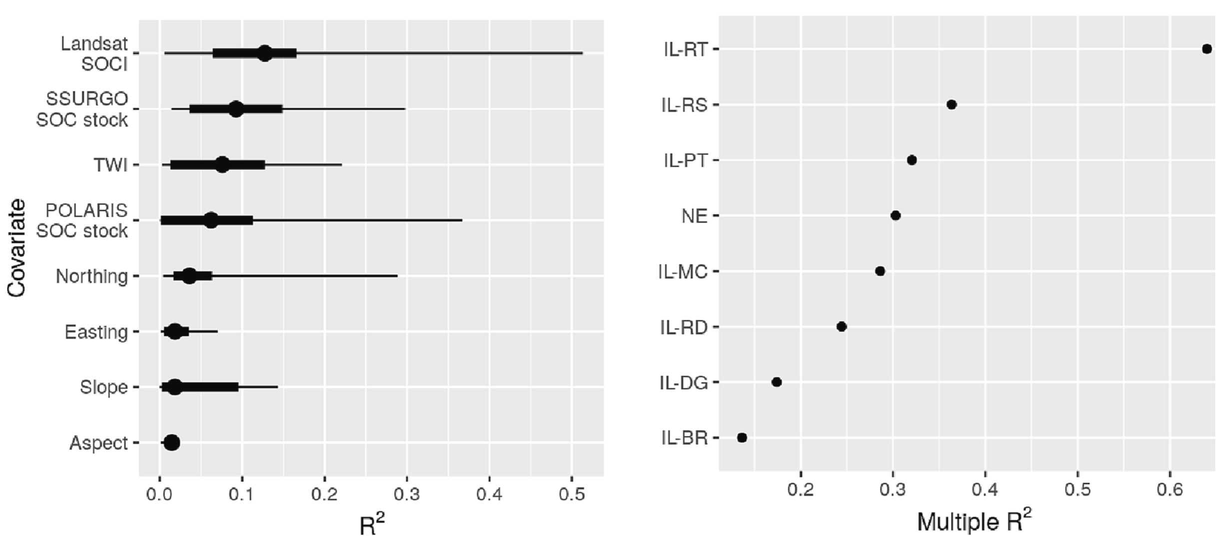

Fig. 3. Left: Distribution across the eight sites of SOC stock variance explained (R^2) by each covariate. Dots and intervals indicate median, 50% and 95% intervals. Right: Variance explained by all covariates (multiple R^2) at each site.

sampling densities (Fig. 5), and averaged across sites (Fig. 6). Using an inverse quadratic relationship between estimation error and sample size (de Gruijter et al., 2006), these relative error reductions were translated into sample size reductions to achieve a given error. Averaged across the eight sites (Fig. 6), stratification using POLARIS predictions of SOC stock would reduce sample sizes by 10% (95% CI, 3% to 16%) compared to SRS. Multivariate stratification would reduce sample sizes by 17% (95% CI, 11% to 24%). Balanced sampling would reduce sample sizes by 25% (95% CI, 20% to 31%). Doubly balanced sampling would reduce sample sizes by 32% (95% CI, 25% to 37%) on average compared to SRS.

The benefit of each sampling strategy for reducing sampling frequency was highly site-specific. For example, while balanced sampling reduced the sample size required at IL-RT by 61% (95% CI, 50% to 68%), the reduction at IL-BR was 16% (95% CI, −5% to 34%). This

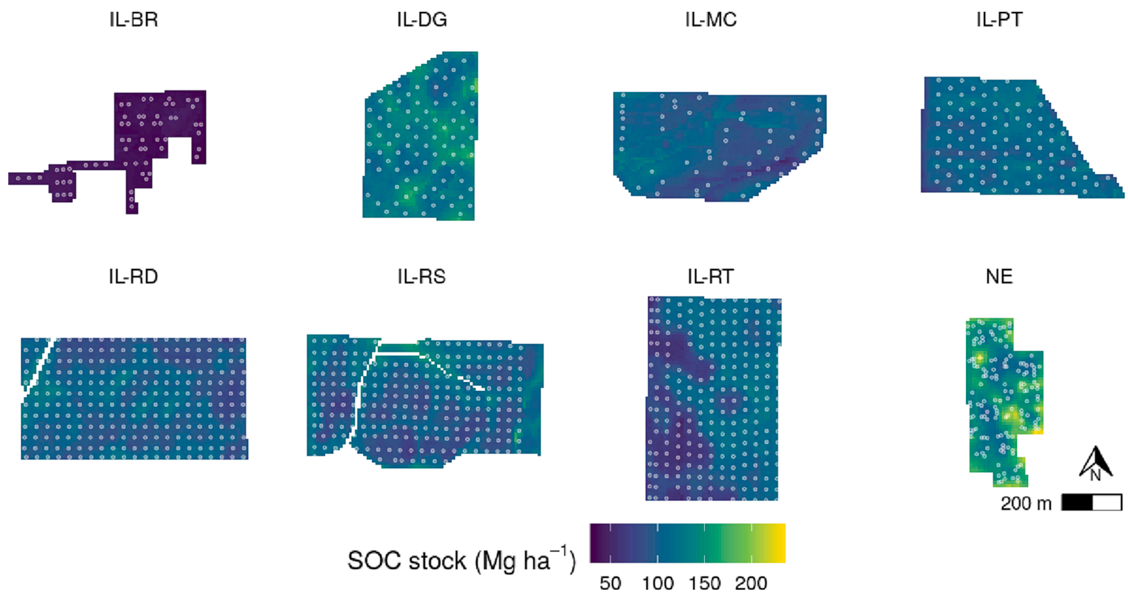

Fig. 4. SOC stock map at each site according to the posterior median prediction using the Kriging with external drift (KED) model. Dots indicate locations at which soils were sampled for measurement of SOC stocks in order to produce these maps.

variability is almost entirely explained by the predictive power of the covariates at each site, as measured by R^2 (Fig. S8). That is, the more predictive the covariates used by a given strategy are of SOC stocks at a site, the better the strategy performed at that site.

4. Discussion

4.1. *Summary of major findings*

In this study we rigorously evaluated field-scale SOC stock sampling strategies across eight commercial agricultural fields in the US Midwest using high-density, deep SOC stock measurements and a Bayesian geostatistical model. The evaluation generated two key findings to inform sampling strategy and sample size selection. First, we provided realistic estimates of estimation strategy performance and found a clear ordering: doubly balanced sampling performed best, followed by balanced sampling, multivariate stratified sampling, univariate stratified sampling, and, lastly, SRS (Fig. 6). Second, the set of covariates available from public databases previously proposed by Potash et al. (2022) explained a substantial proportion of SOC stock variability in most fields, thereby enabling more efficient SOC stock estimation.

Specifically, we found that doubly balanced sampling was the best performing strategy, reducing the sample size required to achieve a given estimation accuracy by about 32% compared to SRS on average

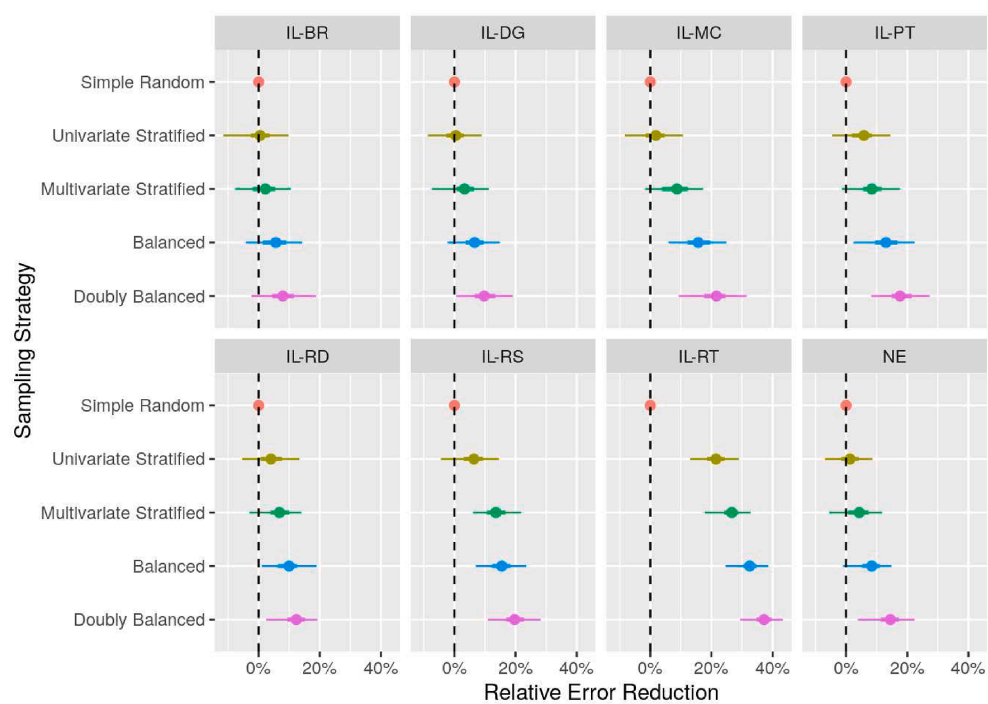

Fig. 5. Relative error bound reduction of each sampling strategy over simple random sampling at each site. Dots and error bars display posterior median, 50%, and 95% intervals.

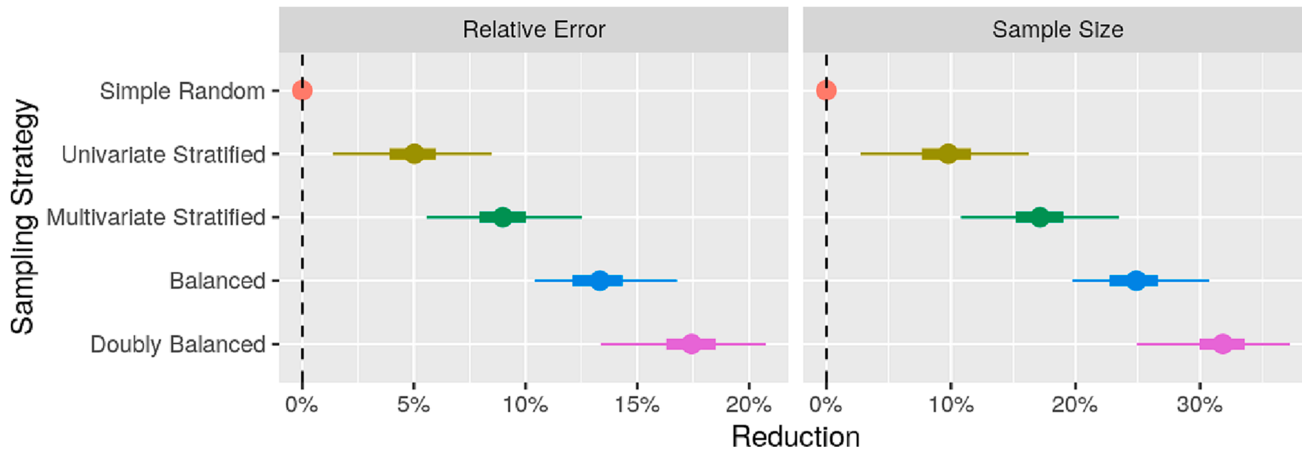

Fig. 6. Reduction in relative error bound and sample size compared to simple random sampling, averaged across sites. Dots and error bars display posterior median, 50%, and 95% intervals.

across our eight sites. The benefits of balanced sampling previously documented at the single IL-RS site (Potash et al., 2022) extended across the range of agricultural fields we sampled in two states in the US Midwest. While multivariate stratified and balanced sampling were previously found to provide similar benefits at the IL-RS site (Potash et al., 2022), our wider study reveals that this site was an outlier: balanced sampling significantly outperformed multivariate stratified sampling averaged across our eight sites (Fig. 6). Moreover, doubly balanced sampling performed even better.

The covariates primarily responsible for the benefits of these strategies were Landsat SOCI, SSURGO SOC stock, and TWI. Notably, the performance of all strategies varied substantially across the sites, though even at its worst-performing site (IL-BR), doubly balanced sampling would likely reduce required sample sizes by 11%. The variability in performance across sites was attributed to the proportion of SOC stock variability explained (R^2) by the covariates across sites, ranging from 14% at IL-BR to 64% at IL-RT (Fig. 3). Moreover, no single covariate dominated in explanatory power (Figs. 2-3) so that in order to obtain the above benefits, a multivariate strategy incorporating all of these covariates should be used. This was in contrast to the previous evaluation by Potash et al. (2022) which, due to its focus on a single site, concluded that a univariate strategy (using SOCI) could provide the same benefits as a multivariate strategy.

The inter-site variability in the relationships between covariates and SOC stocks poses a challenge for modeling SOC stocks in unsampled fields. Indeed, POLARIS predictions of SOC stocks, and hence univariate stratified sampling based on those predictions, were poor in most fields. This may also be due to the fact that the POLARIS model is fitted to relatively sparse soil survey data that rarely includes multiple samples within a given field. Sampling strategies that rely on multiple covariates (e.g. doubly balanced, balanced, and multivariate stratified sampling) are robust to this variability, taking advantage of the correlation between covariates and SOC stocks without assuming any particular relationship between them. Correlations between covariates and SOC stocks enabled such strategies to perform well across the diversity of fields in this study. In contrast, strategies that use predictions of SOC stocks (e.g. univariate stratified sampling using POLARIS) rely not just on covariates being correlated to SOC stocks but on a model that combines those covariates into a single predictor of SOC stocks.

Our findings are sensitive to the parametric assumptions of the KED model used to generate the SOC stock maps used in our evaluation. We chose this model because it can capture spatial autocorrelation, measurement error, and covariate relationships, as well as uncertainty in these parameters. However, to test the sensitivity of our findings to this choice of model we also considered two alternative evaluations: using a Bayesian Additive Regression Trees model (Chipman et al., 2010; see also Potash et al., 2022) as well as a non-parametric approach (Figs. S9-S12). The high degree of qualitative agreement among evaluation results further supports our findings. While the methodology used in this study enabled us to efficiently compare a large set of estimation strategy options without implementing each of them in each field, this ex-ante evaluation does rely on more assumptions than ex-post evaluation. Thus, it may be useful to confirm our findings using ex-post evaluation by comparing some or all of these strategies via field implementation.

4.2. *Expanded contexts*

While our study immediately concerns estimation of absolute SOC stock at the field scale, the findings also have important implications for two expanded contexts: SOC stock change (different estimand) and multifield assessments (larger spatial scale).

4.2.1. *SOC stock change*

There are two basic study designs for estimating SOC stock change: cross-sectional and longitudinal, including both paired and unpaired longitudinal designs. In this section, we discuss (1) how our findings on SOC stock estimation are directly relevant to SOC stock change estimation using cross-sectional and unpaired longitudinal designs, and (2) how our findings are indirectly relevant to SOC stock change estimation using paired longitudinal designs.

In a cross-sectional study (e.g., Yang et al., 2022), also known as space-for-time substitution or chronosequence, SOC stock change due to some treatment (e.g. land use change) is estimated by identifying areas of similar soils that have been treated (e.g. restored from agriculture to prairie) and untreated (e.g. remain in agriculture). By comparing their present-day *absolute* SOC stocks, researchers estimate the effect of the treatment (e.g. prairie restoration) on SOC stock *change*. Cross-sectional studies are advantageous in that they can estimate long-term (e.g. decadal) treatment effects without needing to wait that duration, albeit with more assumptions than a longitudinal design (see below). Our findings on the efficiency of sampling designs for SOC stock estimation can be applied directly to a cross-sectional study: the sampling errors in estimating the treated and untreated SOC stocks are independent, so the squared standard error of their difference is equal to the sum of their squared standard errors. For example, if doubly balanced sampling reduces each SOC stock standard error by about 17% (Fig. 6), it reduces the SOC stock *change* standard error by about 17% as well.

In a longitudinal study, the SOC stock change is quantified by measuring SOC stock at two different time points. The sampling locations at the two time points may be the same (paired) or different (unpaired). In an unpaired longitudinal design, SOC stock change is estimated simply by estimating the absolute SOC stock at each time point and taking their difference. Thus, as with the cross-sectional design above, our findings can be directly applied to an unpaired longitudinal design to measure SOC stock change. One benefit of unpaired designs for regulatory applications is that, unlike paired designs, the follow-up sampling locations may be withheld from farmers so those locations cannot be targeted for treatment, thereby mitigating possible fraud (de Gruijter, 2016; Lawrence, 2020).

However, for longitudinal studies, paired SOC stock designs have two benefits over unpaired designs. First, they enable estimation of SOC stock change at each sampling location, thereby facilitating the creation of an SOC stock change map (see section 4.3.3). Second, a paired design is potentially more efficient than an unpaired design because its variance is reduced by the covariance between the SOC stocks at the two time points (Lark, 2009). However, the magnitude of this benefit may be limited by the combination of remeasurement location error and the large variability of SOC stocks on small length scales (Poeplau et al., 2022). One study found that the covariance between repeated measures of agricultural SOC concentrations across the United Kingdom would reduce the error of SOC concentration change estimation by just 15% (Saby, 2008). A better understanding of this covariance for SOC stocks at the scale of an agricultural field would benefit the design of future SOC stock change measurement projects.

The aforementioned covariance benefits all sampling strategies, including SRS. The additional benefit of sampling strategies such as doubly balanced sampling over SRS in the context of paired sampling for SOC stock change estimation is unclear. We are not aware of any studies on the correlates of within-field SOC stock change. To the extent that our auxiliary information such as TWI is correlated with SOC stock change as well as absolute SOC stock, the strategies presented here will also yield benefits for paired estimation of SOC stock change. We note that our discussion of paired sampling designs also applies to sampling and measuring initial SOC stocks and using a biogeochemical model to estimate the SOC stock change at each location instead of re-measuring SOC stocks.

4.2.2. Multifield assessments

When measuring SOC stocks and their changes across multiple fields, it is natural to stratify at the field level. This reduces estimation error compared to an unstratified design by ensuring that each field receives an appropriate number of samples, commonly proportional to its area. The strategies presented in this article can be used in conjunction with a field-level stratification. The efficiency of such a design can be calculated in terms of the within-field and between-field SOC stock sampling variability. The sampling designs we considered help to reduce the within-field sampling variability. For example, if doubly balanced sampling reduces within-field sampling variability (standard error) by 17% on average (Fig. 6), then by using doubly balanced sampling within each field (stratum) of a multifield project, the total error in the project would also be reduced by about 17% compared to field stratification alone.

In a large multifield project, the sample size within each field may be very small, as few as two samples. It is important to note that the benefits of the sampling strategies we evaluated, while still appreciable, were significantly smaller with smaller sample sizes (Fig. S13). One benefit of (doubly) balanced sampling is that it can be used with as few as two samples per field, whereas stratified sampling requires at least four samples per field (i.e. two strata with two samples each).

4.3. Additional considerations for sampling design selection

Our study provides strong evidence that doubly balanced sampling is the most efficient probability sampling design for estimation of SOC stocks. However, such mean estimation performance is not the only consideration in selecting a sampling design. Here we discuss three other considerations: uncertainty quantification, mapping, and implementation.

4.3.1. Uncertainty quantification

Our main results concern *point* estimation of SOC stocks, i.e. a single number estimate. Arguably more important than point estimation is the quantification of estimation uncertainty, i.e. the construction of a confidence interval around the point estimate. The importance of uncertainty quantification is clear in carbon crediting, where many protocols (e.g. Climate Action Reserve, 2022) issue credits based not on the point estimate of carbon but on the lower bound of a confidence interval in order to reduce the risk of issuing credits for carbon that has not actually been sequestered. Uncertainty quantification is also important in other applications such as field trials, in which an estimate of the uncertainty in the effect of a management practice on SOC stocks is essential. Although less efficient than (doubly) balanced sampling, SRS and stratified sampling have an important advantage: their uncertainty quantification is more directly supported by the central limit theorem, and hence is more rigorous. On the other hand, our study found the uncertainty quantification of (doubly) balanced to be empirically sound (Fig. S14). In some applications this may be sufficient (see section 4.4).

4.3.2. Mapping

Estimating the SOC stock and SOC stock change of an agricultural field is of major importance. These estimands are spatial means (or totals). However, some studies may be interested in other estimands. For example, we may be interested in mapping SOC stocks in order to understand how they vary within a field and with respect to covariates. In this case, the results of this study do not directly apply. However, as discussed by Brus (2015), doubly balanced sampling can support SOC stock mapping due to the fact that its samples are well spread on covariates including spatial coordinates.

4.3.3. Implementation

The practicality of implementation can also be a consideration in the selection of a sampling design. For example, practical considerations are likely a factor in the ubiquity of grid sampling despite the limitations of this non-probability design for rigorous mean estimation. Compared to (doubly) balanced sampling, multivariate stratified sampling requires several additional choices for which there is limited guidance (Oldfield et al., 2021). This lack of information poses a practical impediment to adoption, though recent evaluations (Potash et al., 2022) have narrowed the knowledge gap. Balanced sampling has not been widely adopted and this may be due to its complexity and obscurity relative to SRS or stratified sampling.

4.4. Recommendations

Here we summarize our recommendations given the expanded contexts (section 4.2) of SOC stock change estimation and multifield projects, and the additional considerations (section 4.3) of uncertainty quantification, mapping, and implementation. We describe how our findings can provide concrete guidance for both research and carbon credits.

4.4.1. Research studies

At present, research studies rarely employ any of the probability sampling designs considered in our study. Instead, a non-probability design such as grid sampling is commonly used. Some of the considerations that lead to this choice include: good coverage of the area of interest relative to SRS to improve the accuracy of SOC stock estimates and support the mapping of SOC stocks within the field (section 4.3.2) as well as reduced complexity in field work (section 4.3.3). In fact, when we evaluated grid sampling at our sites, point estimates of SOC stocks were more accurate than any of the probability sampling designs in our study (Fig. S15). However, grid sampling does not support rigorous estimation or uncertainty quantification of SOC stocks or SOC stock changes without a model and a high sample density or strong prior information. Indeed, naively using Student's t distribution to construct confidence intervals for our grid samples, a commonly used but theoretically unfounded procedure, produced poor uncertainty quantification (Fig. S16).

To overcome these shortcomings while continuing to support SOC stock modeling and mapping, we recommend the adoption of doubly balanced sampling. While a complete theoretical understanding of uncertainty quantification using doubly balanced sampling is lacking (section 4.3.1), it is a significant improvement over grid sampling in this regard and our study found that it is empirically sound. For research on and monitoring of SOC stock change (section 4.2.1), doubly balanced sampling using our set of covariates would likely provide significant efficiency gains compared to SRS for cross-sectional and unpaired longitudinal studies. However, more work is needed to understand the efficiency benefits of doubly balanced paired sampling for estimating SOC stock changes (section 4.2.1).

4.4.2. Carbon credits

Major carbon credit protocols currently recommend or require stratified sampling. The efficiency benefit of within-field stratification will be modest at small per-field sample sizes (section 4.2.2). Our evaluation shows that doubly balanced sampling could be significantly more efficient. However, a better understanding of uncertainty quantification for SOC stock estimation under doubly balanced sampling should be developed before this strategy can be recommended for carbon crediting (section 4.3.1). Finally, it is unclear what the benefit of either stratified or doubly balanced sampling would be in conjunction with paired sampling (section 4.2.1).

4.4.3. Sample size selection

Our study provides much-needed guidance on the estimation efficiency benefits of incorporating auxiliary information into SOC stock sampling designs. This information allows a study design to reduce its sample size *relative* to SRS. For example, if a sample size of 1.0 samples ha^{-1} is appropriate for SRS, then the same error on average across our fields could likely be achieved by adopting doubly balanced sampling using 30% lower density, at 0.7 samples ha^{-1}. However, selecting an appropriate sample size for SRS in the first place presents a challenge, since the accuracy of the resulting estimate is a function of within-field SOC stock variability, which varies significantly across fields. For example, we found that the standard deviation of SOC stocks (to various depths, see Table 1) within our 8 fields varied from 6 Mg ha^{-1} (IL-BR) to 35 Mg ha^{-1} (NE) (Fig. S5), with very little of this variability explained by field size (see Lawrence, 2020) or sample depth. However, SOC stock standard deviation was strongly correlated with mean SOC stock (Fig. S17). At present, we are unaware of any proven approaches to predicting SOC stock variability using readily available data, as opposed to preliminary sampling (e.g. de Gruijter, 2016). Future work should explore predicting SOC stock variability to guide sample size selection. Selecting an appropriate sample size is essential to ensuring that SOC stock measurement data can support the analysis sought by investigators.

5. Conclusions

Using a multifield dataset of high-density and deep SOC stock measurements we found that doubly balanced sampling using readily available auxiliary information can significantly reduce the number of samples needed to estimate the SOC stock of an agricultural field. These reductions make doubly balanced sampling using this auxiliary information a promising tool for efficiently monitoring, managing, and researching SOC stocks and SOC stock changes. However, a complete understanding of uncertainty quantification under doubly balanced sampling is lacking so that stratified sampling may be preferred for regulatory applications. We found that stratified sampling can also improve sampling efficiency, but significantly less so than doubly balanced sampling. Finally, we highlight potential next steps for research on SOC stock sampling designs, and have made our SOC stock data available to support continued progress in SOC stock research.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that support the findings of this study are openly available in figshare at https://doi.org/10.6084/m9.figshare.23669304.

Acknowledgements

The authors acknowledge financial support from the DOE ARPA-E SMARTFARM program grant nos. DE-AR0001227 and DE-AR0001382, the NSF Signal-in-Soil program grant no. P008697001, the U.S. Department of Energy, Energy Efficiency and Renewable Energy (EERE), Bioenergy Technologies Office (BETO), grant no. DE-EE0008521, a joint ACES-ICGA funding initiative via USDA Hatch ILLU-802-946, and Agriculture and Food Research Initiative (AFRI) grant no. 2020-67021-32799/project accession no. 1024178 from the USDA National Institute of Food and Agriculture.

Appendix A. Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.geoderma.2023.116587.

References

Beeson, P.C., Daughtry, C.ST., Wallander, S.A., 2020. Estimates of conservation tillage practices using landsat archive. Remote Sens. (Basel) 12 (16), 2665.
Brus, D.J., 2015. Balanced sampling: a versatile sampling approach for statistical soil surveys. Geoderma 253, 111–121.
Chaney, N.W., et al., 2019. POLARIS soil properties: 30-m probabilistic maps of soil properties over the contiguous United States. Water Resour. Res. 55 (4), 2916–2938.
Chipman, Hugh A., George, Edward I., McCulloch, Robert E., 2010. BART: Bayesian additive regression trees. Ann. Appl. Stat. 4 (1), 266–298. https://doi.org/10.1214/09-AOAS285.
Climate Action Reserve. (2022). U.S. Soil Enrichment Protocol: Reducing emissions and enhancing soil carbon sequestration on agricultural lands. https://www.climateactionreserve.org/how/protocols/ncs/soil-enrichment/.
Dalenius, T., Hodges Jr, J.L., 1959. Minimum variance stratification. J. Am. Stat. Assoc. 54 (285), 88–101.
de Gruijter, J.J., et al., 2016. Farm-scale soil carbon auditing. Geoderma.
de Gruijter, J., Brus, D.J., Bierkens, M.FP., Knotters, M., 2006. Sampling for Natural Resource Monitoring. Springer Science & Business Media.
Deville, J.-C., Tillé, Y., 2004. Efficient balanced sampling: the cube method. Biometrika 91, 893–912.
Garten Jr, Charles T., and Stan D. Wullschleger. *Soil carbon inventories under a bioenergy crop (switchgrass): Measurement limitations.* Vol. 28. No. 4. American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America, 1999.
Grafström, Anton and Lisic, Jonathan (2019). BalancedSampling: Balanced and Spatially Balanced Sampling. R package version 1.5.5. https://CRAN.R-project.org/package=BalancedSampling.
Grafström, A., Schelin, L., 2014. How to select representative samples. Scand. J. Stat. 41 (2), 277–290.
Grafström, A., Tillé, Y., 2013. Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. Environmetrics 24 (2), 120–131.
Hbirkou, C., et al., 2012. Airborne hyperspectral imaging of spatial soil organic carbon heterogeneity at the field-scale. Geoderma 175, 21–28.
Hofman, S.CK., Brus, D.J., 2021. How many sampling points are needed to estimate the mean nitrate-N content of agricultural fields? A geostatistical simulation approach with uncertain variograms. Geoderma 385, 114816.
Huang, Z., 1998. Extensions to the k-means algorithm for clustering large data sets with categorical variables. Data Mining and Knowledege Discovery 2, 283–304.
Kravchenko, A.N., Robertson, G.P., 2011. Whole-profile soil carbon stocks: The danger of assuming too much from analyses of too little. Soil Sci. Soc. Am. J. 75 (1), 235–240.
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 (5), 748–756.
Lawrence, P.G., et al., 2020. Guiding soil sampling strategies using classical and spatial statistics: A review. Agron. J. 112 (1), 493–510.

McBratney, A.B., Pringle, M.J., 1999. Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. Precis. Agric. 1 (2), 125–152.

Oldfield, E.E., et al., 2022. Crediting agricultural soil carbon sequestration. Science 375 (6586), 1222–1225.

Oldfield, E.E., A.J. Eagle, R.L Rubin, J. Rudek, J. Sanderman, D.R. Gordon. 2021. Agricultural soil carbon credits: Making sense of protocols for carbon sequestration and net greenhouse gas removals. Environmental Defense Fund, New York, New York. edf.org/sites/default/files/content/agricultural-soil-carbon-credits-protocol-synthesis.pdf.

Olson, K.R., Al-Kaisi, M.M., 2015. The importance of soil sampling depth for accurate account of soil organic carbon sequestration, storage, retention and loss. Catena 125, 33–37.

Poeplau, Christopher, Roland Prietz, Axel, Don, 2022. Plot-scale variability of organic carbon in temperate agricultural soils—Implications for soil monitoring. J. Plant Nutr. Soil Sci. 185 (3), 403–416.

Potash, E., et al., 2022. How to estimate soil organic carbon stocks of agricultural fields? perspectives using ex-ante evaluation. Geoderma 411, 115693.

Saby, N., et al., 2008. Will European soil-monitoring networks be able to detect changes in topsoil organic carbon content? Glob. Chang. Biol. 14 (10), 2432–2442.

Sanderman, J., Hengl, T., Fiske, G.J., 2017. Soil carbon debt of 12,000 years of human land use. Proc. Natl. Acad. Sci. 114 (36), 9575–9580.

Smith, P., et al., 2020. How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. Glob. Chang. Biol. 26 (1), 219–241.

Soil Science Division Staff. 2017. Soil survey manual. C. Ditzler, K. Scheffe, and H.C. Monger (eds.). USDA Handbook 18. Government Printing Office, Washington, D.C.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. Soil Survey Geographic (SSURGO) Database. Available online. Accessed 09/29/2022.

Tautges, N.E., et al., 2019. Deep soil inventories reveal that impacts of cover crops and compost on soil carbon sequestration differ in surface and subsurface soils. Glob. Chang. Biol. 25 (11), 3753–3766.

Thaler, E.A., Larsen, I.J., Qian, Y.u., 2019. A new index for remote sensing of soil organic carbon based solely on visible wavelengths. Soil Sci. Soc. Am. J. 83 (5), 1443–1450.

U.S. Geological Survey, 2018, National Elevation Dataset, accessed 9/29/2022.

Walthert, L., et al., 2010. Determination of organic and inorganic carbon, δ13C, and nitrogen in soils containing carbonates after acid fumigation with HCl. J. Plant Nutr. Soil Sci. 173 (2), 207–216.

Yang, R.-M., et al., 2022. A preliminary assessment of the space-for-time substitution method in soil carbon change prediction. Soil Sci. Soc. Am. J. 86 (2), 423–434.