Article

Optimizing Sampling Strategies for Near-Surface Soil Carbon Inventory: One Size Doesn't Fit All

Charles Bettigole [1,*], Juliana Hanle [2], Daniel A. Kane [3], Zoe Pagliaro [1], Shaylan Kolodney [1,4], Sylvana Szuhay [1], Miles Chandler [1], Eli Hersh [1], Stephen A. Wood [3,5], Bruno Basso [2], Douglas Jeffrey Goodwin [6], Shane Hardy [7], Zachary Wolf [4] and Kristofer R. Covey [1]

1 Skidmore College, Saratoga Springs, NY 12866, USA
2 Department of Earth and Environmental Sciences, Michigan State University, East Lansing, MI 48823, USA
3 Yale School of the Environment, Yale University, New Haven, CT 06511, USA
4 Caney Fork Farms, Carthage, TN 37030, USA
5 The Nature Conservancy, Arlington, VA 22203, USA
6 Texas A&M AgriLife Research, College Station, TX 77845, USA
7 Stone Barns Center for Food and Agriculture, Tarrytown, NY 10591, USA
* Correspondence: cbettigo@skidmore.edu

Abstract: Soils comprise the largest pool of terrestrial carbon yet have lost significant stocks due to human activity. Changes to land management in cropland and grazing systems present opportunities to sequester carbon in soils at large scales. Uncertainty in the magnitude of this potential impact is largely driven by the difficulties and costs associated with measuring near-surface (0–30 cm) soil carbon concentrations; a key component of soil carbon stock assessments. Many techniques exist to optimize sampling, yet few studies have compared these techniques at varying sample intensities. In this study, we performed ex-ante, high-intensity sampling for soil carbon concentrations at four farms in the eastern United States. We used post hoc Monte-Carlo bootstrapping to investigate the most efficient sampling approaches for soil carbon inventory: K-means stratification, Conditioned Latin Hypercube Sampling (cLHS), simple random, and regular grid. No two study sites displayed similar patterns across all sampling techniques, although cLHS and grid emerged as the most efficient sampling schemes across all sites and strata sizes. The number of strata chosen when using K-means stratification can have a significant impact on sample efficiency, and we caution future inventories from using small strata n, while avoiding even allocation of sample between strata. Our findings reinforce the need for adaptive sampling methodologies where initial site inventory can inform primary, robust inventory with site-specific sampling techniques.

Keywords: soil carbon; sampling; grazing; agriculture; stratification; inventory

1. Introduction

The top meter of soil globally contains approximately 1500 Pg C soil organic carbon (SOC). Soils comprise the largest single pool of terrestrial carbon, containing more carbon than the atmosphere and vegetation combined [1–3] with estimates ranging to 3500 Pg C [4,5]. Roughly 116 Pg C has been lost globally in the last 12,000 years due to agriculture, and approximately one third of that within the last two centuries [2]. Open grazing land and croplands comprise approximately 40% percent of ice-free landcover both in the continental United States and globally [6,7]. Because these stocks are large, even small changes in SOC driven by agricultural soil management practices have potentially large impacts at scale [8], suggesting that changes in land management might substantially offset carbon emissions [8,9]. Globally, the application of conservation agriculture could store carbon at similar rates to recent loss [10,11]. Detailed field data are needed to quantify baseline stocks and more accurately model change through time at regional and farm scales, particularly on diversely-managed sites [12].

Effective soil carbon measurement is critical to carbon accounting and sequestration markets, but also of broad agronomic interest. A frequently cited component of soil health initiatives [13] maintains that even relatively small changes in soil carbon concentration are linked to increases in yield and yield stability in agronomic systems [14]. These purported gains may be linked to changes in soil properties such as water holding capacity [15], though a recent meta-analysis suggests this often-cited relationship is uncertain and may be overstated [16].

Collecting soil carbon data at national and property scales is imperative for effective agricultural carbon accounting, understanding links between soil carbon concentrations and management outcomes, calibrating remote sensing models at multiple scales, and projecting change in stocks over time [12,17–21]. The reduced cost of multispectral UAVs, tractor-born sensors, and constellations of small satellites providing high-resolution daily global imagery fostered the recent expansion in precision agriculture [22,23]. While this has brought new resources for land managers, and a suite of predictive geospatial data layers, advances in technology have not occurred in tandem with expanded testing and open reporting of in-field carbon measurements. Data paucity is largely driven by prohibitive sampling costs coupled with high spatial variability [17,24–28], although many countries and governing bodies continue to build systematic carbon inventories at large scales (e.g., Australia, EU, USA).

Approaches to effectively resolve local variability have developed from design-based to model-based methods [29]. While simple random sampling can often produce clustered sample locations, and thus areas with insufficient coverage, more regular grid sampling may be either too sparse to capture local heterogeneity or too cost intensive to be practicable [30–33]. Historical alternatives range from nested sampling designs with variable, pre-determined distances between locations to collecting multiple samples from one location [34–36]. In recent decades, researchers have defined land areas by common characteristics, sampling them as sub-units of larger areas [29].

Drawing from large, often public, datasets and advances in remote sensing, stratification by coarse geospatial covariates has become an accepted practice [17,20]. When the sampled variable is spatially autocorrelated, stratification can increase per sample precision compared to that of non-stratified sampling [25,37–40]. Stratification increases sampling efficiency when grouping strata by variables that closely relate to variation in the variable(s) being sampled; if this correlation is closer, the power of its improvement over non-stratified sampling increases [41–43]. This method produces the greatest improvement in sampling efficiency when it produces strata that comprise small internal variation, reducing total variation per sampled areas; or small size, where strata otherwise might not be sampled sufficiently [29,39]. Multiple national soil carbon monitoring programs have applied stratified sampling designs [41]; however, recommendations for both strata design and sample allocation weights across strata are not often well defined. Researchers recommend its use for both US national monitoring as well as farm-scale soil carbon measurement [41,44]. Stratified soil carbon sampling designs have applied varied geospatial covariates without converging on a standard combination [45,46]. Common stratification covariates include slope, aspect, soil type, modeled soil carbon values, soil minerals, soil mineral maps, apparent soil electrical conductivity, and crop yield [45–47]. We note that modeled soil carbon values may be locally unreliable [41]. Increasing the number of geospatial covariates can result in more efficient sampling, but can also yield too many strata [17,48]. Lack of agreement over the best method to define strata and the complexity of applying robust stratification and sample selection process are perhaps the best explanation for the fact that the majority of on-farm soil inventories still rely on simple regular grid sampling [17].

Following site stratification, sampling designs vary, with different methods recommended by authors of studies across management systems and geographies [47,49]. Within-strata soil sampling design methods include conditioned Latin hypercube sampling (cLHS), Neyman allocation, simple random sampling, stratified random sampling, even sam-

pling, even sampling with pairs, proportional sampling, and modifications of these methods [20,36,49]. Sampling designs have been observed to increase efficiency differently on different sites. However, most studies compare a small number of sampling systems at a time, often on a single site or within a single geography. Sampling designs based on site-specific variability can improve design precision, but require a priori knowledge of that variability [49]. Smith et al. [17] recommend non-intensive pre-sampling data collection to estimate variability. An approach based on a priori estimates of site variability may be increasingly useful as more local data is available, or when data from analogous systems can be applied to increase sampling efficiency at new sites.

There is a need to more robustly assess stratification and within-strata sampling designs across geographies. Identifying an effective, standard approach to sampling for soil carbon concentrations can significantly impact carbon sequestration monitoring efforts: low-cost, scalable, high-resolution soil carbon information that can quantify effects of land management and management transitions is critical to verification, sending robust data to support government and commercial incentive programs [24,50].

To that end, we compare the efficacy of 12 simulated soil carbon sampling design methodologies via a baseline sample optimization using common data from high-density soil carbon sampling campaigns across four agricultural study sites in Virginia, Tennessee, and New York [14,51]. Our objective is to identify techniques that optimize the inventory of near-surface (0–30 cm) soil carbon concentrations. We ask several questions: (a) what combination of geospatial covariates for stratification and within-strata sampling design produce the most efficient sampling systems; (b) how do sampling systems vary in performance across sites; and (c) what drives variation in sampling system performance. This analysis will critically inform field research to address potential links between soil carbon and agronomic outcomes [14] and landscape- and farm-scale carbon stock assessment and budgeting [51].

2. Materials and Methods

We intensively sampled soil carbon at four research farms in the Eastern United States from 2018 through 2020 (Figure 1). Sampling intensity varied across farms, from 1.99 to 5.61 samples per hectare. Stone Barns Center for Food and Agriculture (SB) is located in the southern Hudson River valley (41°06′18.3″ N 73°49′39.3″ W), primarily on fine sandy loam soils, dominated by the Charlton fine sandy loam [52]. Summer 2018 sampling occurred primarily on grazing lands in their first year of new management, with mixed-species grazing from cows, sheep, and ducks. The majority of lands were previously managed as hay fields. Oak Spring Gardens (OSG) is located in Upperville, VA (38°57′42.2″ N 77°51′21.9″ W), in the Piedmont of northern Virginia, primarily on silty clay loam and loam soils, with the Purcellville silty clay loam as the dominant soil type [52]. We sampled at OSG across open grassland (conserved and unmanaged since 2016), orchards, and intact forests in spring 2020. Caney Fork Farm (CFF) is located in Carthage, TN, USA (36°14′ N 85°54′ W), along the Caney Fork River on the Highland Rim, primarily on Armour silt loams [52]. The area sampled in Summer 2019 included warm season and cool season pasture in current management since 2015, along with a commercial chestnut-pasture mix. Lock 7 (L7) is a recently acquired parcel of Caney Fork Farms in Carthage, TN, USA (36°17′44.2″ N 86°00′41.7″ W), historically in conventionally managed annual crops, and as of 2020 in the process of transitioning to regeneratively managed crops and pasture. Soils are primarily on Holston loam [52].
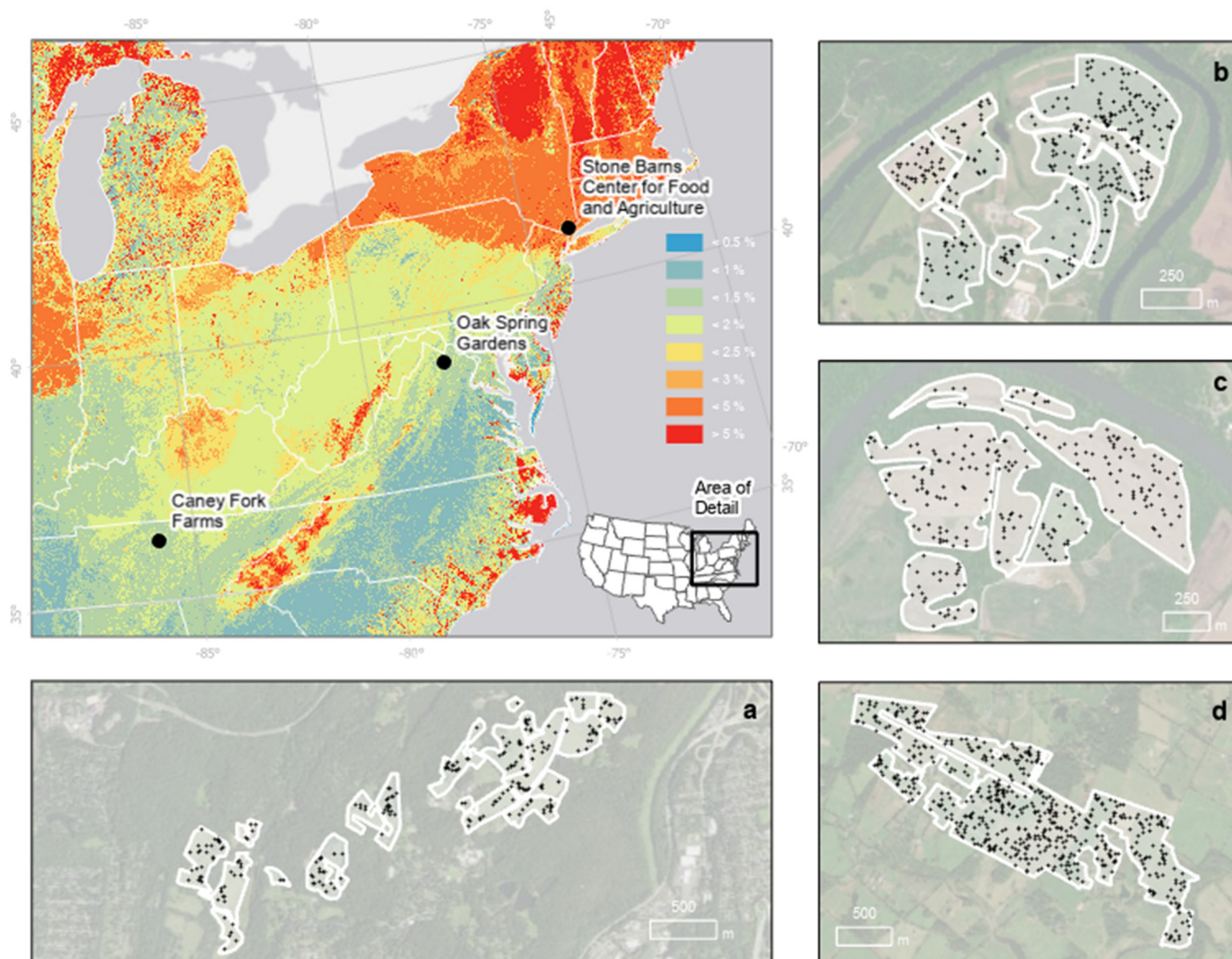
Figure 1. Location and layout of study sites: (**a**) Stone Barns Center for Food and Agriculture; (**b**) Caney Fork Farm; (**c**) Lock 7 at Caney Fork Farm; (**d**) Oak Spring Gardens. Locator map contains POLARIS mean percent organic matter at 5–15 cm depth, highlighting the N-S gradient in soil carbon values.

Soil carbon sampling sites were initially selected through stratified random sampling with the Google Earth Engine web application Stratifi [53]. The application uses an unsupervised classification algorithm (Weka X-Means) that incorporates vegetation indices from Landsat 8 greenest pixel composites (NDVI, NDWI), topography/slope/aspect (National Elevation Dataset) and estimated soil carbon stock from gSSURGO, SoilGrids, and POLARIS [54–57]. It creates a series of "strata" or areas with similar combinations of the above attributes. Weka X-means optimizes the number of strata based on the variability of covariates and we limited possible strata to between three and ten. Random sampling was allocated based on area of strata. Inputs to stratification included NDVI, Slope, Aspect, Elevation, and estimated soil carbon stock (gSSURGO) at all sites. We sampled each site at a density of 2 samples ha^{-1} or greater, with the intention of sampling beyond the sampling intensity required by a simple power analysis. This enabled post hoc model-selection optimization. We navigated to sample locations with the ArcGIS Collector mobile app, using smartphones paired with a Bad Elf Surveyor+ GPS. We sampled to 0–15 cm and 15–30 cm depths using metal collection buckets and 19 mm ship auger drill bits attached to a cordless hammer drill. After air drying, all soils were sent to Ward Laboratories in Kearney, NE, for total carbon (TC)

analysis by dry combustion, following oven-drying. We conducted data analysis with RStudio Version 1.4.1717, ArcGIS Pro 2.7, and Google Earth Engine.

After sampling, we excluded results from all forested areas and redrew study area boundaries to ensure consistency across our predominately open, agricultural study sites. We calculated summary statistics and measures of spatial autocorrelation with semivariograms in R for all study sites and spatial statistics for clustering (Getis-Ord) and autocorrelation (Moran's I) in ArcGIS Pro. We performed a power analysis for each of our full datasets to calculate the ideal number of samples for each property at 90% confidence ($z = 1.645$) where $E = 5\%$ error of whole study site mean carbon values and σ = standard deviation of the sample mean: $\left(n = \left(Z \times \frac{\sigma}{E} \right)^2 \right)$ [25]. We note here that this analysis looks at optimization of soil inventory; to optimize based on the ability to detect a given change in concentrations, sample intensities would be significantly higher.

We conducted a Monte Carlo bootstrapping approach by simulating the application of 12 sampling techniques on our data, iteratively increasing sampling intensity for each methodology from 5 to 245. With 500 iterations at each sample intensity and two clustering treatments (3 strata and 5 strata), this represents 240,000 simulated sampling events for each technique, and 2.88×10^6 sampling events per each site. As our empirical sampling at each site was biased towards stratified-random techniques, we developed a hybrid sampling simulation based on the combination of a raster surface and true empirical samples. For example, if we were to simulate taking 20 simple random samples at CFF in a single iteration, we would place those samples within the site boundary and either (a) extract soil carbon measurement from an existing empirical sample if within a 30 m radius of a given random sampling location, or (b) extract predicted soil carbon values from an interpolated raster surface. With this technique, we were able to avoid the biases inherent in our initial sampling designs in our Monte Carlo simulation—i.e., even though samples were initially collected via stratified random sampling, this a/b approach limits the effects on our non-randomized initial approach.

Predicted, interpolated raster surfaces were generated using Empirical Bayesian Kriging (EBK) (ArcGIS Pro 2.7) for each study site. We calculated RMSE to ensure that models were within the range of variability of each study site. These four EBK raster layers at ten-meter resolution provided a continuous predictive surface of total carbon across each of the study sites (Figure 2).

We generated strata at each site using K-means unsupervised classifications (R stats package). We ran the bootstrapping model with two sets of strata limitations, the first at three strata—the minimum for many sampling protocols (e.g., CAR, FAO, VCS [58–60])—and the second at five strata. We generated these strata with two different sets of input data, the first using a single Polaris organic matter layer and the second with a set of Open Geospatial covariates (Table 1). We ran this analysis in order to compare the efficacy of stratified sampling techniques using (a) a single predictive layer of our variable of interest vs. (b) multi-covariate inputs related to productivity and landform. Input data for Polaris was derived from 5–15 cm log-transformed mean om tiles from POLARIS, a 30 m resolution national-scale probabilistic soil layer [56], and was designed to mimic inputs needed for sampling designs using a Neyman Allocation, such as *Ospats* [20]. Input data for SG included NDVI (Sentinel 2); elevation, slope, and northness (national elevation dataset); and predicted SOC (gSSURGO).
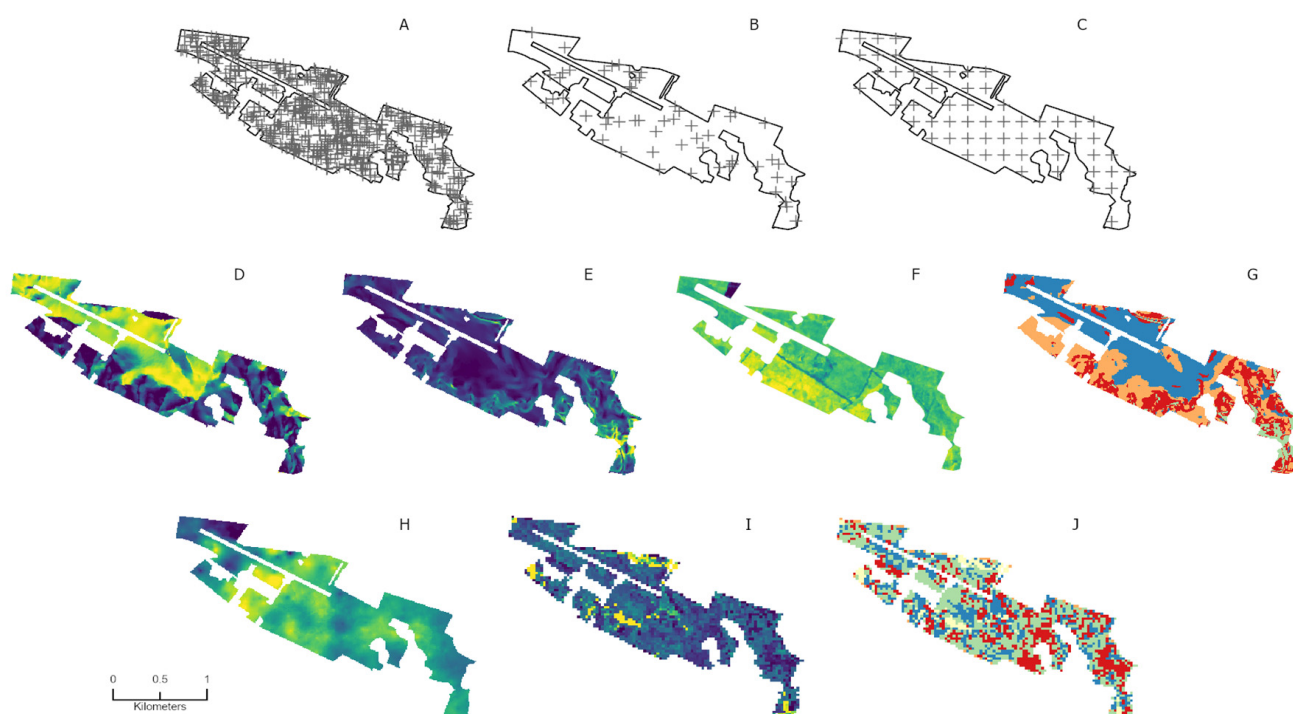
Figure 2. Sampling inputs at Oak Spring Garden. (**A**) all samples on the study site, (**B**) SRS at intensity of the power analysis (*n* = 58), (**C**) Grid at power analysis results intensity, (**D**) Northness (yellow, closer to north, blue, closer to south), (**E**) Slope (yellow is steeper, min: 0, max: 17.3), (**F**) NDVI (yellow is higher NDVI, min: 0.22, max: 0.81), (**G**) Stratified geospatial from (**D–F**), (**H**) Interpolated mean carbon from measured TC using Empirical Bayesian Kriging (EBK) (yellow is higher, min: 0.79, max: 2.71), (**I**) Polaris mean OM (yellow is higher, min: 0, max: 2.77), (**J**) K-means Stratified from I.

Table 1. Information about modeling approaches including inputs to stratification and strata.

Model	Sub-Model Labels	Geospatial Inputs	Number of Strata
Simple Random Sample	SRS	Study Area Boundary	-
Grid Sample	Grid	Study Area Boundary	-
cLHS	cLHS (p), cLHS (g)	p: Polaris Mean OM 5–15 cm g: NDVI (Sentinel 2 L1c), Slope, Northness, Soil C (GSSURGO)	-
K-means	area (p), even (p), bias (p), neyman (p), area (g), even (g), bias (g), neyman (g)	p: Polaris Mean OM 5–15 cm g: NDVI (Sentinel 2 L1c), Slope, Northness, Soil C (GSSURGO)	3/5

For each K-means classification (Open Geospatial and Polaris) we allocated sampling within individual strata using four-substrata random sampling approaches (Table 1): (1) Neyman allocation optimizes sample allocation for a given strata based on total samples (*n*), mean covariate values (e.g., geospatial covariates or Polaris) per strata (x_s), mean covariate across all strata (x), in-strata standard deviation of covariates (σ_s), and standard deviation across all strata (σ): $n_s = n \times \frac{x_s \times \sigma_s}{x \times \sigma}$ [61]; (2) area-weighted sampling (AWS) randomly allocates samples to strata on an equal area basis such that larger strata received a proportionally larger number of random samples; (3) even-weighted sampling randomly allocates an equal number of random samples to strata regardless of strata size;

and (4) mean bias-weighted sampling randomly allocates samples weighted towards strata (n_k) with average laboratory-measured TC values (x_k) closest to the TC mean $(\overline{x}_k)$ for the study site: $n_k = \log\left(\frac{\sum |x_k - \overline{x}|}{x_k}\right)$. While we use laboratory values for mean bias-weighting, future studies could alternatively use ex-ante estimates of predicted or locally observed SOC values.

In addition to the eight K-means sampling approaches outlined above, we simulated three sitewide sampling approaches—simple random sampling (SRS), grid sampling (Grid), and Conditioned Latin Hypercube Sampling (cLHS) (Table 1). We used the spSample tools in the R package sp to create simple random and regular (Grid) sample designs. Conditioned Latin Hypercube Sampling (cLHS) allocates a random sample to optimally represent the distribution of ancillary input data, such as environmental covariates [62]. We used the cLHS R package [63] and the Open Geospatial and Polaris data as inputs—thus creating two cLHS sampling approaches.

For each of these 12 sampling simulations, we performed Monte Carlo bootstrapping of 500 iterations at 240 unique sampling intensities to achieve an estimate of the mean for which the 90% confidence interval was within a 5% margin of error of the true site mean per laboratory measurements (Figure 3). This is the point at which we assume the distribution of the sample is no longer significantly different from the distribution of the carbon samples for the entire study site. We simulated a minimum of five and a maximum of 245 samples chosen from within the study area for each site for each sampling approach.
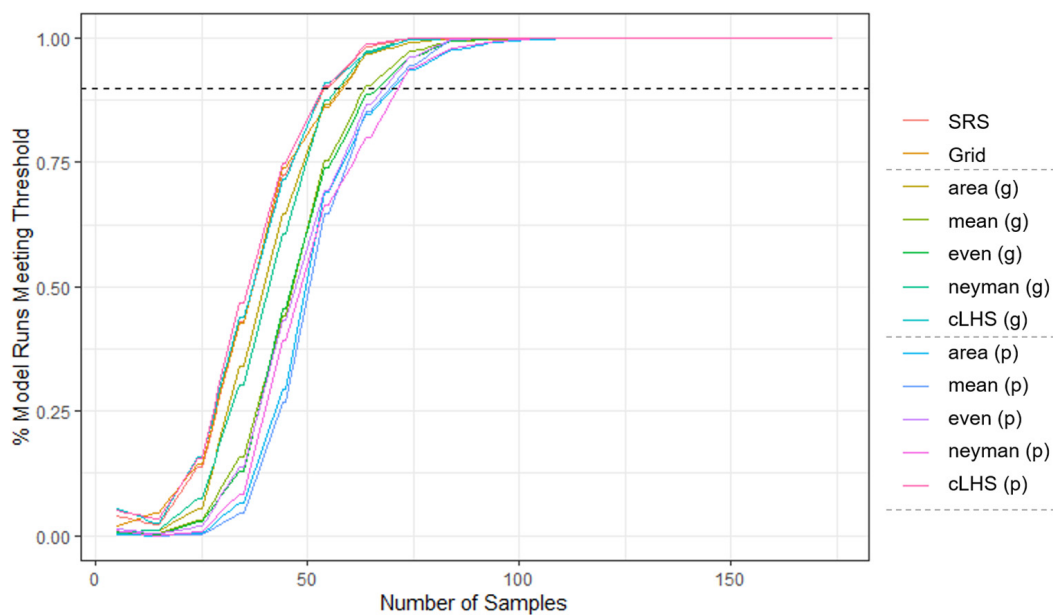

Figure 3. Monte Carlo Bootstrapping at OSG. Each line represents one of 12 different models and shows the percent of 500 model runs that meet the target threshold out of 500 model replicates. The dashed horizontal line sits at 90%.

Simulated points draw their values from the nearest field sampled location using the st_nearest function (sf package CRAN) if within 30 m of a physically sampled site. If further than 30 m, values are assigned to simulated points based on underlying interpolated surface derived from Empirical Bayesian Kriging (Figure 4). At each sample intensity n, and iteration i, we calculated the mean and standard errors and used these values to calculate a 90% confidence interval for each n $\left(ci_n = 1.645 \times \frac{\sigma_{ni}}{\sqrt{n}}\right)$. We recorded each iteration where this 90% confidence interval was less than or equal to 5% of the site population mean: $ci_n \leq E \rightarrow ideal_n$. We set an optimization threshold (e.g., ideal sample size for each sampling technique) at the first sample step n where 90% of iterations were less than or equal to this threshold.
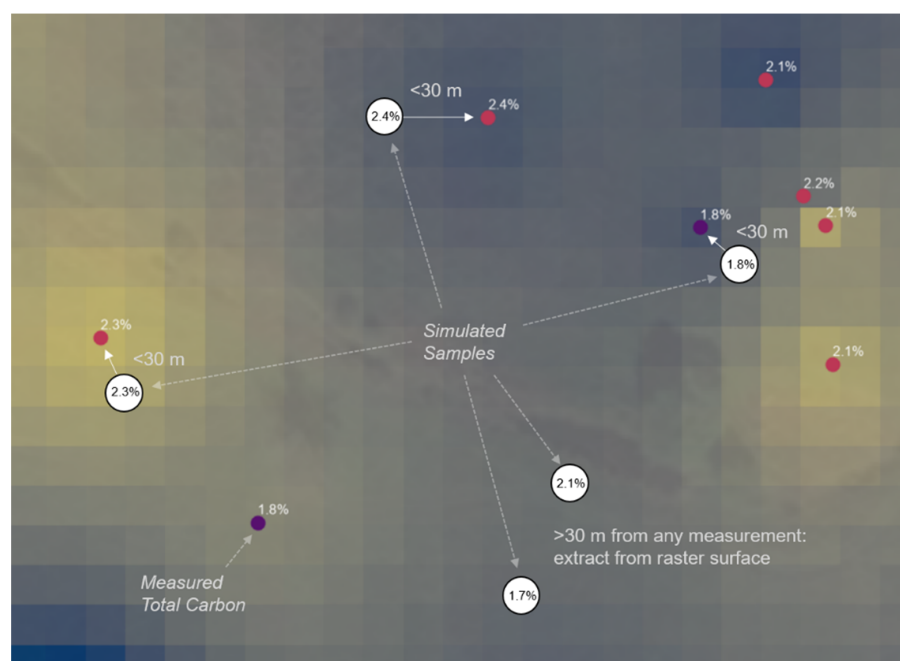
Figure 4. Obtaining carbon values from the nearest known location in simulated sampling. Each of the five points draws its value from the nearest location if within 30 m of a physically sampled site. If further than 30 m, values are assigned to simulated points based on underlying interpolated surface derived from EBK. Dark purple points have lower percent carbon, bright yellow have highest.

We grouped optimization results from our sampling approaches into broad categories based on inputs to stratification (No Strata, Polaris, Open Geospatial) or by in-strata sampling techniques (No Strata, cLHS, Area, Even, Neyman, Biased) for the five-strata models. We normalized the results of the analysis for each site prior to grouping, with values of zero indicating the most efficient sampling approach, and values of one indicating the least efficient. We used these groupings as inputs to ANOVA analysis (multcompvar) to test for significant differences between stratification inputs or in-strata sampling approaches.

3. Results

Soil carbon content of our study sites increased in magnitude from south to north, with low mean carbon values at L7, moderate carbon values at CFF and OSG and high values at SB (Table 2). All sites displayed log-normal distributed carbon values. Site variability (sd) was lowest at CFF, followed by OSG, L7, and SB with the highest. All sites demonstrated significant high-value spatial clustering (Getis-Ord, $p < 0.0001$) and spatial autocorrelation (Moran's I-$p < 0.0001$), indicating the clustering of high-carbon samples. Autocorrelation (range from semivariograms) ranged from 44 m at CFF to 200 m at OSG. While the uncertainty of drawing samples from the EBK surface is not explicitly incorporated into our results, RMSE for our EBK models was generally within the range of variability of our datasets, with the exception of SB, which had a relatively higher RMSE in its EBK.

Per power analysis, ideal samples for all sites ranged from 2.22 samples ha^{-1} at L7 to 0.25 samples ha^{-1} at OSG (Table 2). The farms with the highest SD relative to mean carbon values (L7 and SB) were correlated with higher sample intensity power analysis (Table 2). While OSG (mean: 1.43, sd: 0.33) and CFF (mean: 1.4, sd: 0.29) had similar soil carbon distribution, there is less variability per acre at OSG, and therefore a much lower sampling intensity was produced by power analysis.

Effective K-means sampling intensities differed substantially between sampling in three strata and in five. At five strata, 11 sampling approaches (23%) were less efficient than a power analysis, yet these sampling schemes were restricted to OSG and CFF (Figure 5). At OSG, all approaches using K-means stratification were either equal to or less efficient than

power analysis; at CFF, all K-means approaches (with the exception of Neyman allocation using Polaris) were similarly less efficient than power analysis. At three strata, 15 sampling approaches (31%) were less efficient than the power analysis, spread across all study sites.

Table 2. Summary statistics across sites including sample intensity, soil type, range/sill from semi-variogram, and interpolation error.

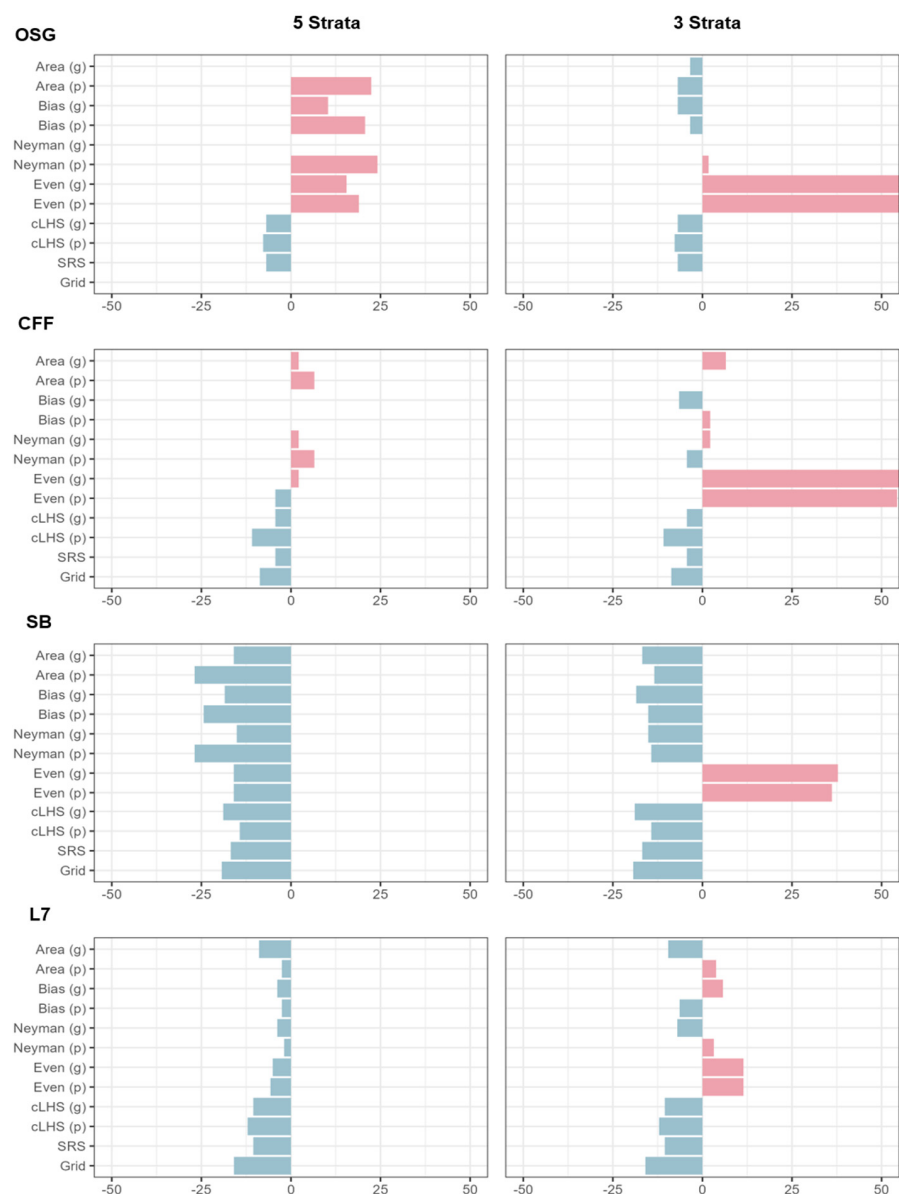| Farm	N-Samp	Ha	Samples ha^{-1}	Power ha^{-1}	Dominant Soil Type	Mean Total C	SD Total C	Range/Psill	EBK RMSE
OSG	569	234.0	2.43	0.25	Purcellville silty clay loam	1.43%	0.33	240.2/0.02	0.283
CFF	344	61.3	5.61	0.75	Armour Silt Loam	1.4%	0.29	44.6/0.02	0.247
SB	207	104.1	1.99	1.14	Charleton Fine Sandy Loam	2.49%	0.82	146.3/0.04	0.758
L7	245	70.7	3.47	2.22	Holston Loam	0.89%	0.34	252.0/0.13	0.256

Figure 5. Results of Monte Carlo analysis for each site at **five** strata (left column) and **three** strata (right column). Blue bars indicate models that outperformed a power analysis. Red bars indicate models that underperformed the power analysis. Values are percent deviation in ideal samples from power analysis.

3.1. Five Strata

At five strata, we see little evidence that any one in-strata sampling approach for K-means stratification outweighs any other. With the exception of K-means stratification using Polaris, all other strata types were significantly more efficient than power analysis, led by Grid and SRS, and followed by Open Geospatial. cLHS significantly outperformed power analysis, either using Open Geospatial or Polaris inputs. All sites had a cLHS model as the most efficient model, with the exception of L7, which had cLHS as the second-best model behind Grid.

Broadly, we see that sites with high variability (SB, L7) see benefits from stratified sampling approaches, while sites with low variability (CFF, OSG) show fewer improvements in sample efficiency with stratified sampling.

3.2. Three Strata

In contrast, with three strata, we see a significant reduction in sampling efficiency when using even allocation within strata (Figure 5). This holds true across all sites, although is most dramatic at OSG and CFF. Unlike five strata, we see K-means stratification with area-weighted, mean-biased, and Neyman allocation outperforming the power analysis. cLHS and No Strata are the most efficient sampling schemes, outperforming all K-means stratification schemes. All sites with three strata stratification have regular grid sampling as the most efficient technique, with the exception of OSG, where cLHS optimizes sample intensity.

On the whole, we see sites with low variability (OSG, CFF) display few increases in efficiency from stratified sampling, with outsized penalties for choosing Even sample allocation approaches. Sites with high variability (SB, L7) still show improvements to sample efficiency with stratified sampling, with the exception of Even sample allocation, yet we note that the penalties associated with this approach are not as large as those seen at low variability sites.

3.3. All Site Summary

Per ANOVA results for five-strata approaches, across all sites we see No Strata (Grid and SRS) and cLHS with both Polaris and Open Geospatial inputs significantly outperform (a) K-means sampling with Polaris and (b) power analysis (Figure 6). K-means sampling with Open Geospatial does not significantly outperform power analysis, yet it is also not significantly outperformed by No Strata or cLHS—as are K-means models with Polaris inputs.

There were three clear patterns that emerged across all four study sites: first, no two study sites display similar patterns across all sampling techniques at either stratification level; second, cLHS and Grid emerged as the most efficient sampling schemes across all sites and strata sizes; and third, the number of strata chosen when using K-means shows that stratification can have a significant impact on sample efficiency.
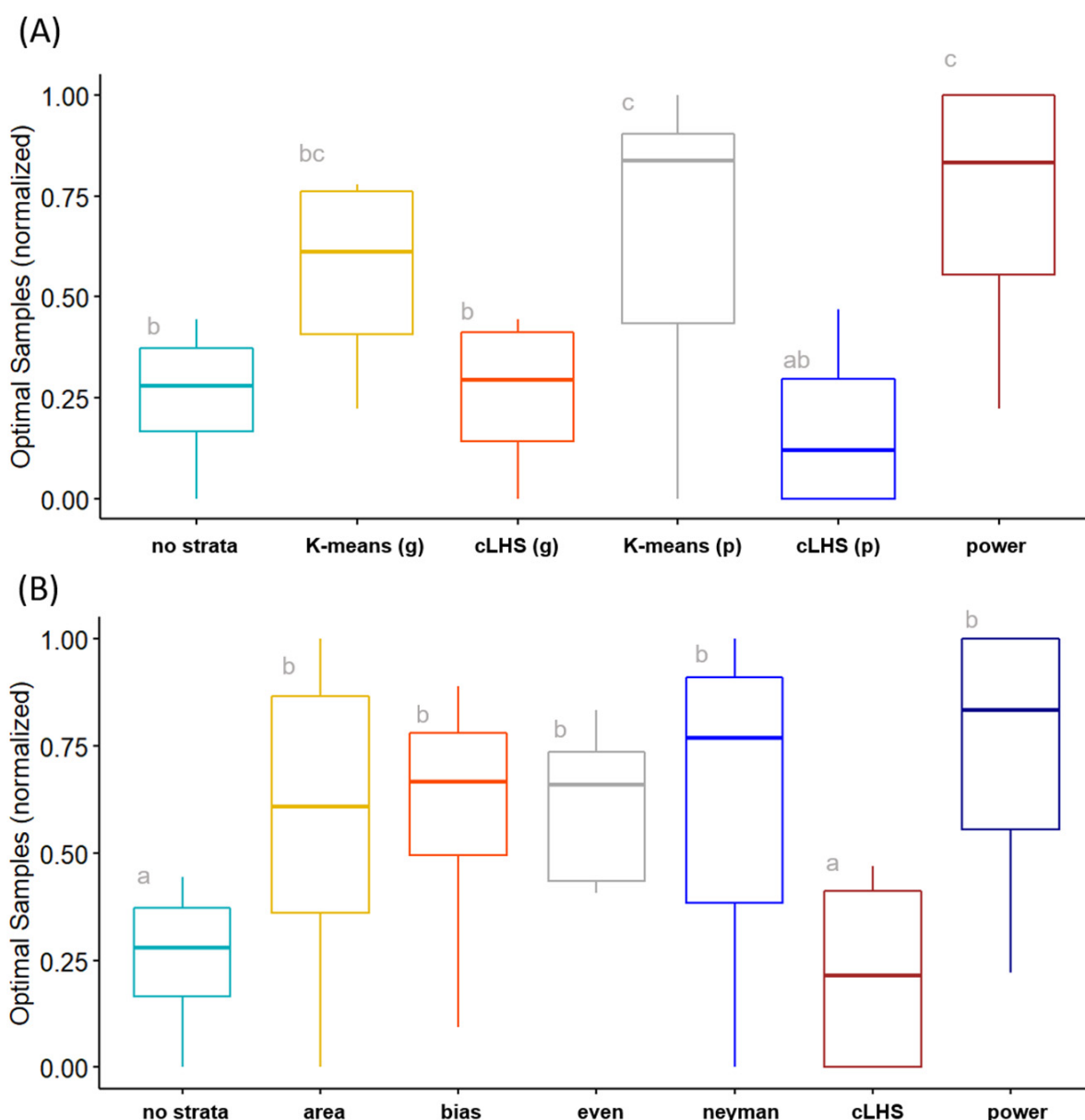
Figure 6. Summaries across all study sites (5 strata only). (**A**) Submodels grouped across study sites and types of inputs to K-means and cLHS. (**B**) Models grouped by in-strata sampling. In both (**A**,**B**), sampling intensities have been normalized at each site from 0 to 1, with zero being the submodel with the lowest sampling intensity and one as the submodel with the highest sampling intensity. Labels on individual boxes reflect results from ANOVA, reflecting significant differences between groups a, b, and c.

4. Discussion

While many agree that sequestering carbon in agricultural soils is a viable natural climate change mitigation strategy [10,64,65], there is continual disagreement about if and how soil carbon stocks can be measured on local and national scales [17,66]. It is clear, however, that regional, national, or even global incentives for agricultural carbon sequestration should be supported by the best-established practices for measuring landscape-scale soil carbon stocks [17].

Our results do not reveal a single optimized sampling scheme applicable to the inventory of soil carbon concentrations at the four study sites considered here. Our analysis does suggest, however, that several general principles can be applied to reduce sampling effort at a given level of desired inventory precision. First, even at the farm-scale, cLHS,

Grid, and SRS generally outperform all forms of stratified sampling. Second, stratification efficiency with both cLHS and K-means is similar when using either the Polaris predictive carbon layer or Open Geospatial predictors, indicating that a single predictive soil carbon layer can be as efficient in designing sampling schemes as a stack of predictive layers. Third, depending on the number of strata, Even-weighted sampling within strata can produce dramatically misleading results and should be avoided at all costs.

4.1. Grid Sampling

While researchers tasked with farm-scale inventory, especially for regulatory purposes, often avoid systematic sampling [67], we show here that regular grid sampling is often among the most efficient sampling methodologies. We posit that at sites similar to ours—with relatively high sample intensities and small study areas—grid sampling is likely to obtain reliable representative samples, while avoiding chance sample clustering around hot-spot carbon locations. While this does not point to the need to perform grid sampling across the board, it indicates that it can be an effective sampling methodology when performed under the right conditions.

We propose that sampling designs for significantly larger regions than those studied here would see a dramatic decrease in the efficacy of Grid sampling. We see the efficacy of Grid sampling fall significantly across our four study sites and as we move into the largest study site (OSG)—the only site where Grid sampling does not outperform our power analysis. Further studies should investigate this trend to confirm the general understanding that design-based methods gain efficiency on larger scales.

4.2. cLHS

Despite broad similarities between cLHS and K-means sampling approaches (i.e., variation across single or multiple input surfaces dictates the placement of sampling locations) [62], cLHS consistently outperforms stratification. Rather than creating strata based on the variability of the study area and then sampling randomly within strata, cLHS skips the intermediate step of defining strata and assigns sampling locations to maximize variability/entropy across all input surfaces. Functionally, sample locations from both techniques should be placed to capture the variability in each system.

However, given the significant effects that number of strata can have on inventory accuracy, cLHS emerges as a methodology more resilient to errors in study design—i.e., choosing an arbitrary number of strata is not possible with cLHS. As both approaches should theoretically accomplish similar goals, it appears to be the prudent choice for our study sites to choose cLHS over stratified sampling for this very purpose. In this, cLHS appears to function equally well using a single predictive carbon layer (Polaris) or a stack of geospatial predictors. We posit that sampling via cLHS may be similarly effective on study sites with similar size and variability.

4.3. Polaris vs. Open Geospatial

Our analysis mirrors others in demonstrating that the choice of inputs for initial stratification can be an important driver of design efficiency [68]. Yet, even as there is wide agreement that stratification is valuable and that soil spatial variability can be described with models such as SCORPAN [69], there is little consensus on which geospatial inputs produce the most precise stratified sampling plan at a given sampling intensity. Potash et al. [70] show that stratified sampling with single covariates (e.g., a Sentinel-2 derived SOC index) or combinations of covariates (Sentinel 2 indices, landform, etc.) leads to reductions in sample error relative to simple random sampling. Across all sites, we see no appreciable difference in sampling efficiency between Polaris and Open Geospatial predictors with cLHS. However, we do show that with the use of K-means stratification, Open Geospatial does slightly outperform Polaris—which significantly underperforms cLHS and Grid sampling (Figure 6). While this, of course, does not hold true for every one of our sites, it highlights the importance of gaining an a

priori understanding of the site level accuracy of predictive geospatial layers in advance of their broad application in sample design. Our results may point to the fact that at small study sites, stratification performs best with empirical input variables that are most accurate at local scales—e.g., slope, NDVI. While national-scale predictive models such as Polaris may be accurate at the national scale [56], their accuracy at the farm scale is often unreliable [41]. While we do not see a correlation between farm size and stratification inputs, we posit that as study area size increases, predictive soil carbon layers may prove to be just as, if not more, efficient in allocating sampling resources.

4.4. Avoiding Small Strata K-Means

We show a dramatic change in sampling efficiency for K-means stratified sampling with even allocation (e.g., even number of samples per strata regardless of area) based on the initial number of strata. K-means is sensitive to outliers [71]; thus, where there are outliers in stratification input layers, clustering can produce one or more relatively small strata by area. If samples are then allocated evenly across all strata (Even) there is then a chance that these "hot-spot" strata are oversampled. For CFF (Figure 7), this becomes especially apparent with Even (g). Strata 3 comprises only 12% of the study site area, yet under an Even sampling, it would receive n/3 samples: this can both skew the mean carbon estimates, and artificially increase variability across the entire site. Especially for CFF, where Strata 2 contains the lowest mean carbon values, this over-allocation of samples to the smallest strata could significantly impact the accuracy of an inventory. This does not present an issue with AWS, as Strata 2 would receive a relatively small number of samples.
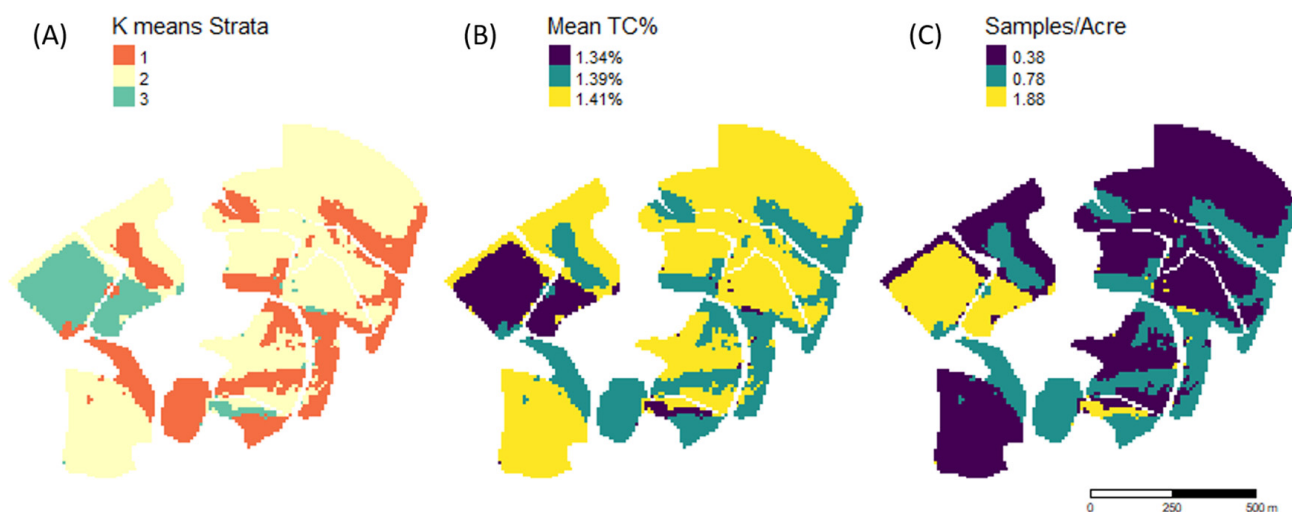

Figure 7. Stratification with certain geospatial surfaces can lead to skewed sampling with certain in-strata sampling techniques. (**A**) Strata for CFF using Open Geospatial; (**B**) mean carbon values (from laboratory data) per strata; and (**C**) number of samples per acre with Even sampling with 100 samples across study area.

This finding is particularly relevant considering that many existing soil monitoring protocols do not explicitly recommend against Even allocation of samples across strata, rather require a minimum number of samples per strata [58,59]. Protocols such as FAO [59] suggest the use of minimum samples and AWS, but leave room for interpretation. Others such as VCS [72] suggest a sample optimization strategy that has the potential to oversample small, yet highly variable strata. However, some protocols do explicitly recommend AWS over Even allocation [60]. While it is possible to optimize the number of strata for a study site (e.g., Elbow Method, Gap Statistic, Silhouette Method), this is difficult without an a priori understanding of population characteristics, and therefore is often overlooked in sample design for soil carbon inventory.

4.5. Cost Penalties

As we show, the difference in precision per number of samples for a given study site can be large. On small or relatively uniform sites where grid sampling is nearly as effective as other methods, or where the difference between sampling approaches is small (e.g., most methods at CFF or L7 with five strata), the sampling cost penalty associated with inefficient design is slight. However, applied to large holdings, regional campaigns, or on sites where design choice considerably impacts sampling efficiency, the cost-penalty for inefficient sampling design quickly adds up (Table 3). Consider our largest study site OSG (234 ha) and a conservative analytical cost of \$20/sample. A land manager looking to obtain a soil carbon inventory with 90% confidence and 5% error would incur a \$90 penalty for choosing sampling intensity from a power analysis (0.25 samples ha^{-1}) as compared to the best performing model (cLHS (p): 0.229 samples ha^{-1}). By employing the worst performing model (Even (p): 0.397 samples ha^{-1}) rather than the best performing model, the cost penalty for inefficient design increases to \$790. These penalties may be larger (as they are for our three other study sites) or smaller depending on the characteristics of the area of interest (Figure 3). At large scales or across cooperatives of multiple smaller operations, these inefficiencies may not be recuperated through management change and/or carbon market payouts.

Table 3. Cost difference for best vs. worst samples (N) scaled up to 1000 ha per site. Parentheses after model names reflect inputs to stratification: geospatial (g) or polaris (p).

	Power Analysis		**Best Model**			**Worst Model**			**Cost Diff Per 1000 ha**	
Farm	**N × ha⁻¹**	**Pwr N**	**Name**	**N × ha⁻¹**	**N**	**Name**	**N × ha⁻¹**	**N**	**Best v Power**	**Best v Worse**
Three Strata										
OSG	0.25	58	cLHS (p)	0.229	53	Even (p)	0.397	93	\$385	\$3376
CFF	0.75	46	cLHS (p)	0.669	41	Even (g)	1.191	73	\$1632	\$10,442
SB	1.14	119	Grid	0.922	96	Even (g)	1.576	164	\$4420	\$13,067
L7	2.22	157	Grid	1.868	132	Even (p)	2.476	175	\$7074	\$12,167
Five Strata										
OSG	0.25	58	cLHS (p)	0.229	53	Neyman (p)	0.308	72	\$342	\$1538
CFF	0.75	46	cLHS (p)	0.669	41	Neyman (p)	0.799	49	\$1632	\$2611
SB	1.14	119	Area (p)	0.836	87	cLHS (p)	0.980	102	\$6149	\$2882
L7	2.22	157	Grid	1.868	132	Neyman (p)	2.179	154	\$7074	\$6225

It is therefore imperative that governing bodies, verifiers, land managers, and others seeking to incentivize wide-spread soil carbon focused practice adoption understand the critical importance of thoughtful sampling design and allocation. Specifically, per our findings, these institutions should explicitly caution participants against (a) small numbers of strata, and (b) the use of even within-strata sample allocation to avoid these costly sampling inefficiencies.

4.6. Adaptive Framework

More broadly, this analysis indicates that none of the commonly applied stratification techniques is likely to emerge as a single optimized sampling scheme applicable to all regions. We therefore recommend an adaptive sampling approach (Figure 8) similar to those recommended by carbon measurement protocols [58–60]. An initial cLHS scheme (or Grid on small properties) can be used to understand spatial auto-correlation, and the variability and magnitude of near-surface carbon values. A power analysis can then identify the appropriate sampling intensity for a given level of precision. Results from this study and future work across more study sties can be used to point out the reduction in sampling intensity from this power analysis for a full sampling campaign.

While not tested in this study, we hope that as more data in a given region becomes available, sampling intensity for AWS or cLHS built from Polaris or Open Geospatial (e.g., with the Stratifi web-app or a K-means/cLHS algorithm in R/Python) or Grid

sampling could be applied directly based on power analysis information gathered from analogous areas (i.e., without pre-sampling). In this way, initial inventories in a given region are likely to incur an 'early adopter' penalty, due to the double sampling necessary to acquire the soil data needed to more efficiently allocate future samples in a given region. Continued collaboration in the acquisition and sharing of farm-level inventories may therefore lead to reduced costs for land managers and an increase in our understanding of sampling design.

Do you have existing soil carbon data for your study area?

No

Pre-sample using best available data to direct sampling intensity (cLHS, Grid or SRS)

Yes

Conduct a Power Analysis:

$$n = (Z \times \frac{\sigma}{E})^2$$

Where n=sampling intensity | σ = standard deviation | E = margin of error

Discount power analysis sampling intensity based on desired modeling framework (e.g. 8% for cLHS)

Create sampling scheme using cLHS

Figure 8. Framework for sampling design. Existing data may be previous soil carbon sampling on study site, on adjacent properties, or local/regional common-practice baselines based on soil type, land management, and landform characteristics.

4.7. Future Work

We recognize several areas where future studies could improve on our work: (a) initial high-intensity sampling designs should be created with either SRS or Grid approaches, rather than the K-means stratification which we used across all study sites; (b) sites should explicitly be chosen across a range of landscape variability, to test these approaches at highly uniform and heterogonous sites; (c) similar techniques should be applied at high intensities across larger study sites (1000+ ha); (d) backfilling our bootstrapped models with interpolated EBK surfaces introduces error into our approach that cannot be quantified—future studies should identify sampling intensities a priori that reduce the need for pulling EBK values; and (e) similar approaches to those presented here should be applied in future studies assessing inventory techniques for soil bulk density and in turn landscape-scale soil carbon stock assessment.

5. Conclusions

Rebuilding soil carbon is emerging as the cornerstone of healthy soil initiatives and as a central component of natural climate solution schemes, yet managing SOC requires agricultural producers to adopt new practices that reverse carbon losses from conventional agriculture (e.g., heavy tillage, continuous overgrazing) [10]. Motivating producers to transition to these practices in support of land stewardship and climate mitigation requires low-cost, scalable, high-resolution soil carbon information that can quantify the costs and benefits of a regenerative agricultural transition [73]. A central challenge limiting widespread adoption of regenerative soil practices therefore is detecting SOC change at management relevant scales (i.e., small changes over large areas) [74].

No clear "best" model emerged from our analysis, however, across all study sites, cLHS appears to be the most effective and resilient sampling methodology, while across

smaller study sites, Grid sampling appears to be just as effective. A single predictive soil carbon layer (Polaris) was as effective as a stack of predictive geospatial layers on these sites (Figure 5). If one were to choose stratified sampling with K-means, we recommend more than three strata and the avoidance of even-weighted sampling.

References

1. Lal, R. Soil Carbon Sequestration Impacts on Global Climate Change and Food Security. *Science* **2004**, *304*, 1623–1627. [CrossRef]
2. Sanderman, J.; Hengl, T.; Fiske, G.J. Soil carbon debt of 12,000 years of human land use. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 9575–9580. [CrossRef]
3. Smith, P. Soils and climate change. *Curr. Opin. Environ. Sustain.* **2012**, *4*, 539–544. [CrossRef]
4. Tifafi, M.; Guenet, B.; Hatté, C. Large Differences in Global and Regional Total Soil Carbon Stock Estimates Based on Soil-Grids, HWSD, and NCSCD: Intercomparison and Evaluation Based on Field Data from USA, England, Wales, and France. *Glob. Biogeochem. Cycles* **2018**, *32*, 42–56. [CrossRef]
5. Scharlemann, J.P.; Tanner, E.V.; Hiederer, R.; Kapos, V. Global soil carbon: Understanding and managing the largest terrestrial carbon pool. *Carbon Manag.* **2014**, *5*, 81–91. [CrossRef]
6. Ramankutty, N.; Evan, A.T.; Monfreda, C.; Foley, J.A. Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Glob. Biogeochem. Cycles* **2008**, *22*, GB1022. [CrossRef]
7. Bigelow, D.; Borchers, A. Major Uses of Land in the United States. 2012. Available online: https://www.ers.usda.gov/webdocs/publications/84880/eib-178.pdf?v (accessed on 8 January 2023).
8. Pezzuolo, A.; Dumont, B.; Sartori, L.; Marinello, F.; De Antoni Migliorati, M.; Basso, B. Evaluating the impact of soil conservation measures on soil organic carbon at the farm scale. *Comput. Electron. Agric.* **2017**, *135*, 175–182. [CrossRef]
9. Liu, L.; Basso, B. Impacts of climate variability and adaptation strategies on crop yields and soil organic carbon in the US Midwest. *PLoS ONE* **2020**, *15*, e0225433. [CrossRef]
10. Griscom, B.W.; Adams, J.; Ellis, P.W.; Houghton, R.A.; Lomax, G.; Miteva, D.A.; Schlesinger, W.H.; Shoch, D.; Siikamäki, J.V.; Smith, P.; et al. Natural climate solutions. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 11645–11650. [CrossRef]
11. Zomer, R.J.; Bossio, D.A.; Sommer, R.; Verchot, L.V. Global sequestration potential of increased organic carbon in cropland soils. *Sci. Rep.* **2017**, *7*, 15554–15558. [CrossRef]
12. FAO. Measuring and Modelling Soil Carbon Stocks and Stock Changes in Livestock Production Systems Guidelines for Assessment. *Livestock Environmental Assessment and Performance (LEAP) Partnership*. 2019. Available online: https://www.fao.org/publications/card/en/c/CA2934EN/ (accessed on 12 August 2021).
13. Lal, R. Soil organic matter content and crop yield. *J. Soil Water Conserv.* **2020**, *75*, 27A–32A. [CrossRef]
14. Oldfield, E.E.; Bradford, M.A.; Wood, S.A. Global meta-analysis of the relationship between soil organic matter and crop yields. *SOIL* **2019**, *5*, 15–32. [CrossRef]
15. Libohova, Z.; Seybold, C.; Willis, S.; Schoeneberger, P.; Williams, C.; Lindbo, D.; Stott, D.; Owens, P.R. Reevaluating the effects of soil organic matter and other properties on available water-holding capacity using the National Cooperative Soil Survey Characterization Database. *J. Soil Water Conserv.* **2018**, *73*, 411–421. [CrossRef]

16. Minasny, B.; McBratney, A.B. Limited effect of organic matter on soil available water capacity. *Eur. J. Soil Sci.* **2018**, *69*, 39–47. [CrossRef]

17. Smith, P.; Soussana, J.F.; Angers, D.; Schipper, L.; Chenu, C.; Rasse, D.P.; Batjes, N.H.; Egmond, V.F.; McNeill, S.; Kuhnert, M.; 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**, *26*, 219–241. [CrossRef] [PubMed]

18. Ward, K.J.; Chabrillat, S.; Neumann, C.; Foerster, S. A remote sensing adapted approach for soil organic carbon prediction based on the spectrally clustered LUCAS soil database. *Geoderma* **2019**, *353*, 297–307. [CrossRef]

19. Conant, R.T.; Smith, G.R.; Paustian, K. Spatial Variability of Soil Carbon in Forested and Cultivated Sites. *J. Environ. Qual.* **2003**, *32*, 278–286. [CrossRef]

20. de Gruijter, J.J.; McBratney, A.B.; Minasny, B.; Wheeler, I.; Malone, B.P.; Stockmann, U. Farm-scale soil carbon auditing. *Geoderma* **2016**, *265*, 120–130. [CrossRef]

21. Mandal, A.; Majumder, A.; Dhaliwal, S.S.; Toor, A.S.; Mani, P.K.; Naresh, R.K.; Gupta, R.K.; Mitran, T. Impact of agricultural management practices on soil carbon sequestration and its monitoring through simulation models and remote sensing techniques: A review. *Crit. Rev. Environ. Sci. Technol.* 2020; *ahead of print*. [CrossRef]

22. Basso, B.; Antle, J. Digital agriculture to design sustainable agricultural systems. *Nat. Sustain.* **2020**, *3*, 254–256. [CrossRef]

23. Basso, B.; Dobrowolski, J.; McKay, C. From the Dust Bowl to Drones to Big Data: The Next Revolution in Agriculture. *Georget. J. Int. Aff.* **2017**, *18*, 158–165. [CrossRef]

24. Paustian, K.; Larson, E.; Kent, J.; Marx, E.; Swan, A. Soil C Sequestration as a Biological Negative Emission Strategy. *Front. Clim.* **2019**, *1*, 8. [CrossRef]

25. Don, A.; Schumacher, J.; Scherer-Lorenzen, M.; Scholten, T.; Schulze, E. Spatial and vertical variation of soil carbon at two grassland sites—Implications for measuring soil carbon stocks. *Geoderma* **2007**, *141*, 272–282. [CrossRef]

26. Garten, C.T.; Wullschleger, S.D. Soil Carbon Inventories under a Bioenergy Crop (Switchgrass): Measurement Limitations. *J. Environ. Qual.* **1999**, *28*, 1359–1365. [CrossRef]

27. Vanguelova, E.I.; Bonifacio, E.; Vos, D.B.; Hoosbeek, M.R.; Berger, T.W.; Vesterdal, L.; Armolaitis, K.; Celi, L.; Dinca, L.; Kjønaas, O.J.; et al. Sources of errors and uncertainties in the assessment of forest soil carbon stocks at different scales—Review and recommendations. *Environ. Monit. Assess.* **2016**, *188*, 630. [CrossRef]

28. Minasny, B.; Malone, B.P.; McBratney, A.B.; Angers, D.A.; Arrouays, D.; Chambers, A.; Chaplot, V.; Chen, Z.; Cheng, K.; Das, B.S.; et al. Soil carbon 4 per mille. *Geoderma* **2017**, *292*, 59–86. [CrossRef]

29. Gruijter, J.J.; Bierkens, M.F.P.; Brus, D.J.; Knotters, M. *Sampling for Natural Resource Monitoring*; Springer: Berlin/Heidelberg, Germany, 2006.

30. Kerry, R.; Oliver, M.A.; Frogbrook, Z.L. Sampling in Precision Agriculture. In *Geostatistical Applications for Precision Agriculture*; Springer: Dordrecht, The Netherlands, 2010; pp. 35–63.

31. Oliver, M.A.; Webster, R. The elucidation of soil pattern in the Wyre Forest of the West Midlands, England. II. Spatial distribution. *J. Soil Sci.* **1987**, *38*, 293–307. [CrossRef]

32. Wadoux, A.M.J.; Brus, D.J.; Heuvelink, G.B.M. Sampling design optimization for soil mapping with random forest. *Geoderma* **2019**, *355*, 113913. [CrossRef]

33. Allen, D.E.; Pringle, M.J.; Page, K.L.; Dalal, R.C. A review of sampling designs for the measurement of soil organic carbon in Australian grazing lands. *Rangel. J.* **2010**, *32*, 227–246. [CrossRef]

34. Youden, W.J.; Mehlich, A. Selection of Efficient Methods for Soil Sampling. *Soil Sci. Soc. Am. J.* **1938**, *2*, 399. [CrossRef]

35. Zhang, Y.; Hartemink, A.E. Sampling designs for soil organic carbon stock assessment of soil profiles. *Geoderma* **2017**, *307*, 220–230. [CrossRef]

36. Wadoux, A.M.J.C.; Marchant, B.P.; Lark, R.M. Efficient sampling for geostatistical surveys. *Eur. J. Soil Sci.* **2019**, *70*, 975–989. [CrossRef]

37. Cochran, W.G. *Sampling Techniques*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 1997.

38. Wallenius, K.; Rita, H.; Mikkonen, A.; Lappi, K.; Lindström, K.; Hartikainen, H.; Raateland, A.; Niemi, R.M. Effects of land use on the level, variation and spatial structure of soil enzyme activities and bacterial communities. *Soil Biol. Biochem.* **2011**, *43*, 1464–1473. [CrossRef]

39. Peltoniemi, M.; Heikkinen, J.; Mäkipää, R. Stratification of regional sampling by model-predicted changes of carbon stocks in Forested mineral soils. *Silva Fenn.* **2007**, *41*, 287. [CrossRef]

40. Rao, P.S. *Sampling Methodologies*; Chapman & Hall/CRC: Boca Raton, FL, USA, 2000.

41. Malone, B.P.; Odgers, N.P.; Stockmann, U.; Minasny, B.; McBratney, A.B. Digital Mapping of Soil Classes and Continuous Soil Properties. In *Pedometrics*; Springer International Publishing: Cham, Switzerland, 2018; pp. 373–413.

42. Parton, W.J.; Schimel, D.S.; Cole, C.V.; Ojima, D.S. Analysis of Factors Controlling Soil Organic Matter Levels in Great Plains Grasslands. *Soil Sci. Soc. Am. J.* **1987**, *51*, 1173. [CrossRef]

43. Peltoniemi, M.; Mäkipää, R.; Liski, J.; Tamminen, P. Changes in soil carbon with stand age-an evaluation of a modelling method with empirical data. *Glob. Change Biol.* **2004**, *10*, 2078–2091. [CrossRef]

44. Spencer, S.; Ogle, S.M.; Breidt, F.J.; Goebel, J.J.; 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**, *1*, 167–178. [CrossRef]

45. Minasny, B.; McBratney, A.B.; Malone, B.P.; Wheeler, I. Digital Mapping of Soil Carbon. In *Advances in Agronomy*; Elsevier Science & Technology: Amsterdam, The Netherlands, 2013; Volume 118, pp. 1–47.

46. Malone, B.P.; Minansy, B.; Brungard, C. Some methods to improve the utility of conditioned Latin hypercube sampling. *PeerJ* **2019**, *7*, e6451. [CrossRef]

47. Singh, K.; Whelan, B.; Goss, M. Soil carbon change across ten New South Wales farms under different farm management regimes in Australia. *Soil Use Manag.* **2020**, *36*, 616–632. [CrossRef]

48. Miller, B.A.; Koszinski, S.; Hierold, W.; Rogasik, H.; Schröder, B.; Van Oost, K.; Wehrhan, M.; Sommer, M. Towards mapping soil carbon landscapes: Issues of sampling scale and transferability. *Soil Tillage Res.* **2016**, *156*, 194. [CrossRef]

49. Shao, S.; Zhang, H.; Fan, M.; Su, B.; Wu, J.; Zhang, M.; Yang, L.; Gao, C. Spatial variability-based sample size allocation for stratified sampling. *CATENA* **2021**, *206*, 105509. [CrossRef]

50. Costa, C.; Dittmer, K.; Shelton, S.; Bossio, D.; Zinyengere, N.; Luu, P.; Heinz, S.; Egenolf, K.; Rowland, B.; Zuluaga, A.; et al. How Soil Carbon Accounting Can Improve to Support Investment-Oriented Actions Promoting Soil Carbon Storage. 2020. Available online: https://samples.ccafs.cgiar.org/how-soil-carbon-accounting-can-improve-to-support-investment-oriented-actions-promoting-soil-carbon-storage/ (accessed on 12 August 2021).

51. Zhou, W.; Guan, K.; Peng, B.; Margenot, A.; Lee, D.; Tang, J.; Jin, Z.; Grant, R.; DeLucia, E.; Qin, Z.; et al. How does uncertainty of soil organic carbon stock affect the calculation of carbon budgets and soil carbon credits for croplands in the U.S. Midwest? *Geoderma* **2023**, *429*, 116254. [CrossRef]

52. Soil Survey Staff Web Soil Survey. Natural Resources Conservation Service, United States Department of Agriculture. Available online: https://websoilsurvey.nrcs.usda.gov/app/ (accessed on 29 August 2022).

53. Stratifi. Available online: https://charliebettigole.users.earthengine.app/view/stratifi-beta-v21 (accessed on 12 June 2020).

54. Natural Resources Conservation Service, United States Department Of Agriculture Gridded Soil Survey Geographic Database (gSSURGO). 2016. Available online: https://www.nrcs.usda.gov/resources/data-and-reports/gridded-soil-survey-geographic-gssurgo-database (accessed on 29 August 2022).

55. Poggio, L.; De Sousa, L.M.; Batjes, N.H.; Heuvelink, G.B.M.; Kempen, B.; Ribeiro, E.; Rossiter, D. SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty. *Soil* **2021**, *7*, 217–240. [CrossRef]

56. Chaney, N.W.; Wood, E.F.; McBratney, A.B.; Hempel, J.W.; Nauman, T.W.; Brungard, C.W.; Odgers, N.P. POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma* **2016**, *274*, 54–67. [CrossRef]

57. Gesch, D.; Oimoen, M.; Greenlee, S.; Nelson, C.; Steuck, M.; Tyler, D. The national elevation dataset. *Photogramm. Eng. Remote Sens.* **2002**, *68*, 5–11.

58. CAR Soil Enrichment Protocol | Version 1.0 |. 2020. Available online: https://www.climateactionreserve.org/how/protocols/ncs/soil-enrichment/ (accessed on 12 August 2021).

59. FAO. FAO GSOC MRV Protocol. 2020. Available online: https://www.fao.org/3/cb0509en/cb0509en.pdf (accessed on 12 August 2021).

60. VCS. VCS Module VMD0021 Estimation of Stocks in the Soil Carbon Pool. 2012. Available online: https://verra.org/wp-content/uploads/imported/methodologies/VMD0021-Estimation-of-Stocks-in-the-Soil-Carbon-Pool-v1.0.pdf (accessed on 12 August 2021).

61. Edgar Bueno OptimStrat: Choosing the Sample Strategy. 2020. Available online: https://cran.r-project.org/web/packages/optimStrat/optimStrat.pdf (accessed on 12 August 2021).

62. Minasny, B.; McBratney, A.B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. *Comput. Geosci.* **2006**, *32*, 1378–1388. [CrossRef]

63. Roudier, P. Clhs: A R Package for Conditioned Latin Hypercube Sampling. 2011. Available online: https://cran.r-project.org/web/packages/clhs/clhs.pdf (accessed on 29 August 2022).

64. Rumpel, C.; Amiraslani, F.; Chenu, C.; Garcia Cardenas, M.; Kaonga, M.; Koutika, L.; Ladha, J.; Madari, B.; Shirato, Y.; Smith, P.; et al. The 4p1000 initiative: Opportunities, limitations and challenges for implementing soil organic carbon sequestration as a sustainable development strategy. *Ambio* **2020**, *49*, 350–360. [CrossRef]

65. Bossio, D.A.; Cook-Patton, S.C.; Ellis, P.W.; Fargione, J.; Sanderman, J.; Smith, P.; Wood, S.; Zomer, R.J.; von Unger, M.; Emmer, I.M.; et al. The role of soil carbon in natural climate solutions. *Nat. Sustain.* **2020**, *3*, 391–398. [CrossRef]

66. Maillard, É.; McConkey, B.G.; Angers, D.A. Increased uncertainty in soil carbon stock measurement with spatial scale and sampling profile depth in world grasslands: A systematic analysis. *Agric. Ecosyst. Environ.* **2017**, *236*, 268–276. [CrossRef]

67. Lawrence, P.G.; Roper, W.; Morris, T.F.; Guillard, K. Guiding soil sampling strategies using classical and spatial statistics: A review. *Agron. J.* **2020**, *112*, 493–510. [CrossRef]

68. Lamichhane, S.; Kumar, L.; Wilson, B. Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma* **2019**, *352*, 395–413. [CrossRef]

69. McBratney, A.B.; Mendonça Santos, M.L.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, *117*, 3–52. [CrossRef]

70. Potash, E.; Guan, K.; Margenot, A.; Lee, D.; DeLucia, E.; Wang, S.; Jang, C. How to estimate soil organic carbon stocks of agricultural fields? Perspectives using ex-ante evaluation. *Geoderma* **2022**, *411*, 115693. [CrossRef]

71. Kaufman, L.; Rousseeuw, P.J. *Finding Groups in Data*; WILEY: New York, NY, USA, 2009.
72. Aynekulu, E.; Vagen, T.; Shephard, K.; Winowiecki, L. *A Protocol for Modeling, Measurement and Monitoring Soil Carbon Stocks in Agricultural Landscapes*; World Agroforestry Centre: Nairobi, Kenya, 2011.
73. Bradford, M.A.; Carey, C.J.; Atwood, L.; Bossio, D.; Fenichel, E.P.; Gennet, S.; Fargione, J.; Fisher, J.R.B.; Fuller, E.; Kane, D.A.; et al. Soil carbon science for policy and practice. *Nat. Sustain.* **2019**, *2*, 1070–1072. [CrossRef]
74. Davies, L.L.; Uchitel, K.; Ruple, J. Understanding barriers to commercial-scale carbon capture and sequestration in the United States: An empirical assessment. *Energy Policy* **2013**, *59*, 745–761. [CrossRef]