Verified Carbon Standard

Draft VCS Guidance Document

VM0042

SOIL SAMPLING AND ANALYSIS HANDBOOK

Draft Version 1.0

Draft for public consultation: 2 March 2026

Sectoral Scope 14

Version 1.0 of this document was prepared by Agricarbon (Helaina Black, Daniela Cuba, Shibu Ebrahim, Annie Leeson and Alan Strong), Seqana (Ahmad Awad, Julian Kremers, Tobias Horstmann) and Verra. Seqana developed Sections 2 (Core Concepts) and 3 (Sampling Design), while Agricarbon developed Sections 4 (Field Sampling) and 5 (Laboratory Analysis). Further technical contributions were provided by Jaclyn Kachelmeyer (formerly Terra Carbon, Perennial), Dan Kane (formerly Terra Carbon), Jim Kellner (Perennial) and Alexandre Wadoux (INRAE).

CONTENTS

1 INTRODUCTION AND PURPOSE

This handbook is a guidance document for projects quantifying soil organic carbon (SOC) stocks and stock changes as part of a VCS project. It focuses on the implementation of *VM0042 Improved Agricultural Land Management*. The presented guidance on soil sampling and analysis to plan and execute monitoring strategies is generally applicable to other relevant methodologies.

The statistically grounded design of sampling strategies is fundamental for minimizing errors when quantifying SOC stock changes and deriving emission reductions and removals. Following strict standard operating procedures and avoiding common sources of error and bias in field and laboratory procedures ensures robust monitoring.

This handbook defines key terms and concepts, provides illustrative examples, and outlines optionality and recommendations for decisions around sampling design, field and laboratory protocols. The purpose of this document is to provide guidance to project proponents to expand on the requirements defined in the applied methodology with a focus on *VM0042*. During project validation and verification, it is not required that validation and verification bodies (VVBs) assess projects against this project document. It is however, a valid resource to evaluate the robustness of applied soil sampling and analysis strategies and plans in a VCS project.

2 Core Concepts and Key Terms

2.1 Quantification approaches

Soil sampling is required in VCS projects focusing on quantifying SOC stock changes. The quantification of SOC stock changes may involve the use of models. Under *VM0042*, both eligible quantification approaches for SOC stock changes require soil sampling in different modalities:

Quantification Approach 1 (QA1) - *Measure and Model:* Requires soil sampling to estimate SOC stocks at project start date (t_0 ± 5 years), and at minimum every five years thereafter to:

1. Initialize the applied biogeochemical model
2. Re-estimate model prediction error, potentially re-calibrate and (re-)validate the applied biogeochemical model

Quantification Approach 2 (QA2) - *Measure and Remeasure*: Requires soil sampling at project start date and at every verification event as direct measurement is the basis to quantify SOC stock changes.

Table 1. Quantification Approaches and Sampling Purposes

Quantification Approach	QA1 – *Measure and Model*	QA2 – *Measure and Remeasure*
Purpose of soil sampling	Provide model input data at t_0 and at least every five years after initial sampling via re-measurement.	Direct quantification of SOC stock changes based on sampling at t_0 and prior to any verification in baseline control sites and the project area.
Type of sampling	Paired sampling.	Independent or paired sampling.
Sampling requirements	Sampling at t_0 for model initialization must provide sufficient unbiased estimates of SOC content or SOC stock mean and variance in different strata within the quantification unit to initialize the model and appropriately assess model output uncertainty. Sampling at remeasurement events to re-estimate model prediction error and/or recalibrate the applied biogeochemical model must provide sufficient unbiased paired samples to enable re-calibration and re-validation of models based on sampled data.	The results from soil sampling must be unbiased and sufficient in size to distinguish spatial variability from real change (reduce uncertainty/variance of the estimates).

2.2 Soil sampling, statistical sample and soil sample

Soil sampling is the planning and executing of the collection of a subset (sample) of the soil in the defined project area to estimate soil properties (e.g. mean/total SOC stock).

A (statistical) **sample** is a subset of the target population. The term **soil sample** refers to a single portion of soil extracted from a land area, e.g., soil core, that can be handled and analyzed. This is distinct from the statistical concept of a *sample*. The term *sample* is merely a subset of numbers that are part of a larger total set of numbers (population) that are the numerical values of a soil property (target variable). The goal of sampling is to make estimates with a specific precision and associated uncertainty about the larger total set (population) derived from the smaller subset (sample).

2.3 Sampling unit

The sampling unit represents an individual of the target population that is eligible for sampling and may be selected as part of the sample. The sum of all sampling units is equal to the entire population (e.g. the complete soil of the project area to the target depth). Sampling units are non-overlapping, such that each element of the population belongs to one and only one sampling unit (Cochran, 1977). In SOC measurement, reporting and verification (MRV) the sampling unit may be defined as an individual point or a defined area, depending on the sampling design.

1. **Individual Point:** A specific geographic location[1] represented by a pair of longitude and latitude coordinates. The individual point can be representative of a single soil core or a physical composite sample, where soil from multiple cores within a radius is physically combined into a composite sample for a single laboratory analysis. Both are eventually geographically represented as individual points.
2. **Area:** A designated area within which multiple soil cores are collected. These cores are analyzed individually in the laboratory and then the results are statistically aggregated (yielding a mean and standard error over the area). Commonly used sampling units for soil sampling are plots of 10x10m or 20x20m or entire management units (i.e., crop fields).

2.4 Independent and paired sampling

In a VCS project focused on SOC, the objective is to quantify changes in carbon stocks due to implementation of project activity over a given crediting period. SOC content and soil mass can only be directly measured at single points in time. To derive change over time, the quantified SOC stocks from different points in time need to be subtracted. This subtraction can be carried out with the population level estimates or at the individual element level (sampling unit), when elements from two samples over time can be paired. In the former approach, the two samples from the populations to be subtracted are considered to be independent; in the latter the samples are paired.

[1] As a longitude/latitude coordinate pair (coordinate reference system WGS84) measured with a GPS device with a defined and documented precision. If other coordinate reference systems are used, they must be reported with their corresponding EPSG-code, see https://spatialreference.org/.

Paired Sampling aims to retain the same sampling unit for sample collection across initial and all remeasurement events. When paired sampling is employed, the spatial variance of SOC stock change is the relevant target variable for statistical analyses, not the spatial variance of SOC stocks at a single point in time.

Independent Sampling allows the sample units to be randomly re-distributed at every re-measurement events. There is no expectation of returning to the same locations of the initial sampling when resampling. For independent sampling, statistical calculations require the spatial variance of SOC stocks measured at each individual point in a given year, rather than the variance of SOC stock change between two points in time (e.g., between t_0 and t_1).

2.5 Probability sampling

Probability sampling is a type of random sampling where every sampling unit in the population has a known strictly positive probability of being sampled at random (i.e. >0). In the context of SOC MRV, this approach is critical to avoid bias and ensure that the results accurately reflect the variability of SOC stocks across the entire project area. By using probability sampling, project developers can produce scientifically credible and objectively verifiable estimates of SOC changes over time.

2.6 Multistage sampling

Multistage sampling, also known as hierarchical, cluster, or nested sampling, is a structured approach for obtaining representative soil samples where the project area into two or more stages of sampling units. The lowest level corresponds to point/area of the chosen sampling type[2] where soil samples are collected. Higher-level units are clusters defined by the project proponent, such as regions, farms, or fields. At each sampling stage, a random selection of sampling units is performed to ensure representativeness. However, this process generally increases uncertainties and hence sample size requirements.

In the context of *VM0042*, the project area may be subdivided into quantification units, or the entire project area may represent a single quantification unit. To apply the requirement of stratified random sampling, the entire quantification unit is stratified, and samples could be taken from anywhere within the quantification unit. When applying a multistage sampling strategy, each quantification unit is subdivided into clusters before applying stratification and randomly distribute the sampling units.

[2] Sampling type can be either a) individual sample, b) physical composite sample or c) statistical aggregate sample, see Section 3.5

2.7 Target population, target variable, and variance

The **target population** (population) is the complete collection of observations that could be made about a target variable (Lohr, 2010). In the context of SOC MRV, this typically refers to the collection of all the measurable SOC stock (changes) within the project area.

The **target variable** is the quantity of interest. In a VCS project quantifying SOC stock changes, this refers to SOC stock[3] at a particular point in time (snapshot, formally referred to as SOC_t and $SOC_{t+\Delta t}$) or SOC stock changes over time ($\Delta SOC_{t,\Delta t}$). It may also represent the difference between SOC stock changes in the project scenario and a counterfactual dynamic baseline scenario of SOC stock changes over time ($\Delta SOC_{pc,t,\Delta t} - \Delta SOC_{bsl,t,\Delta t}$). Additionally, when model-based estimates need to be evaluated, the target variable can be the model prediction error of the specific model used on its respective target variable.

In the context of VM0042, the target variable is different depending on the Quantification Approach applied as shown in Table 2.

Table 2. Quantification approaches and corresponding target variables.

Quantification approach and stage	Target variable	Variance of the target variable
QA1 - model initialisation	SOC stock snapshot	Spatial variance of SOC stock at a single point in time
QA1 - model true-up	Model prediction error of the biogeochemical model	Spatial variance of the model prediction error of the biogeochemical model
QA2 - paired sampling	SOC stock change	Variance of SOC stock changes
QA2 - independent sampling	SOC stock snapshots	Spatial variances of SOC stock at two points in time

2.8 Snapshot and change

This handbook distinguishes between *snapshot* quantities and *change* quantities for the target variable (e.g., SOC stock, model prediction, or model prediction error).

A snapshot is the value of a target variable at a single point in time *t* for a defined population, stratum, or sampling unit. Examples include:

- SOC stock (tC/ha) in 0–30 cm at time *t* at a specific sampling unit, or
- a model prediction error of a biogeochemical model at time *t*.

[3] The target variable SOC stock itself refers to a product of the variables SOC content, bulk density, fine earth ratio or the equivalent soil mass as detailed in Section 6.

When this handbook refers to a SOC stock snapshot, snapshot variance, or snapshot prediction, it always means the value of the target variable at a single point in time, and the spatial variability of those values across sampling units, not a difference between two times.

A change is the difference between two snapshots of the same target variable for the same population, stratum, or sampling unit, taken at two points in time, t and $t + \Delta t$.

When this handbook refers to SOC stock change, variance of change, or change predictions, it always means a temporal difference over a defined monitoring period (e.g., between initial sampling at t_0 and a first remeasurement at t_1).

- Under paired sampling, the change is calculated for each paired sampling unit (e.g., point, or field) as the difference between its two snapshots; the relevant variance is then the variance of these unit-level changes across space.
- Under independent sampling, change is obtained by subtracting two independent snapshot estimates (e.g., mean SOC stock at $t + \Delta t$ minus mean SOC stock at t); in this case the variance is calculated as the sum of the independent snapshot variances at each point in time..

The same snapshot–change terminology applies when the target variable is a model-based quantity (e.g., SOC stock predictions or model prediction errors) rather than a directly measured SOC stock.

2.9 Net SOC stock change

Net SOC stock change is a second-order target variable expressing the difference between two SOC stock changes over the same monitoring period (Δt), one under the project scenario and one under the baseline scenario. Formally:

$$NSC_{t,\Delta t} = \Delta SOC_{wp,t,\Delta t} - \Delta SOC_{bsl,t,\Delta t}$$

NSC represents the additional SOC stock change attributable to the project, i.e. the net climate-relevant SOC effect over the monitoring period.

When this handbook mentions net SOC stock change, variance of the net SOC stock change, or model-based net SOC stock change, it always refers to:

- the difference between two SOC change quantities (project and baseline) defined over the same time interval; and
- the associated uncertainty of this second-order difference.

Net SOC stock change is obtained by subtracting independently estimated SOC changes for project and baseline; its variance is calculated as the sum of the variances of the two change estimates, unless a modeled correlation structure is specified. Determining a relevant sample size often requires ex-ante estimates of the net SOC stock change. In this document, such ex-ante estimates of net SOC stock change are referred to as the expected net SOC stock change.

This definition of net SOC stock change applies whether the underlying target variable is measured SOC stock or a model-based quantity (e.g. predicted SOC change or prediction error).

2.10 Variance in SOC stock change quantification

The variance is defined as the square of the standard deviation. It is a measure of the variability of the target variable. In the context of SOC MRV, there are three relevant target variables associated with different variances:

1. **Spatial variance of SOC stock snapshots**: The variability of SOC stocks measured across different locations at a single point in time. This variance is used for statistical calculations when independent (unpaired) sampling is conducted. It also plays a role in initializing biogeochemical models. The precision of the SOC stock estimate is influenced by this variance.
2. **Variance of SOC stock change**: The spatial variability of the change in SOC stocks between two points in time. This variance is used for statistical calculations when paired sampling (i.e., repeated sampling at the same locations) is performed.
3. **Spatial variance of model prediction error**: The variability of model prediction errors across different locations specific to the type of model used (biogeochemical model or DSMs). This variance is relevant for the calculation of model prediction uncertainty.

Appendix 7.4.5 provides guidance on estimating these variances *ex-ante*.

Sample size requirements are driven by the variance in the target variable rather than the physical size of the quantification unit or project area. A higher variance indicates greater heterogeneity, which in turn requires a larger number of samples to accurately characterize the target variable. A large project area with lower variance of the target variable requires fewer samples than a smaller area with higher variance. Therefore, accurate *ex-ante* estimates of the variance may support effective sample size determination.

2.11 SOC strata and stratified random sampling

SOC strata refers to subgroups of a quantification unit (QU) based on the relative homogeneity of the target variable, in this context, SOC stocks. The SOC stocks within one stratum are more homogenous than the SOC stocks of different strata.

Stratification is the process of dividing an area into Strata.

Stratified Random Sampling describes the random allocation of sampling units across the entirety of each stratum (see section 3.1). It ensures that sampling of each strata is representative and unbiased.

2.12 Depth increments and equivalent soil mass

Equivalent Soil Mass (ESM) is a concept used to represent each depth increment by the *mass* of the *mineral soil* present in that layer at baseline, rather than by sampling to a fixed soil depth. This accounts for bulk density changes due to ALM practice changes by comparing the change in SOC stock for the same mass of soil at later remeasurement events.

Depth Increments are the contiguous segments of each soil core defined for soil sampling and analysis. Reporting of SOC stock changes in the VCS Program must occur on an equivalent soil mass (ESM) basis. It is required to sample at least two depth increments to enable accounting on an ESM basis.

2.13 Coupled and decoupled sampling for SOC content and mineral soil mass

Coupled sampling occurs when SOC content and mineral soil mass are measured from the same soil sample. In contrast, **Decoupled sampling** occurs when SOC content and mineral soil mass are determined from different soil samples. Coupled sampling generally reduces measurement error.

3 Sampling Design

3.1 Creating Efficient and Effective Sampling Designs

Developing a robust sampling design requires making critical assumptions grounded in reliable data and comprehensive background information about the project area, the quantification unit, the impact of the planned project activities and economic considerations associated with project management. The expected net SOC stock change and variance of the mean and variance of SOC stock change across the quantification unit and the project area guide sample size determination. Appendix 5.5 presents various data sources to derive *ex-ante* estimates of these target variables. Furthermore, economic considerations related to project management influence the budget available for soil sampling campaigns.

The sampling design should be driven by the goals of the sampling campaign, as these determine the target variable and population for the sampling design. Table 3 presents the different types and goals of soil sampling depending on the QA applied. The factors driving the variance relevant for sampling design and sample size are specific to each QA.

Table 3. Goals of sampling and driving factors to determine the sampling design and sample size

Quantification Approach	QA1 – *Measure and Model*		QA2 – *Measure and Remeasure*	
Type of sampling	Paired sampling.		Paired sampling	Independent sampling
Goals of sampling	Low model prediction uncertainty due to model initialization	Sufficient validation data for estimating model prediction error of the biogeochemical model at remeasurement (i.e: “true-up”)	Low SOC stock change estimate uncertainty	Low SOC stock snapshot estimate uncertainties
Driving factor/variance for sampling design and sample size	Model sensitivity to initialization + Variance of SOC stock snapshot	Variance of the model prediction errors of the biogeochemical model	Variance of SOC stock change	Variance of SOC stock snapshots

Every sampling design starts with defining the project area and potentially dividing it into quantification units. When applying VM0042 projects, stratified simple random sampling is the required type of sampling design. Sample sizes for stratified simple random sampling differ depending on which quantification approach (QA1/QA2) is chosen and whether DSMs are used or not. When planning remeasurement campaigns, several factors must be considered, including their implications for the initial sampling campaign (section 3.3). Before conducting sampling campaigns in the field, several critical decisions regarding the sampling specifications must be taken (Section 3.4). Finally, multi-stage sampling may be considered as an option for project proponents to balance the economics of sampling just a subset of the quantification unit with the increased uncertainty that this entails (section 3.5).

3.2 Stratified Simple Random Sampling Design

What is stratified simple random sampling?

Stratified simple random sampling is a sampling strategy where the population is divided into distinct subgroups, or strata, based on homogeneity of the target variable (e.g. SOC Stock). As a result, the variance of the target variable within each stratum is smaller than the variance across the entire population in the quantification unit, leading to more precise estimates of the overall population parameters mean and total (Cochran, 1977, Lohr, 2010, Thompson, 2012). By ensuring each stratum is proportionally (to its size and/or variance) represented in the sample, stratified simple random sampling reduces bias and enhances the precision of estimations. The process of defining strata is called stratification.

Why stratified simple random sampling?

Unlike grid sampling, which can introduce biases due to its systematic nature, stratified simple random sampling is unbiased and ensures that the variance in the population is appropriately captured (Lohr, 2010). Unlike simple random sampling, where the target variable's variability may be underrepresented, stratified simple random sampling addresses this variability by explicitly ensuring that different subgroups (strata) are represented, leading to more accurate and reliable estimates (Cochran, 1977). Furthermore, stratified simple random sampling tends to require fewer samples than simple random sampling to achieve the same level of precision, a phenomenon known as the design effect (Kish, 1965).

Stratification and design effect

The concept of the design effect (deff) is crucial for improving sampling efficiency. By stratifying the population, the variance of the estimate of the target variable (e.g., mean SOC stock) is reduced. This can either reduce the sample size required to achieve the same precision (ex-ante) or allow for more precise estimates without increasing sample size (ex-post). The magnitude of the design effect depends on how well the strata reflect the variability in the target variable. Formally, the design effect is given by the following equation[4]:

$$deff_{stratification} = \frac{VAR_{Stratified}}{VAR_{Simple}} \approx \frac{\sum_{h=1}^{H} \frac{Ah}{A} S_h^2}{S^2} \approx \frac{\sum_{h=1}^{H} \frac{Ah}{A} S_h^2}{\sum_{h=1}^{H} \frac{Ah}{A} \left[S_h^2 + \left(\underline{\Delta \cdot_h} - \underline{\Delta \cdot} \right)^2 \right]}$$

Where:

$deff_{stratification}$	The design effect due to stratification

[4] Refer to Brus, 2023 and Lohr, 2010

$VAR_{Stratified}$	The variance of the stratified simple random sampling estimator
VAR_{Simple}	The variance of the simple random sampling estimator
A_h	The area of stratum h
A	The total area of the quantification unit
S_h^2	The sample variance of the stratum
S^2	The simple random sampling sample variance for the quantification unit
$\underline{\Delta \cdot}_h$	The mean estimate of the target variable in the stratum
$\underline{\Delta \cdot}$	The weighted mean estimate across all the strata

The real design effect can only be determined after the samples have been taken (*ex-post*), so it is recommended to calculate sample sizes without considering the design effect initially and rather benefit from lower uncertainty deductions *ex-post*. This approach ensures that sufficient samples are taken, even if the design effect is overestimated during the planning phase.

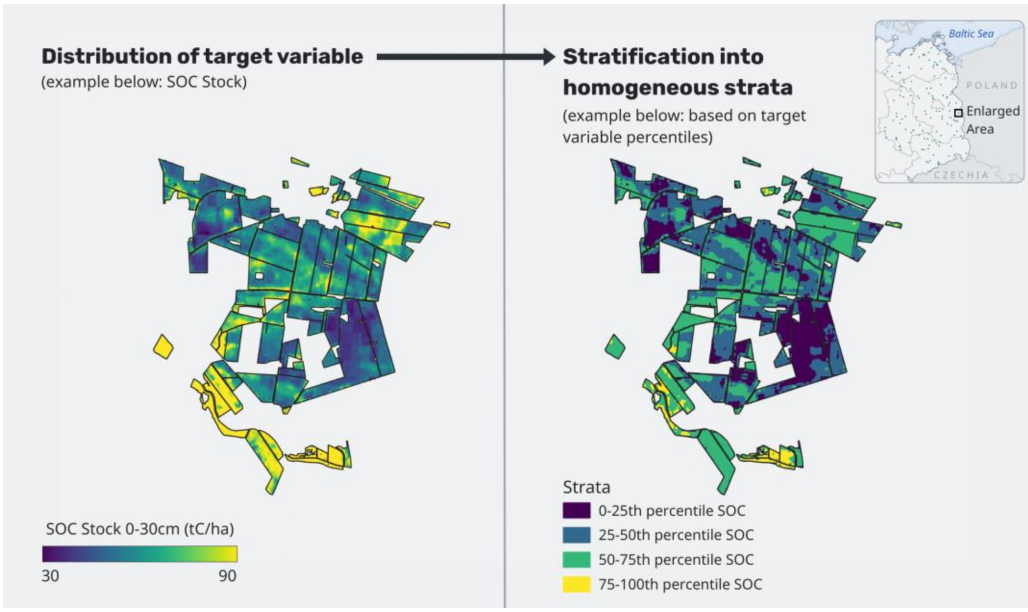
How to stratify?

The process of stratification involves geographically dividing the quantification unit based on the homogeneity of the target variable (e.g., SOC stock or expected SOC stock change). In general, prior to sampling the actual distribution of SOC stock is only known based on publicly available soil maps (e.g., SoilGrids) or can be estimated based on biophysical variables that correlate with SOC stock distribution.

Stratification based on SOC stock heterogeneity and expected SOC stock change

Whenever precise estimates of SOC stock distribution, i.e., the target variable itself are available, it is recommended to use them for stratification. Figure 1 provides an example of this simple stratification. Stratification based on the target variable itself typically results in better design effects and more efficient sampling. However, the resolution of these estimates of the target variable plays a critical role. For instance, estimates with a resolution greater than 100 meters may not be suitable for agricultural applications due to the potential inclusion of surrounding land features that do not accurately reflect the target variable's variability within agricultural fields (Croft et al. 2012).

Figure 1. Example of stratification based on 4 percentiles of SOC stock distribution



Stratification based on environmental covariates
Stratification may be based on proxies correlated with the target variable. These are often referred to as covariates. For SOC stocks, relevant proxies include:

- field boundaries,
- soil texture,
- climate,
- topography (slope, elevation, curvature),
- historical or current land use,
- management history proxies (e.g., manure application)
- vegetation,
- parent material, and
- soil type.

For SOC stock change, additional proxies related to differences in land management practice changes that will be introduced, such as cover crops and tillage practice, should be considered as well.

Stratification based on combined information on SOC stock distribution and environmental covariates
Strata may be defined using a combination of spatial SOC predictions (e.g., SoilGrids) and environmental covariates. This is recommended best practice because of the combined information leverage of available SOC maps that capture broad landscape heterogeneity patterns and environmental covariates that explain sub-field variability. This stratification approach often results in strong variance reduction.

The process for stratification combining these information sources may be:

a) Rule-based, e.g., by combining SOC stock classes (low/medium/high) with slope (flat/steep), 6 strata are produced.

b) Multivariate clustering–based, e.g., applying a suitable clustering algorithm (e.g., k-means[5]) is then applied to group spatial units into a predefined number of clusters, each of which defines a sampling stratum.

Consideration of Baseline Control Sites under QA2 for Stratifications

All baseline control sites for one quantification unit need to be representative to match this quantification unit and must be stratified for sampling using either:

1. the same stratification approach as in the quantification unit **or**
2. baseline control sites must be selected to meet the requirements of the applied methodology (e.g., in *VM0042* they must be paired with individual strata from the each quantification unit in line with the similarity criteria requirements for baseline control sites).

3.2.1 Sample Allocation and Placement to Strata

Sample Allocation

Once the QU is stratified and the overall sample size is determined, samples are allocated to each stratum. In the absence of prior information on the variability of the target variable within strata, *proportional allocation* is recommended, distributing samples in proportion to the size of each stratum. If reliable estimates of within-stratum variability are available, *Neyman allocation* offers a more statistically efficient alternative. This method assigns more samples to strata with greater variability, enhancing the precision of overall estimates without increasing the total sample size. The choice between proportional and Neyman allocation should be based on the availability and reliability of information on the target variable's variability. Table 4 compares the two approaches.

Table 4. Comparison of proportional and Neyman sample allocation

Sample Allocation Type	Benefits	Parameters	Formula
Proportional Allocation	Requires minimum information	A_h and n	$n_h = n \times \frac{A_h}{\sum_h^H \quad (A_i)}$
Neyman Allocation	Can minimize sample size but requires a priori information on Strata variances	A_h, s_h and n	$n_h = n \times \frac{A_h \times \sqrt{s_h^2}}{\sum_h^H \quad \left(A_h \times \sqrt{s_h^2}\right)}$

The parameters listed in Table 4 are:

1. A_h is the area of stratum h
2. n is the number of samples in the project
3. h refers to a specific stratum

[5] See Potash et al. (2023) for an example of applying k-means to create geographically compact strata that minimized variability within clusters, supporting more efficient sampling for SOC stock estimation.

4. H is the total number of Strata
5. s_h^2 is the variance of stratum h

Sample Placement
After the right number of sampling units has been allocated to their respective Strata, the placement of these sampling units within the Strata must consider the following:

1) Sampling units must be placed using a random selection method, and the procedure used to ensure randomness must be clearly documented and demonstrable.
2) There must be a defined buffer of 10 m between the samples and artefacts, such as field boundaries, roads, buildings, walking paths, etc.

3.2.2 Improving stratification over the course of a project

While strata must remain consistent within a single monitoring period, project proponents are encouraged—both for compliance and improved precision—to refine stratification ahead of each remeasurement sampling campaign. There are several reasons why updating stratification is beneficial:

- Changes to quantification units: Quantification units may evolve over time, either voluntarily (e.g., through merging, see Section 3.2) or involuntarily (e.g., due to evolving project area by addition of project activity instances). When such changes occur, the associated stratification must be revised accordingly.
- Improved understanding of the target variable or proxies: Initial sampling campaigns generate valuable data that can be used to refine stratification in subsequent campaigns, resulting in more precise and representative sampling. In addition, ongoing improvements in modeling capabilities can lead to more precise *ex-ante* estimates of the target variable, further enhancing the effectiveness of stratification over time.

The process of improving stratification must consider if the project designed sampling with independent or paired sampling:

- Independent sampling: there is no need to maintain consistency with previous strata or sampling units. A new stratification or updated stratification can be developed for each campaign.
- Paired sampling: Previous sampling units must be revisited, and sample allocation must adhere to proportional or Neyman allocation rules (see Section 3.4.4). In some cases, it may be necessary to:
    - Discontinue certain previously sampled units, resulting in the loss of pairing for those units in the overall sample, or
    - Introduce new units that can only be paired in future campaigns

Despite these challenges, improving stratification under paired sampling remains a valuable endeavor. Further guidance on Stratification and sampling strategies is found in Aynekulu et al. (2011), FAO (2019), de Gruijter et al. (2006), de Gruijter et al. (2016), Hengl et al. (2003), ISO (2018, p. 18), Maillard et al. (2017), Mudge et al. (2020), and Vanguelova et al. (2016).

3.3 Sample Size Determination

The required sample size depends primarily on the purpose of the sampling campaign, which corresponds to the chosen quantification approach. It is also influenced by the specific target variable, which varies based on both the quantification approach, DSM usage and the sampling strategy employed.

Appendix 7.1 presents online available sample size calculators focusing on different approaches based on minimum detectable difference or the applicable uncertainty deduction for *VM0042* projects.

The following factors influence the determination of sample size and must be estimated *ex-ante*—before the sampling campaign begins:

1. Expected SOC stock change (tC/ha):
   The ability to distinguish real SOC stock change rate depends on the relation between expected SOC stock change and variability of SOC stocks in a given area. When natural soil variability is much larger than the annual SOC stock increase as a result of implementing the project activity, a high number of samples are required to confidently detect change.
   The spatial heterogeneity of soils and temporal SOC dynamics unrelated to management are often relatively large compared to annual SOC stock change rates.

2. Magnitude and type of variance:
   Higher variance in the target variable results in higher number of samples required to represent the spatial variability and enable the detectability of changes over time. The relationship is linear, but the slope depends on the variance type[6]:
   a. *SOC stock change variance* (requires paired sampling), with an impact of a factor of 1. Example: A twofold increase in variance of change results in a twofold increase of the required sample size.
   b. *SOC stock snapshot variance* (applicable to independent sampling and paired sampling), which involves a factor of 2. Example: A twofold increase in spatial variance results in a fourfold increase in required sample size.
   c. For model-based estimations (DSMs or biogeochemical models), the *variance of the model prediction error* of the respective model is the variance that influences sample size requirements.

3. Uncertainty deduction rules:
   Lower sample sizes generally lead to higher uncertainty deductions. The detailed implications are determined by the applied methodology.

---

[6] Note that there are several sources of variance that contribute to either of these types of variance, such as small-distance variability in both SOC stocks and SOC stock changes, sampling errors (e.g. incomplete cores), compositing errors (incomplete homogenization), measurement uncertainties (which should be provided by lab analysis providers). Eventually all these sources manifest in the overall variance of the sample. The appropriate probability sample is assumed to be unbiased if potential sources of bias are minimized and accordingly documented.

4. Design effect:
   Design effects can improve sampling efficiency by reducing the variance of the estimate of mean SOC stocks. This can be leveraged either *ex-ante* to reduce sample size assumptions or *ex-post* to reduce uncertainty after sampling. Design effects are achieved through:
   a. *Stratification*: Effective stratification can substantially reduce required sample sizes compared to simple random sampling. The magnitude depends on the quality of stratification (see section 3.2 with the respective formula).
   b. Model-assisted *ex-post* estimator: Additional data sources can be used to refine the estimates from the samples with a model-assisted estimation (see Appendix 6.2 for more details)
   c. Timing of design effect application: To avoid underestimating sample size due to overly optimistic design effect assumptions *ex-ante*, it is recommended to calculate sample sizes initially without applying a design effect. The design effect can then reduce uncertainty deductions ex-post, once sampling data is available.
5. Applied biogeochemical models or DSMs:
   The structure of biogeochemical models may condition higher sample sizes depending on the modality of SOC inputs and the sensitivity of the applied model to SOC input uncertainty. The remeasurement rules and requirements of the applied methodology impacts the required sample size. See further guidance in Appendix 7.2.

3.3.1 Initial and true-up sampling considerations QA1 in *VM0042*

Under QA1, soil sampling serves two purposes:

1. **Model initialisation:** provide sufficient unbiased estimates of SOC stock mean and variance in different domains within the quantification unit to initialize the model and appropriately assess model output uncertainty.
2. **Model true-up:** provide sufficient unbiased paired samples to enable re-calibration and re-validation of models based on sampled data.

Since the model true-up is required for QA1, the initial sampling units are paired with the sampling units at the true-up, so that the model can be initialized on measured ground observations for specific sampling units (locations) and the model predictions for these sampling units at true-up time can be used with remeasured ground observations to calculate the project-specific model prediction error of the biogeochemical model.

The most important criterion at model true-up is to show that the model-based estimates are conservative in comparison with the measured data from the project site. Since the baseline scenario is modelled and not measured, no model verification is possible on project-related measured baseline data.

Initial sample size considerations

Before the biogeochemical model used in QA1 is validated with data from within the project area (i.e., model validation after the first remeasurement), project proponents may estimate appropriate sample sizes under QA2 requirements (see Section 3.2.2), even if QA1 is planned. This approach allows quantifying SOC stock changes within the project area with design-unbiased estimates and allows project proponents to switch from QA1 to QA2 at a later project stage.

This approach also acknowledges and mitigates the risks that the validation of the biogeochemical model may:

- **fail** after sample remeasurement is carried out or
- result in **high uncertainties** following the true-up of the biogeochemical model.

Sample size determination after the first remeasurement (3rd sampling campaign onwards)

Once the biogeochemical model has been successfully validated using data from within the project area, the sample size for subsequent sampling campaigns may be reduced. The reduced sample size must still meet the requirements for both QA1 purposes: 1) (re-) initialization of the biogeochemical model, and 2) true-up validation (and re-calibration). The model true-up test is mandatory based on the latest version of the *VCS Methodology Requirements*.

For some biogeochemical models that are highly sensitive to initial SOC stocks, an optimal sample size may be higher than the sample size required for biogeochemical model true-up to reduce uncertainty deductions. This depends on the specific biogeochemical model used in the project, the overall SOC stock variance at project start and the sensitivity of the model to its initialization.

Sample size requirements for biogeochemical model true-up

The relevant sample size criterion that must be passed at true-up is that the model-based estimates do not lead to over-crediting. It must be shown with a Welch's test that the model bias on the practice change scenario is smaller or equal to the model bias on the baseline scenario with a tolerable error or minimum detectable difference of the magnitude of the uncertainty deductions applied in the project until this point. This ensures that the bias of the final model-based estimate (practice change - baseline) never exceeds the applied uncertainty deduction and hence does not lead to overcrediting. Since the prescribed Welch's test must pass with a minimum power of 95% (equivalent to a 5% error rate of accepting $\underline{BIAS}_{wp,t,\Delta t} \leq \underline{BIAS}_{bsl,t,\Delta t}$ even though it is false), the sample size for the test must be large enough to achieve this statistical power. The number of sampling units $n_{wp,t,\Delta t}$ on the project site that is re-measured for Model true-up must be chosen such that the power of the test described above is at least 95%. Power at an ex-ante assumed sample size *n* is estimated prior to sampling using the following power equation:

$$power(n_{wp,t,\Delta t}) = 1 - F_{NCT}\left(t_{\alpha(1),\nu}\ ;\ \nu, \lambda\right)$$

with:

$$\lambda = \frac{UD_{t,\Delta t-1}}{\sqrt{\frac{s^2(\epsilon_{wp,t,\Delta t})}{n_{wp,t,\Delta t}} + \frac{s^2(\epsilon_{bsl,t,\Delta t})}{n_{bsl,t,\Delta t}}}}$$

$$\lambda = \frac{UD_{t,\Delta t-1}}{\sqrt{\frac{s^2_{pooled}(\epsilon_{VMD0053,bsl})}{n_{wp,t,\Delta t}} + \frac{s^2_{pooled}(\epsilon_{VMD0053,bsl})}{n_{VMD0053,bsl}}}}$$

$$UD_{t,\Delta t-1} = \sqrt{s^2_{\underline{\hat{\Delta}}\cdot t,\Delta t-1}} \times t_{\alpha(1)=0.667}$$

Where:

$power(n_{wp,t,\Delta t})$	=	The anticipated power $(1-\beta)$ of the bias test if $n_{wp,t,\Delta t}$ sampling units are re-measured at the project site for model true-up.
$F_{NCT}\left(t_{\alpha(1),\nu}\ ;\ \nu, \lambda\right)$	=	The cumulative distribution function (CDF) of the **noncentral t-distribution** with ν degrees of freedom and noncentrality parameter λ, evaluated at the critical value $t_{\alpha(1),\nu}$.
λ	=	Non-centrality parameter of the noncentral t-distribution.
$UD_{t,\Delta t-1}$	=	Uncertainty deductions applied at the latest project verification event before the model true-up.
$s^2_{\underline{\hat{\Delta}}\cdot t,\Delta t-1}$	=	Variance of the mean ERR estimate from each GHG or C pool • at the latest project verification event before the model true-up.
$t_{\alpha(1)=0.667}$	=	Critical t-value of a one-sided 66.67 confidence interval.

Any number of samples $n_{wp,t,\Delta t}$ that yields $power(n_{wp,t,\Delta t}) > 0.95$ is allowed for Model true-up. The minimum required sample size can be obtained by iteratively computing $power(n)$ for increasing values of $n = 2, 3, \ldots$ until $power(n) > 0.95$. Appendix 7.1 includes an exemplary calculation.

It must be noted that it is a strict true-up requirement to show that the bias of model-based change estimates of practice change is smaller than the bias in baseline change estimates. If this requirement is not met at true-up, the model cannot be used for model-based estimations in the context of verification events and credit issuance until it is successfully recalibrated and revalidated as per *VMD0053* and passes the bias test on the data from within the quantification unit.

To avoid failing the bias test because of insufficient sample size, it is advised to err on the side of caution when estimating the model prediction error variance, which means to rather overestimate this variance.

Buffering Sample Size for Inaccessible Sampling Units at Remeasurement

For model re-calibration and validation, the critical factor is the number of paired samples, i.e., samples taken at the same sampling unit during both the initial- and remeasurement campaigns. Therefore, it is recommended to anticipate and estimate the number of sampling units that may become inaccessible at remeasurement due to factors such as farmer turnover, land use changes, or adverse weather conditions. To address this, a buffer should be added to the initial sample size to compensate for potential inaccessibility at remeasurement. This buffer ensures that a sufficient number of paired samples will be available to support robust model re-calibration and validation.

3.4 Remeasurement Considerations

There are three major considerations when planning remeasurement:

1. Results of initial sampling at t=0
2. Changes to the quantification unit (i.e: loss or additions of farms/fields between t=0 and $t+\Delta t$)
3. Inaccessibility to sample units for paired sampling

The **first consideration for planning remeasurement** is the **new information provided by the initial sampling event** at t=0, which can be investigated in a post-hoc analysis of the data. In the post-hoc analysis, project proponents can review the following:

- Re-calculate the variability of SOC stocks. If the first sampling event over- or underestimated SOC stock variance, it is possible to recalculate the number of samples at t_1.
    - Practical consideration for *paired sampling* only: Any initial sampling unit that is not re-sampled during the remeasurement campaign is effectively lost as a paired sample for detecting SOC changes. Similarly, any new sampling unit added at remeasurement that was not part of the baseline campaign cannot be used as a paired sample for the initial quantification period.
- Re-assess Stratification and optimize sample size, allocation, and placement. As outlined in Section 3.2, refining stratification over time is recommended, as it can lead to more favorable design effects and improved precision in estimating the target variable. This, in turn, reduces the sample size required for future remeasurements.
- QA1: Project proponents must investigate the effect on the true-up requirements for the biogeochemical model if fewer paired samples are taken.

The **second consideration for planning remeasurement** is to adjust for potential **changes to the quantification unit**, e.g. loss of farms/fields or adding new farms/fields. Over each distinct monitoring period the quantification unit must remain stable. When parts of the quantification unit leave the project during the monitoring period, the quantification for that monitoring period must retrospectively be adjusted at the next verification event. This has ramifications for the sampling design and estimation, as previously sampled sampling units might not be part of the population anymore. At the time of resampling, it must be demonstrated that the sampling design has been updated appropriately and that any statistical implications have been addressed. To mitigate such issues, project proponents are strongly encouraged to consider and plan for this scenario during the initial sampling design phase.

The **third consideration for planning remeasurement**, relevant for **paired sampling,** any **potential inaccessibility** of sampling units during the baseline campaign must be accounted for when designing the sample size and sampling strategy. This ensures that the targeted number of paired samples can be achieved in future resampling campaigns. For example, if 1,000 paired samples are targeted and 10% of sampling units were inaccessible during the baseline campaign, this 10%, plus any anticipated changes in the project area (see point two above), should be included as a buffer in the sample size for remeasurement. This approach helps guarantee that sufficient paired samples will be available for future model re-calibrations and validations after the second remeasurement campaign.

3.5 Defining Specification of Sampling

Sampling specifications significantly influence uncertainty deductions, overall project costs and risks. Key design dimensions include:

1. Independent vs. Paired Sampling
2. Individual Samples vs. Statistical Aggregates vs. Physical Composites
3. Depth, Depth Increments, and SOC Stock Calculation
4. Coupled vs. Decoupled Measurement of SOC content and Soil Mass

There are infinite possible combinations across these four dimensions. Combinations that are scientifically robust typically result in lower uncertainties and, therefore, lower uncertainty deductions, but often at higher sampling and laboratory costs. Conversely, lower-cost approaches may lead to higher uncertainty deductions due to weaker scientific rigor.

Selecting the Right Sampling Specification

Project proponents are encouraged to select sampling specifications that uphold the highest standards of scientific rigor. However, recognizing that this may not always be economically feasible, it is often necessary to balance scientific rigor with economic efficiency. Project proponents are encouraged to quantify both the uncertainty deductions and sampling costs for each feasible sampling specification. The optimal approach is the one that minimizes the sum of uncertainty deductions and sampling costs for the given project. Appendix 7.3 discusses further considerations for paired sampling.

Table 5. Sampling specifications

Sampling specification	Individual sample	(Statistical) Aggregate sample	(Physical) Composite sample		
Sampling specification options	1: Individual sample single point	2: Multiple individual samples statistically aggregated over an area	3: Individual composite sample from sub-samples	4: Multiple composite samples statistically aggregated per area	5: Field level composite

Sampling specification	Individual sample	(Statistical) Aggregate sample	(Physical) Composite sample		
Sampling unit	Individual point (single core)	Areal aggregate (e.g. field level or 10m*10m square)	Individual point (multiple cores within radius of e.g. 10m)	Areal aggregate (mostly field level)	Field
Requires multistaged sampling?	No	No	No	Yes	Yes
Compatibility	Independent sampling and paired sampling[7]				
QA applicability	QA1 & QA2				
Available statistics	Mean, standard deviation (stdv)		Composite mean, stdv between composites	Composite mean; field mean; stdv between composites, stdv for management unit	Mean per management unit, stdv between management units
Considerations	Enables calculation of all relevant statistics to assess uncertainties Results in lower uncertainties and reduced uncertainty deductions (assuming equal sample sizes) Involves higher laboratory costs Often superior to composites, leading to reduced uncertainty deductions		Does not offer insights into SOC stock variance between the sub-samples of the composite, leading to unknown uncertainty; conservative estimates of this uncertainty is propagated8 Introduces additional uncertainty due to the compositing process which must be accounted for propagated Reduces laboratory costs by requiring fewer samples for analysis		

[7] Paired sampling refers to the sampling unit which has to be paired. Within a sampling unit, independent samples can be taken. For example, if the sampling unit is a field, the fields measured in t_0 are re-measured in t_1, but the locations sampled within the field may be independent. Therefore, they can differ between t_0 and t_1.

[8] Uncertainty propagation follows a similar approach as with mutli-stage sampling (see Appendix 7.6 on multi-stage sampling): accounting for both the within composite sub-sample variance and the between variance of the composites.

Variations within each of the 5 options described in Table 5:
Each main option includes several design variables that impact both cost and uncertainty:

- Number of cores per statistical aggregate or composite (approach 2-5 above). Higher numbers of cores reduce uncertainty deductions but at the same time increase costs.
- Sampling depth and increments
- Coupled vs. decoupled measurement of SOC content and soil mass and subsequent ESM calculations.
- Size of aggregation area (approach 2 and 4)

More rigorous sub-options generally lower uncertainty, but increase cost.

Depth and Depth Increment Considerations:
Sampling depth and increment selection should be aligned with the anticipated depth of SOC Removals and Reductions based on land use, management practices, and project objectives. The reporting depth for SOC stock changes is defined at the methodology level. Key considerations include:

- *Minimum depth of 30cm:* In the VCS program, SOC stock changes must be reported down to 30 cm.
- *Management influence depth*: Projects should align depth with the likely depth for SOC change due to ALM practices (e.g., tillage depth, rooting depth, manure incorporation).
- *Expected SOC dynamics*: Deeper increments (>30 cm) may be justified for practices like agroforestry or deep rooting systems.
- *Standard benchmarks*: Sampling depth intervals may be defined to enable consistency across reporting systems.
- *Economic Optimum Depth*: While deeper sampling enables the detection of SOC Removals and Reductions at greater depths, particularly relevant for practices like agroforestry or deep-rooted perennials, SOC stock variance (of change) typically increases with depth. This leads to higher sample size requirements to maintain the same level of precision and uncertainty deductions. Projects should carefully evaluate this tradeoff and aim to determine an economic optimum depth: the depth at which the expected additional sequestration quantified justifies the increased sampling costs or uncertainty deductions. This analysis can be informed by modeling the relationship between expected SOC sequestration and SOC stock variance (of change) at different depth increments, and evaluating how these factors influence downstream sampling costs and uncertainty deductions.

3.6 (Optional) Multistage Sampling Design

Multistage sampling—also known as hierarchical, cluster, or nested sampling—is a structured approach for obtaining representative soil samples without the need to sample the entire project area. The project area is divided into two or more stages of sampling units within the Quantification Unit (QU). The lowest level corresponds to point/area of the chosen sampling type[9] where soil samples are collected. Higher-

[9] Sampling type can be either 1: individual sample, 2: physical composite sample or 3: statistical aggregate sample, see Section 3.5

level units are clusters defined by the project proponent, such as regions, farms, or fields. At each sampling stage, a random selection of sampling units is performed to ensure representativeness.

Examples:

- **Two-stage SOC sampling design:** In a two-stage design, fields may serve as the primary sampling units (PSUs). The project proponent randomly selects a proportion of fields (e.g., 30%) across the QU, using a probability proportional to each field's size. Within these selected fields, soil sampling points form the secondary sampling units (SSUs), which are then randomly selected for soil collection.
- **Three-stage SOC sampling design:** In a three-stage design, provinces could be defined as the primary sampling units (PSUs), while fields within those provinces could serve as the secondary sampling units (SSUs). The project proponent first randomly selects a proportion of provinces (e.g., 30%), using probabilities proportional to each province's size. Next, within the selected provinces, a proportion of fields (e.g., 40%) is randomly selected, again using probabilities proportional to each field's size. Finally, soil sampling points within these selected fields form the tertiary sampling units (TSUs), which are randomly chosen for soil collection.

Benefits of multistage sampling are that the approach allows sampling of a subset of clusters (e.g., fields), reducing the number of clusters that must be visited. It also enables higher sampling density within selected clusters, potentially improving local data resolution.

Disadvantages of multistage sampling are:

a) introduction of additional statistical uncertainty, requiring a larger overall sample size compared to sampling without stages
b) complexity to the design and verification of the sampling strategy

Sampling Units must be randomly selected at every stage and will still be distributed across the entire quantification unit. For example, if fields are used as a sampling unit, a random selection—weighted by area—of all fields must be sampled, *not* just those that are convenient or nearby. Multistage sampling results in multiple layers of sampling uncertainty added at every stage, leading to a higher required sample size compared to not using multistaged sampling. See appendix 7.5 for further details.

Multistage sampling is permitted under *VM0042* but is often misunderstood or misapplied. Its advantages, such as reducing the number of field visits, should be carefully weighed against the downsides of increased sample size and design complexity. As a result, project proponents should assess its suitability on a case-by-case basis. A detailed explanation is provided in Appendix 7.5.

4 FIELD SAMPLING

4.1 Key Considerations

The primary aim of soil sampling is to support the reliable measurement and monitoring of SOC stock changes and/or other soil properties over time. Therefore, the tools and techniques deployed in sampling must support representative, repeatable and reproducible sampling and/or analysis of soil to defined depths at initial measurement and remeasurement events.

Best practice in soil sampling includes a wide variety of tools and techniques which can be deployed for different requirements and environments. Therefore, there is no *one-size-fits-all* approach for soil sampling in the field that can cover distinct site- and project-specific conditions. Rather, tools and procedures should be selected to be *fit-for-purpose* for the specific conditions of the project and able to meet these requirements:

- Deep enough to sample a minimum depth including topsoil and subsoil -, i.e., ≥30 cm
- Supports reporting SOC stock changes on an Equivalent Soil Mass (ESM) basis
- Capable of sampling at, the minimum of, two depths increments
- Capable of sampling a consistent volume of soil from each depth increment
- Captures samples representing both coarse fragments (> 2mm) and fine earth fractions (< 2mm)
- Sufficient soil (mass, e.g., grams) sampled to enable laboratory analyses of the required soil properties
- Reproducible across repeated measurement events at different times and project areas
- Suitable for the practical and logistical aspects of field sampling and subsequent laboratory analysis
- Cost-efficient for the scope and scale of the project
- Measurement errors, from field sampling, sample preparation and laboratory analysis as relevant, are known and quantified to fulfill the uncertainty deduction requirements

Traditionally, soil sampling obtains a defined volume or mass of soil to support analytical determination of soil properties in a laboratory (direct measurement). It is preferable to determine SOC content, bulk density and soil mass on the same soil sample from the same location since this minimizes measurement error.

Subsections 4.2-4.4 address direct measurement, and Section 4.5 addresses alternative analysis using in-field sensor technologies.

4.2 Choosing Field Tools for Soil Sample Collection

Certain tools, equipment and methods will be better suited to specific soil conditions; projects should consider the physical characteristics of the soils in the project area, and baseline control sites where relevant, such as:

- stone content in soils,
- soil penetration resistance,
- soil depth (thickness of the soil layer)

There will also be practical and logistics considerations in selection which could include:

- equipment conditions and availability
- access to the project area
- capacity to maintain and repair tools
- costs of operating and renting equipment
- labor costs
- laboratory costs per sample
- time needed for sampling and analysis of samples
- cost of sample storage—if needed

Projects should use the same tools for sampling soils for SOC content, bulk density and soil mass since this minimizes measurement error and ensures a consistent sample size (n) for calculating SOC stocks. If separate tools are used for sampling properties e.g. SOC content and bulk density, the project should account for the related measurement errors.

Table 6 describes suitable tools, methods, and best practices for collecting soil samples for the determination of soil properties, with an emphasis on SOC content (SOC%), soil mass (dry weight) and bulk density. This table is not exhaustive but is intended as a description of the most common, practical sampling tools for VM0042 projects and key considerations for their use.

Table 6. Description and best practice recommendations of soil sampling field tools

Name	Description and best practice
Soil Probes and Corers	Hollow cylinders where the internal diameter determines the volume of sample collected while the corer lends itself to marking off defined depth increments.

Name	Description and best practice
	Amongst the most widely used sampling tools since they can be deployed effectively in most agricultural soil environments. Can be used to obtain individual soil samples. Probes are typically narrow diameter samplers (e.g. 5–30 mm) while corers are generally of a wider diameter (e.g. +30 mm). The core diameter must be wide enough to obtain a representative soil sample for the range of soil properties expected across the project. Wide diameter corers are recommended for bulk density and coarse fraction mass sampling. Small changes to the core diameter will introduce errors into mass and volume measurements and therefore measured core diameter should be recorded. Maintenance procedures should indicate the frequency of core diameter measurements and thresholds for coring bit replacements. Small diameter probes (i.e. ½ inch / 12.7mm) are not suitable for representative sampling of coarse fraction materials and therefore fine soil mass determinations. Risk of sample compression, or expansion, due to the sampling action of driving a cylinder into soil, can be mitigated by selecting suitable diameters for the soil type and good quality management protocols in the field. Some sample compression / expansion is unavoidable but can be mitigated through depth corrections using coring depth measurements taken in the field. The core is split into appropriate soil depth increments in the field or laboratory. It is recommended to split cores in the laboratory to reduce measurement error and to address core expansion / compression. Soil corers and probes exist in two main forms - manual and powered - which are described in the next two rows.
Powered probes or corers	Hydraulic jack systems, often vehicle-mounted, to assist with removing intact soil cores. Recommended if a project is sampling to greater depth (e.g. >30cm). Commonly used to obtain individual cores from georeferenced locations. The core is split into appropriate soil depth increments in the field or laboratory. If vehicle-mounted: impact on farmers' fields to access sampling locations should be considered.
Hand Operated Probes and Corers	Hand-operated push or hammer probes or short corers are typically used for manual collection of soil samples at shallow depths. In some soil types, sampling below 30cm with hand tools is challenging and increases the risks of sample compaction and loss

Name	Description and best practice
	Most commonly used for compositing samples across fields for agronomic purposes. However, if of sufficient diameter, they may be used to obtain individual samples
Soil Augers	A helical / spherical blade that cuts through the soil and a helical 'screw' shape that excavates the soil and carries to the soil surface. Particularly suitable for sampling soils where the soil conditions make it impractical to drive a soil probe or corer into the soil without significant compaction, or in situations where the soil has low cohesion. Include both manually operated hand tools as well as powered tools which are driven by hand-drills, presses or hydraulic systems. Where the auger diameter is sufficient, suitable for a wide suite of soil properties and can be used for both composite or individual soil samples. If used for soil mass / bulk density sampling, the auger diameter must be wide enough to obtain a representative sample of soil coarse materials The diameter of the actual auger hole and the sampling depths achieved must be recorded for accurate bulk density calculation. Small diameter augers, such as hand drills, can be used for sampling for SOC content but should not be used for fine soil mass or bulk density. Augers are open-sided and therefore complete and accurate recovery of soil samples at defined depth increments requires careful sampling procedures. When using a dutch auger, it is recommended that it be used with a sampling plate to ensure the full volume of soil is collected (see Vågen and Winowiecki 2023).

Other approaches (e.g., soil pits and horizontal bulk density rings) are also allowable, and practical considerations make them less common for sampling in VCS projects where large sample sizes are critical for accurate estimates of spatial variability in SOC stocks.

However, there may be situations in which it is a reasonable reason for not using probes, corers or augers. The use of these methods should be justified and supported by evidence ensuring that the use of procedures will produce unbiased and reproducible results at all remeasurement events, and that measurement error can be determined accurately. Beem-Miller et al. (2016) provide a useful approach to ensuring high-quality sampling in rocky agricultural soils

Clod Method: The clod method is suitable for measuring bulk density in surface and subsoils exposed in trenches. The typical sample volume for the clod (undisturbed lump of soil) method is between 400 and 700 cm^3. The clod method can vary, and some approaches involve coating the clod with a waterproof material, while others measure the clod's volume without coating. The specific method used should be clearly and accurately reported (Prihar and Hundal, 1971; Grossman and Reinsch, 2002; Antille et al., 2021.). The primary source of inaccuracy with the clod method is sampling bias caused by the propensity of clods to fracture along interpedal voids.

Excavation Pits Method: This method is typically applied in stony, gravelly, or structured soils where cores or clods are impractical. It involves digging out a large hole or pit to a target depth and collecting the entire soil mass from the hole in pre-determined depth increments, oven-drying the soil at 105 ºC, lining the hole with plastic film, and filling it completely with a measured volume of water or sand (Grossman and Reinsch, 2002). The excavation method requires significant time and labor, particularly when measuring the bulk density of subsoil. The excavation method is suitable for soils with gravel content >25%. It is only recommended in soils with very high (>25% by volume) coarse fragment content.

4.3 Field Procedures

Projects must use consistent field procedures for every sampling event. Projects must ensure that these are documented clearly, including which tools are used (and why), and how they are used in the field. The following, when included in these procedures, will support best practice and quality management in field sampling at repeated sampling events.

4.3.1 Preparation

1) **Sampling timing windows** chosen to minimize the influence of seasonal agricultural management practices and/or schedules (e.g. machinery operations, manure applications, liming or tillage) on SOC stock measurements
2) **Operating protocols** must provide clear instructions to field personnel
3) **Personnel** carrying out field sampling must have suitable expertise and / or training to deploy the tools and carry out the procedures as required
4) Suitable means to **communicate**[10] correct sampling locations and specifications to the core extraction personnel
5) Agreed **permissions** from landowners and establish transparent and open communication with landowners

4.3.2 Field Work

1) Clear **processes** to secure field records, ensuring that data can be accurately linked to soil samples and their respective sampling locations.
2) Suitable **equipment** to record **GPS or** other Global Navigation Satellite Systems **(GNSS)** that provide positioning, navigation, and timing; projects must record both the planned GPS coordinates and the 'actual' GPS coordinates in case the original point can't be sampled (and the reason why it was moved or missed). GPS precision is *especially* important when doing paired sampling, see appendix 7.3.

[10] For example, providing translation, if necessary.

3) Guidance must be specified to direct if/how Sample Points can be moved in the field, if a planned **sampling point is inaccessible,** in a way that doesn't compromise the sampling design. One possible approach is to specify alternative locations to be used if the original point cannot be accessed. These rules must ensure that no bias is introduced.
4) Clear instructions on **when to abort sampling** due to poor field conditions such as waterlogged or frozen ground and poor weather conditions
5) **Deviations** from field procedures, including sampling dates and times, land use, accessibility issues, difficulties with sampling (e.g. weather conditions), changes to coring bits, etc., must be unbiased and justified
6) **Quality control of extracted cores and samples** must be in place (e.g. rules for: secure packaging and storage of individual soil cores or samples; discarding inadequate cores due to e.g. soil moisture, soil compaction and/or loss of soil). Measurement of the actual sampling depth should be recorded to enable adjustment for compression or expansion of sample core length.
7) **Quality checks** and maintenance of **field equipment** should be embedded in operating procedures e.g. checks on coring bit diameter, field sampling depth recording, cleaning of equipment between sampling points).

4.3.3 Post-field operations

1) **Records** for registering by whom, where and when field sampling was carried out
2) **Deviations** from field procedures e.g. sampling dates and times, land use, land cover, accessibility issues, difficulties with sampling (e.g. weather conditions, soil moisture, erosion), changes to coring bits, etc. must be unbiased and justified in a report

This list is not exhaustive, and projects should identify where there may be risks to SOC stock measurements from project conditions, and to mitigate those risks. The project should ensure that documented field procedures are retained for each sampling event and review previous field procedures at each remeasurement event to ensure consistency in sampling over time.

4.4 Packaging, Storage and Transport

Handling of soil cores and samples prior to analysis must avoid cross-contamination and minimize the potential for changes in soil properties to occur. The following requirements must be met:

1) Soil samples should be **packaged** in **cleaned and secured** air-tight bags, containers or cores to minimize loss of soil material in transport
2) All packaging should be **permanently labelled** with unique identification codes that will enable tracking of individual soil samples from field through laboratory analysis to data reports and SOC stock calculations
3) **Where intact cores are transported**, they should be stacked in a way that reduces compression or expansion in transit, and avoids cross-contamination
4) **When cores are split** into individual depth increments or composite samples prior to transport, then care should be taken not to lose or mix soil material while cutting and transferring soil to

individual sample packaging, e.g. by using clean collection trays or buckets and tools for each processing step, e.g, by cleaning with water and/or thorough brushing between use.

5) The depth **splitting procedure**, whether in the field or at the laboratory, should account for compression / expansion that occurs during field sampling and transport using the actual field depth sampled e.g . comparing the soil elevation inside the cylinder with the original soil surface outside the cylinder (FAO, 2023).
6) Sub-sampling of composite samples should be carried out in the laboratory to reduce measurement error.

Soil samples must be **shipped within five days** of the completion of a sampling campaign. Until then, samples may be temporarily stored on-site in a location protected from sunlight, humidity, and precipitation, avoiding mixing of the different soil materials and any other product (e.g., chemical or organic agronomic products).

4.5 In-Field or In-Vitro Sensor Technologies

There is growing interest in the application of sensor technologies (proximal sensing) for the determination of soil properties *in-situ*. The following proximal sensing techniques are allowed: infrared spectroscopy, including near infrared (NIR), visible near infrared (Vis-NIR), and mid-infrared spectroscopy (MIR); laser-induced breakdown spectroscopy (LIBS); and inelastic neutron scattering (INS, also known as neutron-stimulated gamma ray analysis or spectroscopy). Detailed criteria for evaluating the use of proximal sensors (IR spectroscopy, LIBS, and INS) must be adhered to as defined in the applied methodology.

The choice between direct and proximal measurement will depend upon the specific requirements of the project including, for example:

- the range of soil properties to be determined
- the level of accuracy needed
- the ability to detect change over time
- the feasibility of deploying the methods in the field environment

Whilst direct measurement is preferred for its accuracy and capacity to detect change in SOC stocks, there may be circumstances where proximal measurement could provide a useful alternative e.g. when direct measurement is not feasible or practical (England & Viscarra Rossel, 2018).

When using proximal measurement tools, it is important to understand the limitations of the methods and have a clear understanding of how the measurement error will be determined (in combination with laboratory analyses for calibration, if appropriate) for use in uncertainty deduction calculations.

5 LABORATORY ANALYSIS

5.1 Soil Properties and Analyses

There are few soil properties that must be reported from laboratory analytical procedures to provide the data necessary for the calculation of SOC stocks by fixed depth SOC stocks and Equivalent Soil Mass. These are listed in Table 7 with standard and alternative methods used by laboratories.

The determination of fine earth fraction soil bulk density is different from the standard laboratory determination of soil bulk density. Standard soil bulk density is widely used for general soil compaction assessments (determined using total soil mass and volume) but it is fine earth fraction bulk density that is required for SOC stock calculations. Fine soil mass is also required for ESM calculations; note this is not a common laboratory determination.

Table 7. Soil properties to be reported from field sampling and laboratory analyses of soil samples. FD = fixed depth SOC stock calculations; ESM = SOC stock and change calculations by Equivalent Soil Mass (optional depending on ESM method applied).

Soil Property	Unit	Use	Standard Laboratory Method	Sources of Measurement Error
Total mass of soil	g	(FD) (ESM)	Balance weight; oven-dried or corrected for gravimetric moisture content	Moisture content; Balance accuracy
Mass of coarse fraction >2mm	g	FD ESM	2 mm sieving; balance weight; oven-dried or corrected for gravimetric water content	Moisture content; Balance accuracy; Incomplete sieving;
Mass of fine earth fraction < 2mm (Fine Soil Mass)	g	ESM FD	2 mm sieving; Balance weight; oven-dried or corrected for gravimetric water content	Moisture content; Balance accuracy; Incomplete sieving;
Soil organic carbon content	% or g/kg	ESM FD	oven-dried samples; Elemental analysis (DUMAS-C); proximal sensor alternatives	Analyser accuracy; Moisture content; Sub-sampling; inorganic carbon content

Soil Property	Unit	Use	Standard Laboratory Method	Sources of Measurement Error
Area of sample	cm^2	ESM FD	Calculate from diameter of field sampling tool	Damage to sampling tool
Sampling depth(s)	cm	FD ESM	Measurements from the field	Compression or expansion of sample
Sample volume	cm^3	FD (ESM)	Calculate from area of sample and sampling depth	Damage to sampling tool; Sample loss; Inaccurate field measurements
Fine earth fraction bulk density	g/cm^3	FD (ESM)	Derive from mass and volume of fine earth fraction and gravimetric moisture content	See fine fraction mass and sample volume
Soil organic matter content (SOM%)	% or g/kg	ESM	Loss-on-ignition (LOI) or SOC content to SOM content conversion using standard function	Oven temperature and duration; Accuracy of SOC:SOM function
Mineral mass of soil < 2 mm	g	ESM	Measure via combustion analysis, or derive as fine earth fraction dry mass minus SOM	Oven temperature and duration or see fine earth fraction and SOM

Projects may include additional analyses for modelling or other purposes such as texture, clay content, pH, etc. These can be determined from the same soil sample when there is sufficient soil mass available. Practical consideration: Projects proponents following QA1 (measure and model) should carefully review the input parameters required for their biogeochemical model or other models to determine what soil properties may require to be analyzed in the lab in addition to the soil properties listed in Table 7.

5.1.1 Selecting a Laboratory and Quality Assurance

Accreditation: Projects should use an analytical laboratory that is ISO/IEC 17025 accredited or equivalent, for the specific analysis process they are providing (e.g. Elemental Analysis) via established national and/or international accreditation body.

Alignment with Standards: The methods and procedures must be documented and detailed in line with ISO requirements and with specific details of anywhere those differ from ISO Standard Methods.

Quality Audits: The laboratory should carry out regular audits to report on quality assurance, including internal quality control procedures, for example the use of soil reference materials, testing documentation according to quality cards, monitoring of variation in analysis (repeatability tests) and error thresholds.

Quality Performance: Evidence of laboratory analytical quality performance evaluation should be provided by participation in inter-laboratory trials or ring-testing (e.g., through participation in the North American Proficiency Testing program or WEPAL) or participation in the Global Soil Laboratory Network scheme (GLOSOLAN). This information must be made available to the project and contribute to laboratory error analysis.

Continuity: All soil samples collected should be analyzed using the same methods in the same laboratory across the project lifetime to minimize measurement errors and discrepancies that can occur in different laboratories (Even et al., 2025). If a project changes the analytical laboratory, method or equipment within or between measurement or remeasurement events, the project must demonstrate that changes do not bias SOC stock or change data in the monitoring report.

Chain of Custody: Laboratories should have labelling and sample tracking systems that maintain the integrity of soil sample identification from sample to data reports. Laboratories should retain soil sub-samples until data reports have passed quality control procedures and there is assurance that no rerun analyses are required. Archiving several sub-samples from each measurement event can aid in quality assurance across extended time periods, particularly if a project needs to change laboratories and/or methods between resampling events.

Laboratory Error: The analytical laboratory should quantify and report analytical measurement error statistics to project proponents on a regular basis. These results will be required by projects for use in uncertainty calculations and may be produced from audit reporting under ISO ISO/IEC 17025 for accredited labs. Otherwise, FAO (2019) provides examples of the different statistics that laboratories can use to report analytical measurement errors.

Archived samples: Analytical laboratories or projects should retain a proportion of soil or soil samples (e.g. 5%) to enable reanalysis at repeat sampling events to ensure that results are comparable across all sampling events.

5.2 Sample Preparation

5.2.1 Storage at Laboratory Facilities

On arrival at the laboratory and prior to sample processing, samples should be stored under environmental conditions that minimize soil biological activity e.g. cool dry environments, refrigeration, but not frozen.

The duration of storage before analysis should not exceed 3 months. Soil samples may be air-dried and then stored for longer periods prior to laboratory analysis. Where projects retain archived soil samples, these should be stored under secure cool, dry and dark conditions.

5.2.2 Storage at Laboratory Facilities

The following steps are common to the laboratory analyses required by projects:

1. **Sample Recovery.** Packaged *soil sample or whole core*
    - Choose appropriate tools that minimize the risks of losing soil material in processing.
    - Clean tools and equipment between samples
    - Ensure that all samples can be traced through laboratory analyses via unique sample IDs
    - When processing packaged soil samples ensure that all soil material is transferred to the next step.
    - When processing intact cores, carefully separate or cut Depth Increments to the required lengths, accounting for any compression or expansion that has taken place during sampling and/or transport to the laboratory.
    - All subsequent steps should be carried out on each soil sample separately.

2. **Sample sieving.** Sieving prepares soil for drying and further analysis by removing coarse material (i.e. roots and rocks) from the sample
    - Sieve each sample to a final sieve size of 2mm. It is generally recommended to sieve field moist soils prior to drying, as it can be very difficult to sieve dry hard soils particularly if they have high clay content.
    - Care should be taken to completely remove all noticeable plant material from samples. Plant material has a high carbon content that can inflate measurements of SOC content.
    - Record the weight of the fine earth soil material (<2mm) from each sample
    - Record the weight of the coarse soil material (> 2mm) from each sample
    - These weights will be used for soil mass and bulk density calculations
    - Balances used should be regularly calibrated.

3. **Sample drying.** Samples should then be oven dried to remove water content (i.e. gravimetric water content).

- Standard practice is to dry soil samples at 105°C in a drying oven for 24 hours or until they are no longer losing mass. Samples can be dried at lower temperatures or air-dried at room temperature, but drying times are longer.
- Record the oven-dry weight of the soil sample (<2mm) and determine gravimetric water content.

Sample processing procedures must be reported in detail since effectiveness of soil processing will influence the measurement error associated with analytical results (Even et al., 2025). Procedures should remain consistent throughout the entire project lifetime. If there is a change in analytical laboratory, the project should demonstrate that there is no bias and that any changes to measurement errors are fully accounted for.

5.3 Determining Fine Earth Bulk Density and Soil Masses

Soil fine earth fraction bulk density (FE-BD) and soil mass (total, fine soil mass, coarse fraction) should be determined on the same soil sample using the same method and procedures. These soil properties require careful weighing of sequential fractions of soil from sieving with an adjustment for water content. While there is not an ISO standard for soil fine earth bulk density, laboratory methods for soil fine earth bulk density should reflect the guidance on equipment and procedures to produce reliable measurements of soil masses and water content which are detailed in ISO 11272:2017.

Soil fine earth bulk density and fine earth fraction are required for the standard calculation of SOC stock from fixed depth. While some ESM methods for calculating SOC stock or change do not require bulk density values, accurate measurements of soil mass and sample volume, adjusted for coarse fragments and SOM, are still required. The methods used to determine bulk density provide a practical and commonly used way to derive these properties, assuming a dry bulk density method is used (i.e. not a clod method). This section details how to process soil samples collected for bulk density and/or mass determinations.

1. **Sample Mass and Volume.** Once drying is complete, weigh the sample
    - Measure the following masses of each sample from the sequential steps in the sieving procedure using a calibrated balance with recommended precision of at least 0.01 grams:
        - Total soil mass (prior to sieving)
        - Coarse fraction (> 2mm)
        - Fine fraction (<2mm)
    - Measurement of mass at each step can reduce overall measurement error.
    - Determine the volume of the sample based on the internal dimensions of the sampling tool used. Note that for some tools, like augers, it may be better to instead determine the sample volume based on the size of hole made. In this case, those measurements must be collected in the field at the time of sampling.

- When processing packaged soil samples ensure that all soil material is transferred to the next step.

2. **Fine Earth Bulk density calculations.** Correct total dry soil mass and volume for coarse fragments to calculate bulk density on the fine earth fraction.
    - Use this formula to calculate fine earth bulk density:

    FEBD = Dry Fine Fraction Mass (g)/ volume of fine fraction in the soil sample (cm^3).

    Where:

    - Volume of fine fraction = volume of soil sample – volume of coarse material
    - Volume of coarse material = 2.65 x mass of coarse material

    - Standard practice is to use 2.65 g cm^{-3}, which is the density of quartz, but other values could be used if the specific type of rock in the soil is known.

A typical range for standard soil bulk density for surface, mineral agricultural soils is 1.1-1.7 g cm^{-3}. Values outside this range are possible but might indicate measurement error. Soils with highly organic soil layers may have much lower BD (<1 g cm^{-3}). Since published data are lacking on data for fine earth bulk density, projects should ensure that field and laboratory procedures are accurately measuring volumes and fraction masses to reduce measurement error and bias.

5.4 Analysis of Soil Organic Carbon Content

All soil samples must be analyzed for SOC content (SOC%). VM0042 specifies the approved methods for determining SOC content and outlines special considerations for each (e.g., error quantification requirements for soil spectroscopy). SOC content with known uncertainty should be measured using dry combustion elemental analysis (i.e., Dumas method). Walkley-Black (wet) oxidation and loss on ignition (LOI) are not recommended due to accuracy concerns but may be applied where no other method is available . Proximal sensing techniques are allowed if defined in the applied methodology.

The following considerations reduce errors when analysing SOC content:

1. **Sample processing.** Substantial error can be introduced into SOC content analyses at sample processing and homogenization.
    - Individual soil samples or composite samples must be sieved and homogenized prior to sub-sampling e.g. rotary sampling, cone-quartering.
    - Homogenisation, particularly for composited samples, should be carried out thoroughly to prevent bias being introduced into sub-sampling.
    - Any subsequent sampling of sub-samples e.g. after pulverising or for elemental analysis, must also use reliable sub-sampling techniques.
    - Homogenisation and sub-sampling introduce additional errors which must be accounted for.

- Sub-samples for SOC% analysis should be weighed into containers suitable for the elemental analyser, with sample weights recorded to enable calculation of SOC content.
- Sample weights should be maintained within a pre-defined range appropriate to the elemental analyser in use.
- Samples should be kept dry until the instrument analysis is completed. Reabsorption of water into the dried soil sample can increase the mass of the sample aliquot that is analyzed by the instrument.

2. **Removal of soil inorganic carbon.** Analyses for soil organic carbon must also correct for soil inorganic carbon (SIC, i.e. carbonates) where present.

- Projects should inform laboratories prior to sample processing if SIC could be present in soil samples and the laboratory should take action. This can be supported by checking local soil type maps, local soil data and/or local experts.
- In cases of uncertainty, a soil fizz test can be applied where small amounts of hydrochloric acid (HCl) are dripped onto a soil sample. If the sample produces a fizzing reaction, this indicates the presence of carbonates (i.e., SIC).
- Projects should have a plan to correct SOC content analyses accordingly. Two common approaches are generally used. For both SIC approaches, determine whether these procedures are applied to all samples or only a subsample. If only a subsample is analyzed, explain how the values from that subsample are applied to appropriately adjust all other SOC measurements.

Option 1 Pre-treat sample aliquots with an acid (typically HCl) to remove SIC prior to SOC analysis. This method assumes that SIC is fully removed before analysis, ensuring the SOC value is accurate. Methods for this approach are described in Bisutti et al., 2004; Nelson and Sommers, 1996.

Option 2 Determine the amount of SIC via the pressure calcimeter method. This approach similarly uses HCl to dissolve carbonates in the sample, but the process is done in a sealed vessel with a pressure gauge. As CO_2 off-gasses from the reaction of HCl and carbonates, the pressure inside the vessel increases, and the change in pressure can be used to estimate SIC content. Methods are described in Sherrod et al., 2002; Wagner et al., 1998; Loeppert and Suarez, 1996. Typically, this approach involves running total carbon analysis without pre-treatment and then using the SIC value to correct the total carbon estimate to SOC content.

3. **Elemental analysis.** The exact procedure for elemental analysis of SOC content will vary by analyser type and should follow the manufacturer's guidelines for SOC content analysis. Laboratories should maintain regular checks on analyzer performance which will include the use of Certified Reference Materials, Internal Reference Materials and sample blanks at regular intervals across sample batches. The analyzer should be calibrated regularly following manufacturer's guidelines using the relevant Certified Reference Materials. Accuracy and precision of laboratory analyses for SOC content can be

determined using results from repeat analyses of Certified Reference Materials and/or Internal Reference Materials (c.f. FAO, 2020)

6 EQUIVALENT SOIL MASS

6.1 Use of the Equivalent Soil Mass Method

Equivalent Soil Mass (ESM) is a mathematical method which is used as an alternative to fixed depth accounting of SOC stocks and changes. ESM accounts for bias in SOC stock and/or change calculations from fixed depth sampling due to changes in soil bulk density. By standardizing to an equivalent mineral mass of soil, the ESM method aims to ensure that changes in SOC stocks across sampling locations, soil depths or sampling events are due to actual carbon gain or loss, and not artefacts of physical shifts in the volumes of soil sampled. ESM is not appropriate for all circumstances, particularly where there may be significant soil erosion and where soils are sampled to full depth. The use of ESM should account for the specific errors associated with the application of predictive mathematical functions and conversions.

The following are specifically required to support the application of ESM methods

- Sampling at least 2 depth increments
- Determination of soil mineral mass from soil samples
- Establishment of Reference Soil Mass
- Application of a mathematical interpolation function

Projects should maintain documentation of the ESM methods applied at the initial and resampling events including the reporting of these specific requirements and any relevant calculation workbooks used to apply ESM. Projects should use the same ESM procedures at each resampling. Projects must provide documented justification for changing the ESM methods including the interpolation technique if this becomes necessary during the project lifetime. Projects that use publicly available and published ESM workbooks must ensure that these workbooks meet VM0042 requirements.

6.2 Soil Mineral Mass or Fine Soil Mass

When SOC stocks are calculated on an ESM basis, it is the mineral mass of soil or fine soil mass (i.e. the fine earth fraction dry mass minus soil organic matter content) that anchors the comparison across time and management interventions. ESM methods rely on the cumulative mineral mass of dry fine earth fraction soil per unit area, which requires accurate determination of dry soil mineral mass within each depth increment. The mineral soil mass can be determined from reported soil properties (see Table 7) using either of these two approaches:

- Mineral soil mass is calculated on an area basis for each depth increment using the cross-sectional area of the corer or auger e.g. Wendt & Hauser (2013). This procedure requires no additional bulk density sampling but still requires the determination of mineral mass, and therefore fine fraction mass and coarse fraction mass, from defined depth increments. If using the approach the projects must make modifications to the calculations provided in Wendt & Hauser (2013) to account for the coarse fraction.

- Mineral soil mass is calculated using fine earth fraction bulk density, increment length and SOC content for each depth increment, as outlined in von Haden et al. (2020), Peng et al. (2024), Fowler et al (2025), FAO (2020). These publications provide worked examples with some calculation worksheets available online or on request from the authors.

The following provides worked examples of the calculations necessary to determine soil mineral mass using either of these approaches. Projects must justify and document the approach used and the calculations applied.

Example Calculations: Mineral Soil Mass using Bulk Density

SOC stocks and mineral soil mass across the soil profile can be calculated using bulk density (corrected for coarse fragments and moisture) and SOC concentration as follows (Poeplau et al., 2017):

Step 1: Calculate total dry soil mass in each depth layer.

$$M_{n,dl,soil} = d_{n,dl} \times BD_{n,dl,corr} \times 100 \quad (1)$$

Where:

$M_{n,dl,soil}$	=	Total dry mass of soil sample *n* in depth layer *dl* (t/ha)
$d_{n,dl}$	=	Length of depth layer *dl* in sample *n* (cm)
$BD_{n,dl,corr}$	=	Bulk density of soil sample *n* in depth layer *dl,* corrected for coarse fraction and moisture (g/cm^3)
100	=	Conversion from g/cm^2 to t/ha

Step 2: Calculate total SOC mass in each depth layer.

$$M_{n,dl,SOC} = \frac{SOC_{n,dl}}{100} \times M_{n,dl,soil} \quad (2)$$

Where:

$M_{n,dl,SOC}$	=	Soil organic carbon mass of sample *n* in depth layer *dl* (t/ha)
$SOC_{n,dl}$	=	Soil organic carbon concentration of sample *n* in depth layer *dl* (%)
$M_{n,dl,soil}$	=	Total dry mass in soil sample *n* in depth layer *dl* (t/ha)
100	=	Conversion from % to decimal

Step 3: Calculate total SOM mass in each depth layer.

$$M_{n,dl,SOM} = \frac{SOM_{n,dl}}{100} \times M_{n,dl,soil} \tag{3}$$

Where:

$M_{n,dl,SOM}$	=	Soil organic matter mass of sample *n* in depth layer *dl* (t/ha)
$SOM_{n,dl}$	=	Soil organic matter concentration of sample *n* in depth layer *dl* (%)
$M_{n,dl,soil}$	=	Total dry mass in soil sample *n* in depth layer *dl* (t/ha)
100	=	Conversion from % to decimal

Soil organic matter mass is derived from soil organic matter concentration and the total soil mass. Where SOM concentration is not measured, it may be calculated from SOC concentration, using the SOC-to-SOM conversion factor of 0.58 g SOC / g SOM (von Haden et al. 2020).

$$SOM_{n,dl} = SOC_{n,dl} \div 0.58 \tag{4}$$

Where:

$SOM_{n,dl}$	=	Soil organic matter concentration of sample *n* in depth layer *dl* (%)
$SOC_{n,dl}$	=	Soil organic carbon concentration of sample *n* in depth layer *dl* (%)
0.58	=	van Bemmelen factor converting SOC to SOM

Step 5: Calculate total *mineral* soil mass in each depth layer.

$$M_{n,dl,mineral} = M_{n,dl,soil} - M_{n,dl,SOM} \tag{5}$$

Where:

$M_{n,dl,mineral}$	=	Mineral soil mass of sample *n* in depth layer *dl* (t/ha)
$M_{n,dl,soil}$	=	Total dry mass of soil sample *n* in depth layer *dl* (t/ha)
$M_{n,dl,SOM}$	=	Soil organic matter mass of sample *n* in depth layer *dl* (t/ha)

Example Calculations: Mineral Soil Mass using Composite Samples and Soil Mass

Wendt and Hauser, 2013, provide calculations for cases in which 1) soil mass—not bulk density—is measured in the lab and 2) composite soil samples are collected. In these cases, the total mineral soil mass per unit area can be calculated using the soil mass, diameter of the probes, and the number of cores collected and composited into a single sample. Wendt and Hauser do not perform these calculations on a mineral soil mass basis, so modification is required to comply with VM0042. Additionally, Wendt and Hauser (2013) do not correct for coarse fragments. Such corrections must be made by VM0042 projects.

Step 1: Calculate total dry soil mass per unit area in each depth layer based on the mass of the composite sample (measured in the lab, adjusted for SOM) and the cross-sectional area sampled across N cores using a probe of diameter D.

$$M_{n,dl,soil} = \frac{M_{n,dl,mineral}}{\pi \left(\frac{D}{2}\right)^2 \times N} \times 10{,}000 \qquad (1)$$

Where:

$M_{n,dl,soil}$	=	Total dry mass of soil sample *n* in depth layer *dl* (t/ha)
$M_{n,dl,mineral}$	=	Mineral soil mass corrected for coarse fragments of sample *n* in depth layer *dl* (g)
D	=	Inside diameter of probe or auger (mm)
N	=	Number of cores composited into sample *n* (unitless)
10,000	=	Conversion factor from g/mm^2 to t/ha

Step 2: Calculate total SOC mass in each depth layer.

$$M_{n,dl,SOC} = \frac{SOC_{n,dl}}{100} \times M_{n,dl,soil} \qquad (2)$$

Where:

$M_{n,dl,SOC}$	=	Soil organic carbon mass of sample *n* in depth layer *dl* (t/ha)
$SOC_{n,dl}$	=	Soil organic carbon concentration of sample *n* in depth layer *dl* (%)
$M_{n,dl,soil}$	=	Total dry mass in soil sample *n* in depth layer *dl* (t/ha)
100	=	Conversion from % to decimal

6.2.1 Reference Mineral Soil Mass

Establishing a reference soil mass is a key step in using Equivalent Soil Mass. The reference soil mass is a benchmark mineral mass of soil in a defined soil Depth Increment which is then used to calculate comparable SOC stocks and stock changes for the same equivalent soil mass across samples or locations over time, as illustrated in Figure 2. The mineral mass of soil samples from initial and / or resampling events is adjusted to match the reference mineral mass using a mathematical interpolation function. Reference Soil Mass can be determined in the following ways to calculate SOC stocks and changes to use in ESM.

Single Value Reference Soil Mass for a defined area

Reference Soil Mass is determined from the distribution of mineral mass of soil samples collected from a defined depth at the initial time point e.g. 0-30 cm at t_0. This single Reference Soil Mass is then used to convert all samples at t_0 and at all resampling events to an equivalent soil mass using an interpolation function. Reference Soil Mass at t_0 is generally set as the lowest soil mineral mass or from a low percentile point (e.g. 10%) to ensure that soil mass collected at subsequent resampling points is greater than this reference soil mass. SOC stocks at initial sampling and changes at resampling are reported according to this single Reference Soil Mass value e.g. 100 Mg /ha SOC reported to an equivalent mineral soil mass of 1000 Mg ha^{-1} at 0-30 cm (Reference Soil Mass = 1000 Mg ha^{-1}). The approach can produce consistent SOC change data for a defined area but does not allow direct comparison, or integration, of SOC stock or change data between areas or with other SOC stock or change data sources.

Reference Soil Mass applied at Resampling Events

Mineral mass of samples or locations at t_0 are used to adjust Fixed Depth calculations of SOC stock changes at resampling events to account for bulk density changes. The mineral mass from resampling events is adjusted to match the initial (reference) soil mass from that paired sample or location e.g. stratum. This approach will also involve interpolating the field data using a mathematical function, commonly a linear function e.g. Fowler et al 2025; FAO (2020). SOC stocks and changes at resampling are reported as fixed depth corrected by ESM e.g. 100 Mg/ha SOC at 0-30cm. SOC stocks at the initial sampling are reported at fixed depths with no ESM correction. The calculation approach is generally simpler than that for a single value reference soil mass approach although it can produce more variable SOC change data. Here data will be comparable and interoperable between projects and other data sources.

Projects should justify the selection of an appropriate Reference Soil Mass approach and document the calculations used to establish the Reference Soil Mass.

The following outlines steps in establishing a single value Reference Soil Mass in paired and independent sampling :

- Where Paired Sampling is used, in Quantification Approach 1 or 2, the single value reference mineral soil mass is determined for each sampling location based on t_0 measurements at the target sample depth (≥30 cm) within a project area, including the linked baseline control site(s) for Quantification Approach 2, where relevant.
- For Independent Sampling (only allowed for Quantification Approach 2), the single value reference mineral soil mass is determined as the average mineral soil mass at the target depth within the baseline control site at t_0 (von Haden et al., 2020). Here, SOC stocks in a project area are reported relative to the reference mineral soil mass of the linked baseline control site(s).

In both instances, the initial (t_0) SOC stocks for the project area, and baseline control sites if relevant, are determined using an interpolation function for the sampling points using the relevant Reference Soil Mass. At subsequent resampling events, SOC stocks are calculated for time *t*, using the same reference soil mass established at t_0 and the same interpolation function. The following provides examples of these calculations.

Example Calculations: Reference Mineral Soil Mass in Paired Sampling for Quantification Approach 1 and 2

In Paired Sampling, the reference mineral soil mass at t_0 is equivalent to the cumulative mass of mineral soil at each sample unit, calculated as:

$$M_{ref,n,min} = M_{n,cml,min} = \sum_{dl=1}^{dl} M_{n,dl,mineral} \tag{6}$$

Where:

$M_{ref,n,min}$	=	Reference mineral soil mass for sample *n* (t/ha)
$M_{n,cml,min}$	=	Cumulative mineral soil mass of sample *n* to the target depth at t_0 (t/ha)
$M_{n,dl,mineral}$	=	Mineral soil mass of sample *n* in depth layer *dl* (t/ha)

Since SOC stocks are monitored at these same units over time, initial SOC stocks are also calculated for each sample unit as:

$$M_{n,cml,SOC} = \sum_{dl=1}^{dl} M_{n,dl,SOC} \tag{7}$$

Where:

$M_{n,cml,SOC}$	=	Cumulative mass of SOC of sample *n* to the target depth at t_0 (t/ha)
$M_{n,dl,,SOC}$	=	Mass of SOC of sample *n* in depth layer *dl* (t/ha)

Example Calculations: Reference Mineral Soil Mass in Independent Sampling for Quantification Approach 2

In Independent Sampling designs, the reference mineral soil mass in a project area is the average mineral soil mass at the target depth across Sample Points in the baseline control site(s) linked to that project area, e.g.:

$$M_{ref,i,min} = \frac{1}{N_i} \sum_{n=1}^{N_i} \sum_{dl=1}^{dl} M_{n,dl,SOC} \tag{8}$$

Where:

$M_{ref,i,min}$	=	Reference mineral soil mass for project area *i* (t/ha)

$M_{n,dl,,SOC}$ = Mass of SOC of sample *n* in depth layer *dl* (t/ha)

N_i = 1, 2, 3 … N Sample Points in the baseline control site(s) linked to project area *i*.

dl = 1, 2, 3 … dl depth layers in sample point *n*.

6.2.2 Applying Interpolation Functions to calculate SOC stock and changes on an ESM basis

Interpolation functions are mathematical techniques which are used to predict a property, e.g. SOC stock, at a specified Reference Soil Mass. The functions use a mathematical relationship between SOC and soil mineral mass to support this determination, which is illustrated in Figure 2 from Wendt and Hauser (2013).

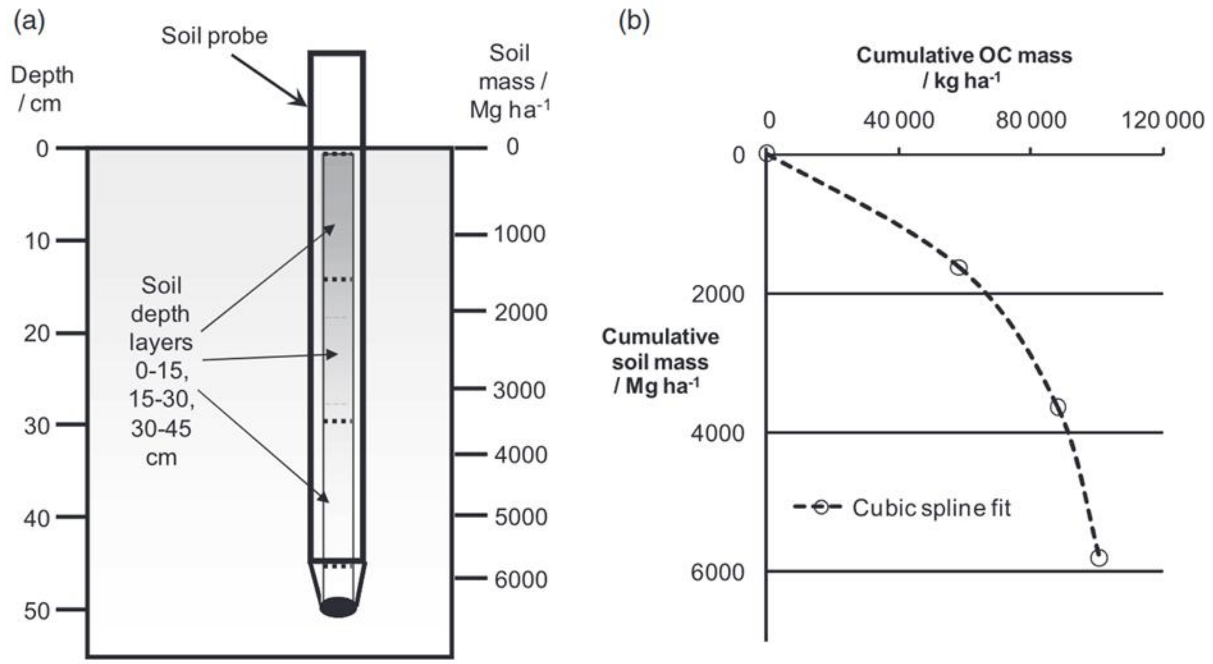
Whilst this ensures that SOC stock and change comparisons are made on an equal mass basis, eliminating error biases caused by different interpolation functions will result in different SOC stock and change values and have specific prediction errors to be accounted for. Projects should document the selection and application of interpolation functions, including prediction errors, along with relevant workbooks.

The following are the most common interpolation functions used in ESM:

- Linear interpolation (e.g. Fowler et al., 2023; GFAO GSOC).
- Cubic Spline interpolation (von Haden et al., 2020; Wendt & Hauser, 2013)
- Cumulative Coordinates Approach (Rovira et al., 2022)

The linear interpolation function must be used where there are only 2 sampling depths available (c.f. Fowler et al 2023). Linear, cubic spline or cumulative coordinates functions can be used when samples are collected in three or more depth increments (Wendt and Hauser 2013; von Haden et al 2020). The R script provided by von Haden et al., 2020, performs cubic spline interpolation and is suited to projects that measure bulk density and organic carbon at the same sampling points. Wendt and Hauser (2013) provide a workbook that automates the cubic spline interpolation calculations. However, project proponents must ensure that calculations are based on mineral soil mass and appropriately corrected for coarse soil fractions. Projects must use the same interpolation function and procedure across the project crediting period and account for calculation and prediction errors associated with the interpolation function used e.g. by using error propagation. As indicated previously, publications may provide worked examples with some providing calculation worksheets online or on request from the authors.

Figure 2. Illustration showing an estimation of SOC masses using a cubic spline fit. Derived from Wendt and Hauser (2013).



Projects should use linear interpolation techniques for 2 or more depth increments (c.f. Fowler et al 2023; Equations 5, 6 & 7) or alternatively cubic spline interpolation when samples are collected in three or more depth increments (Wendt and Hauser 2013; von Haden et al 2020).

Projects must use the same ESM interpolation procedure across the project crediting period and account for measurement errors associated with the specific ESM procedure used. Documentation of the ESM procedure applied, and associated workbooks, should be maintained to ensure that the same ESM procedure will be applied at each remeasurement event. Projects must ensure that publicly available and published ESM workbooks meet VM0042 requirements. Projects must provide documented justification for changing the interpolation technique, if this becomes necessary throughout the project lifetime.

A workbook developed by Wendt and Hauser (2013) automates the cubic spline interpolation calculations is available for download in the VM0042 webpage. However, project proponents must ensure that calculations are based on mineral soil mass and appropriately corrected for coarse soil fractions. This spreadsheet is best suited to projects that measure soil mass in the laboratory and base calculations on the cross-sectional area sampled

The R script provided by von Haden et al., 2020, performs cubic spline interpolation and is best suited to projects that measure bulk density and organic carbon coupled at individual sample points. VM0042 requirements and recommendations generally follow von Haden et al., 2020, so project proponents may find the R script more readily compatible.

7 APPENDICES

7.1 Sample size calculations

There are numerous approaches available for estimating sample sizes. In the context of the Voluntary Carbon Market (VCM), and specifically under *VM0042*, it is essential to choose a sample size estimation approach that ensures the reliable detection of change in SOC stocks.

Technical service providers may support sample size determination. This appendix outlines four different approaches to determine sample size. Seqana, for example, provides freely available online calculators to support sample size estimation applying different approaches illustrated in Figure 3:

1. Economic Optimum Number of Samples. Seqana's calculator available at: https://www.seqana.com/resources/eons-calculator

2. Minimum Detectable Difference. Seqana's calculator available at: https://www.seqana.com/resources/mdd-sample-size-calculator

3. Margin of Error at a Confidence Interval.

4. Model True-Up Based Sample Size Calculation. Seqana's calculator available at: https://www.seqana.com/resources/true-up-sample-size-calculator

Figure 3. Comparison of EONS, Minimum Detectable Difference (MDD), and Margin of Error (MoE) approaches

	1: EONS (Economic Optimum Number of Samples)	2: MDD (Minimum Detectable Difference)	3: Confidence Interval and Margin of Error
Question: How many samples are needed to…	… balance the *financial and scientific priorities* of the project, while being standard compliant?	… claim a *positive* SOC sequestration as statistically significant?	… estimate a SOC stock change with a given *Margin of Error* at a certain level of confidence?
Example Answer	At 543 samples, the marginal value of each additional sample does not pay off anymore. Below 50 samples you are not standard compliant.	With 690 samples, the *actual* SOC stock change will be correctly detected to be >0 tC/ha…[1]	With 267 samples, we are 95% confident that the *actual* SOC stock change lies within +-1 tC/ha around the *estimated* SOC stock change.
Statistical concept	Builds on uncertainty deductions, which in the case of VM0042 are based on a 1-sided 67% confidence interval	Hypothesis testing and power analysis	Confidence Interval with Margin of Errors

[1] … in more than 90% of the cases (=Power) when it at least 2.1 tC/ha (MDD) are sequestered, at a significance level of 5%.

7.1.1 Economic Optimum Number of Samples

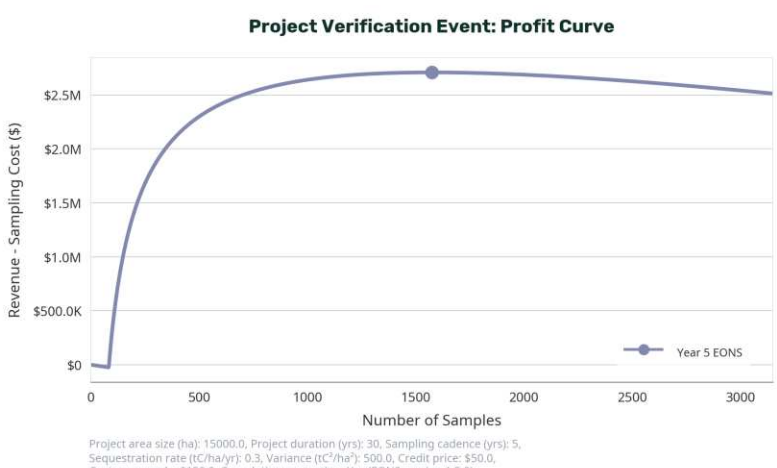
Seqana's concept of Economic Optimum Numer of Samples (EONS) is defined as the point where the **additional revenue from reduced uncertainty deductions** is no longer greater than the **additional cost of collecting more soil samples.** It represents the economically optimal balance between sampling effort and financial return.

Two key factors influence the statistical determination of EONS:

1. **Minimum Sample Size for Credit Eligibility**: The VCS Methodology Requirements state that the half-width of the two-sided 90% confidence interval must be smaller than the estimated net GHG benefit (in this context, the SOC stock change) over a 10-year period. If this criterion is not met, the project becomes ineligible for crediting, regardless of the sampling effort already undertaken. Failing to meet it effectively results in 100% uncertainty deductions, meaning no credits can be issued despite incurring costs. The EONS framework incorporates this requirements setting its fulfillment as the minimum sample size threshold for credit eligibility.
2. **Uncertainty deductions:** Seqana's EONS framework calculates the point where **net revenue** (creditable revenue minus sampling costs) is maximized and is based on how uncertainty deductions in the applicable methodology are calculated. The EONS tool assumes that a higher sampling size will increase sampling costs and simultaneously reduce uncertainty deductions.

   The net-revenue curve shown in Figure 4 represents a hypothetical scenario of a project's first verification event after five years. The calculation takes into account expected net SOC stock change, variance of SOC stocks and/or SOC stock changes, VCU prices, measurement uncertainty, a linear increase of sampling cost with increasing number of samples, size of project area, sampling cadence and project lifetime.

Figure 4. Relationship between sample size and net financial benefit (revenue after uncertainty deductions minus sampling costs) in QA2 after 5 years.



7.1.2 Minimum Detectable Difference

The Minimum Detectable Difference (MDD) sample size calculation approach is a type of power analysis based on the statistical concept of hypothesis testing. It is frequently used in research and referenced in previous versions of VM0042. It is used to calculate the minimum sample size at which a targeted statistical claim can be made.

Using the Minimum Detectable Difference (MDD) approach to calculate sample size helps answer a key question:

> "How many samples do I need in my quantification unit to demonstrate a statistically significant positive SOC sequestration?"

Example answer:

> "With 690 samples, the *actual* SOC stock change will be correctly detected as greater than 0 tC/ha in more than 90% of cases (i.e., statistical power), provided that at least 2 tC/ha (the defined MDD) has been sequestered. This result is valid at a 5% significance level."

Key Inputs for MDD-Based Sample Size Calculation

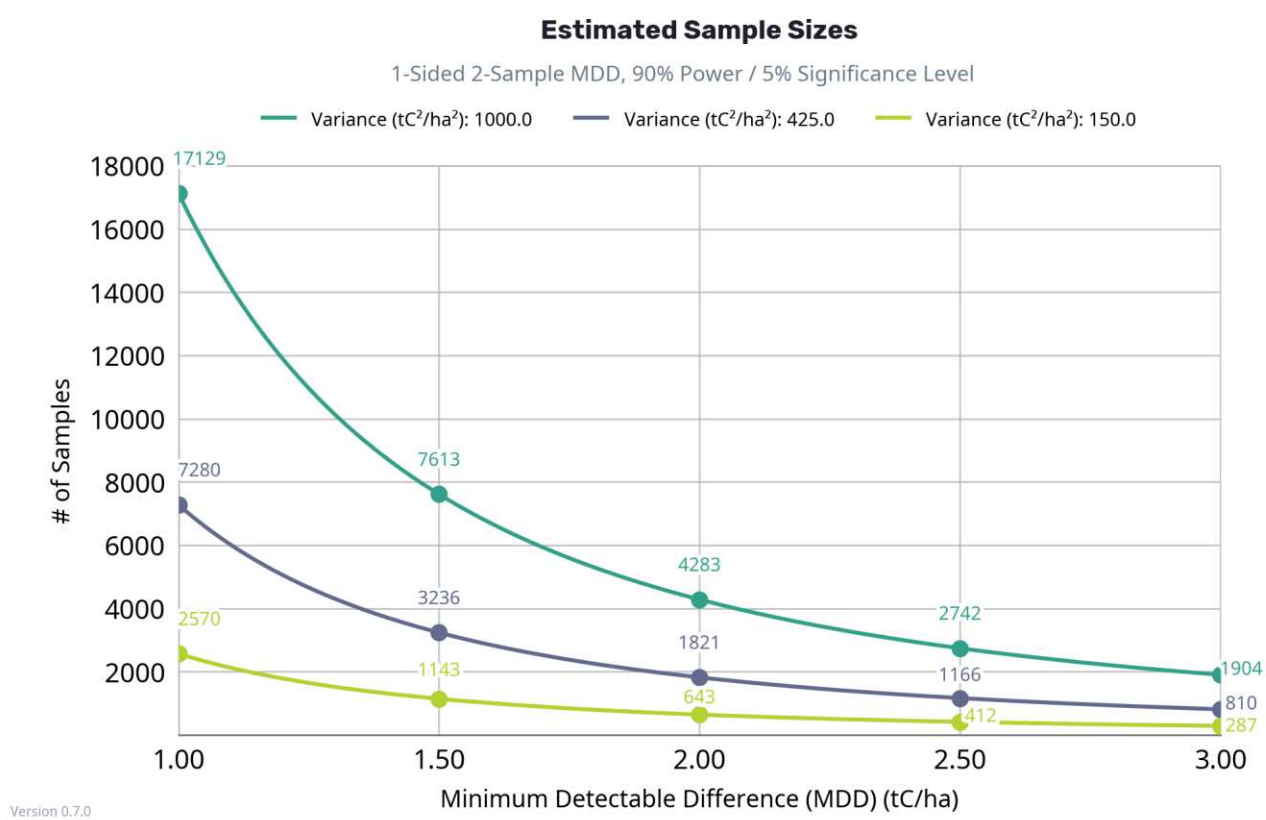
Calculating sample size using the MDD approach involves several important inputs and parameters, which are:

- **One-tailed or Two-tailed test:**
  - One-tailed test is recommended, allowing for directional claims (e.g., SOC sequestration > 0tC/ha).
  - Two-tailed tests are common in literature but not required in the context of VM0042 projects. They require higher sample sizes.
- **Null Hypothesis (H_0)**: Recommended to be set at "Actual SOC Stock change is less than or equal to 0tC/ha". If the null hypothesis can be rejected, the statistical claim that can be made becomes: There was a statistically significant positive SOC Stock change. The MDD can then be interpreted as the smallest SOC sequestration required to detect a statistically significant positive SOC stock change for a given number of samples.
- **Paired- or independent-sample:**
  - *Paired-sample t-test*: Requires an estimate of the variance of SOC Stock change and paired sampling.
  - *Independent-sample (2-sample) t-test*: Requires an estimate of the variance of SOC Stock snapshot at two points in time and requires independent sampling.
  - 1-sample t-tests using the variance of a snapshot once are only applicable under VM0042 if used for DSM validation, as they are not designed to detect change but rather to measure a snapshot.
- **MDD value (tC/ha)**: Should be set conservatively and need to be ≤ Expected Net SOC stock change. Overly optimistic MDD values can lead to underestimated sample sizes. Figure 5 illustrates the sensitivity of the sample size requirements on the MDD level chosen.

- **Variance estimate (tC^2/ha^2):** Depending on the type of t-test, this needs to be the variance of change or a snapshot. Variance can vary significantly across quantification units. Conservative estimates are strongly advised to not underestimate the sample size.
- **Significance level (%):** Typically set at 5%. This defines the probability of a false positive (Type I error)—i.e., detecting change where none exists.
  - Common erroneous assumption: The lower significance level, the more samples are required.
  - Validation bodies, auditors and offtakers would want significance to be low.
- **Statistical power (%):** Typically set at 80–90%. This defines the probability of correctly detecting a true change (1 – Type II error).
  - Common erroneous assumption The higher Power, the more samples are required.
  - Project proponents would want power to be high.
- **(optional) Design effect:** The better the design effect, achieved through high quality stratification or model assisted estimations, the lower the sample size required to achieve the same precision.

Figure 5 shows a range of sample size calculations following the MDD approach.

Figure 5. Example illustrating effect of different MDD levels and SOC Stock variances on sample size requirements.



7.1.3 Margin of Error approach (MoE) to calculate sample size

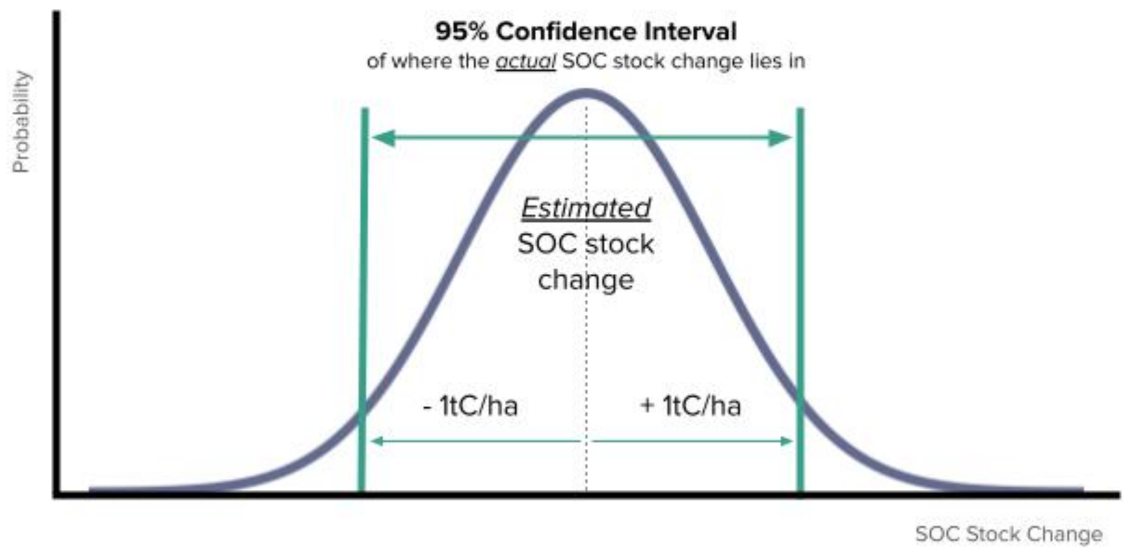
Applying the Margin of Error (MoE) approach, illustrated in Figure 6, helps answer the question:

> "How many samples are needed to estimate a SOC stock change with a targeted precision/MoE at a certain confidence level?"

Example answer:

> "With 267 samples, we are 95% confident that the *actual* SOC stock change lies within ±1 tC/ha of the estimated value."

Figure 6. Visual representation of the confidence interval and Margin of Error concept (Example: MoE ±1 tC/ha at 95% confidence).



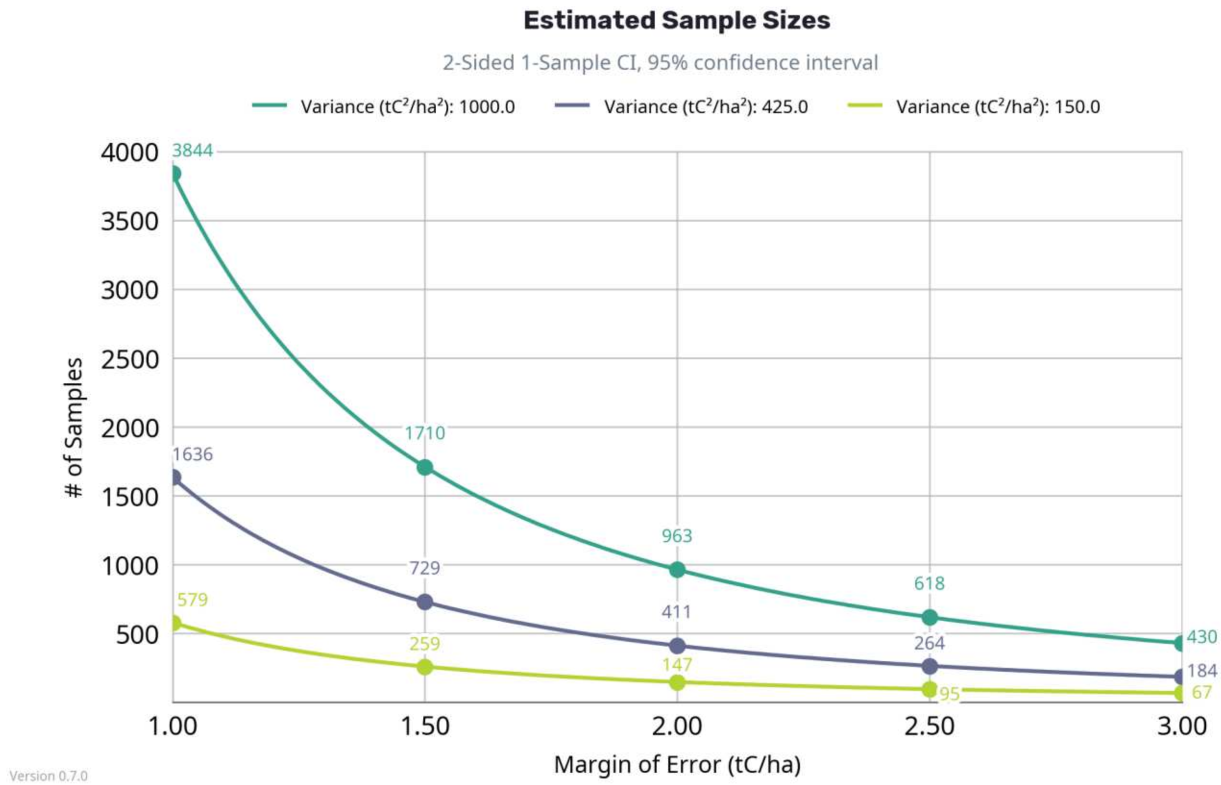
Inputs needed for calculation:

- **Number of tails:**
  - A *two-tailed confidence interval* is recommended, allowing for claims of ± MoE around the measured mean.
  - A *one-tailed confidence interval* is not recommended, even though it may result in smaller sample sizes, as the statistical claims it supports are not particularly useful in the context of VM0042 projects.
- **Targeted Margin of Error (MoE, tC/ha):** The MoE should be set conservatively and must be ≤ expected SOC Stock change. Overly optimistic MoE assumptions lead to underestimated sample sizes. Figure 7 shows how sample size requirements are affected by the MoE level.
- **Variance estimate (tC^2/ha^2):** Depending on the sampling method, this could either be the variance of change (for paired sampling) or a snapshot (for independent sampling). Variances can differ significantly for different types of quantification units and have to be estimated on a case-

by-case basis. Conservative estimates are recommended to avoid underestimating the sample size.

- **Confidence interval (%):** Typically 90%-95%. The higher the confidence interval, the more samples are required.
- **(optional) Design effect**: The better the design effect, achieved through high quality stratification or model assisted estimations, the lower the sample size required to achieve the same MoE.

Figure 7. Example illustrating effect of Margin of Error levels and different SOC Stock variances on sample size requirements at 95% confidence interval.



7.1.4 Model True-up Sample Size Example

Below is a description and the equations for calculating the sample size requirements for the true-up of a biogeochemical model.

Per VM0042, the sample size may only be determined using the true-up logic described below *after* the biogeochemical model has passed its first validation using data from within the project area. In practice, this means it can only be applied starting from the second remeasurement campaign, provided that the true-up has been passed based on the change data obtained from the initial sampling campaign and the first remeasurement campaign.

Numerical example calculation for estimating sample size for the Bias test at True-up:	
Step 1	Determine the following quantities from the model validation process/report in accordance with the VMD0053 procedure: $n_{VMD0053,wp}$: Number of samples in the validation dataset representing the project scenario $n_{VMD0053,bsl}$: Number of samples in the validation dataset representing the baseline scenario $s^2_{pooled}(\epsilon_{VMD0053,wp})$: Variance of the model prediction errors in the validation data sample representing the project scenario $s^2_{pooled}(\epsilon_{VMD0053,bsl})$: Variance of the model prediction errors in the validation data sample representing the baseline scenario. e.g. $n_{VMD0053,wp} = 150; n_{VMD0053,bsl} = 150; s^2_{pooled}(\epsilon_{VMD0053,wp}) = 10;\ s^2_{pooled}(\epsilon_{VMD0053,bsl}) = 10;$ From these the sampling variance of the mean bias estimates can be estimated: $s^2_{\widehat{BIAS}_{VMD0053,wp}} = \frac{s^2_{pooled}(\epsilon_{VMD0053,wp})}{n_{VMD0053,wp}} = \frac{10}{150} \approx 0.0667$ and $$s^2_{\widehat{BIAS}_{VMD0053,bsl}} = \frac{s^2_{pooled}(\epsilon_{VMD0053,bsl})}{n_{VMD0053,bsl}} = \frac{10}{150} \approx 0.0667$$
Step 2	Determine or assume the uncertainty deductions ($UD_{t,\Delta t}$) from the relevant project verification event. In practice at the Model True-up this value will be known from the model uncertainty calculated with the analytical or Monte Carlo approach. Before the end of the monitoring period this value may be assumed or estimated to obtain an estimate for the required number of samples at model true-up. For this example, we assume an uncertainty deduction of 1 tC / ha ($UD_{t,\Delta t} = 1\ tC/ha$).
Step 3	Iterate over a range of possible values for $n_{wp,t,\Delta t} = 2, 3, \ldots$, then for each $n_{wp,t,\Delta t}$:
Step 3a	Calculate the estimate of degrees of freedom with equation 23 $$\nu(n_{wp,t,\Delta t} = 2) = \frac{\left(s^2_{\widehat{BIAS}_{wp,t,\Delta t}} + s^2_{\widehat{BIAS}_{VMD0053,bsl}}\right)^2}{\frac{\left(s^2_{\widehat{BIAS}_{wp,t,\Delta t}}\right)^2}{n_{wp,t,\Delta t} - 1} + \frac{\left(s^2_{\widehat{BIAS}_{VMD0053,bsl}}\right)^2}{n_{VMD0053,bsl} - 1}} = \frac{\left(\frac{10}{2} + 0.0667\right)^2}{\frac{\left(\frac{10}{2}\right)^2}{(2-1)} + \frac{0.0667^2}{150-1}} \approx 1.027,$$
Step 3b	Calculate the non-centrality parameter λ, with equation 25b: $$\lambda(n_{wp,t,\Delta t} = 2) = \frac{1\ tC/ha}{\sqrt{\frac{10}{2} + \frac{10}{150}}} \approx 0.444,$$
Step 3c	Calculate the power with equation 24: $$power(n_{wp,t,\Delta t} = 2) = 1 - F_{NCT}(t_{\alpha(1)=0.95,\nu=1.03} = 6.313\ ;\ \nu = 1.027, \lambda = 0.444),$$ $$power(n_{wp,t,\Delta t} = 2) \approx 1 - 0.921 \approx 0.079$$
Step 3d	As long as $power(n_{wp,t,\Delta t}) < 0.95$, repeat steps 3a-3c with increasing $n_{wp,t,\Delta t}$, until: …

Numerical example calculation for estimating sample size for the Bias test at True-up:	
	$$\nu(n_{wp,t,\Delta t} = 396) = \frac{\left(s^2_{\widehat{BIAS}_{wp,t,\Delta t}} + s^2_{\widehat{BIAS}_{VMD0053,bsl}}\right)^2}{\frac{\left(s^2_{\widehat{BIAS}_{wp,t,\Delta t}}\right)^2}{n_{wp,t,\Delta t}-1} + \frac{\left(s^2_{\widehat{BIAS}_{VMD0053,bsl}}\right)^2}{n_{VMD0053,bsl}-1}} = \frac{\left(\frac{10}{396} + 0.0667\right)^2}{\frac{\left(\frac{10}{396}\right)^2}{(396-1)} + \frac{0.0667^2}{150-1}} \approx 268.654,$$ $$\lambda(n_{wp,t,\Delta t} = 396) = \frac{1\ tC/ha}{\sqrt{\frac{10}{396} + \frac{10}{150}}} \approx 3.298,$$ $$power(n_{wp,t,\Delta t} = 396) = 1 - F_{NCT}\left(t_{\alpha(1)=0.95,\nu=268.654} = 1.650\ ;\ \nu = 268.654, \lambda = 3.298\right),$$
	$$\boldsymbol{power(n_{wp,t,\Delta t} = 396) \approx 1 - 0.050\ \approx\ 0.950}$$

In this example, with 150 data points in the VMD0053 validation dataset and an uncertainty deduction of 1 tC/ha, a minimum sample size of 396 samples would be required for the True-up to pass the Bias test with a power of 95%.

7.2 Guidance for Sample Size Determination under VT0014 Estimating Organic Carbon Stocks Using Digital Soil Mapping

VT0014 permits the use of DSM to:

- initialize and/or true-up any model that requires an estimate of SOC content, SOC stock, or bulk density (BD).
- predict SOC content and/or SOC stock under measure and model or measure and re-measure approaches.

This Appendix provides guidance on sample size and sampling design determination that supports the use of *VT0014*. Sections 1 – 3 describe key concepts and factors influencing sampling decisions under *VT0014*. Sections 4 and 5 provide specific guidance for determining the minimum sample size for validation and uncertainty assessment given a calibrated digital soil mapping (DSM) model, and variogram-specific sampling guidelines.

7.2.1. The Role of Sampling under Model-based Statistical Approaches

DSM is a model-based form of statistical inference, which differs fundamentally from traditional design-based inference, such as a measure and re-measure approaches under *VM0042*. Under design-based inference, estimates of SOC stock and stock change are calculated directly from the sampled data. The validity of those estimates and their uncertainty is derived from the sampling design. In a design-based framework, probability sampling is essential because it ensures that estimates are unbiased and provides a basis for quantifying sampling variance.

In contrast, *VT0014* operates under a model-based statistical paradigm in which inference is derived from a statistical model rather than from the sampling design and data alone. The model produces spatially explicit predictions of SOC stock at a specific point in time. Estimates at two points in time are combined to derive the predicted SOC stock change. Because inference is driven by the model, uncertainty is a function of the model architecture and calibration, not the sampling design itself. Instead, model-based statistical inference requires demonstrating that model predictions accurately estimate uncertainty, that the model has predictive power, and that the average model prediction error is not significantly different from zero. As a result, probability sampling is not required for estimation of SOC stock, stock change, or uncertainty estimation under a model-based approach. A probability sample can be used to construct a representative calibration and validation dataset, but alternative sampling strategies may also be appropriate and, in some cases, more efficient, as long as they allow rigorous evaluation of model performance (validation) and model prediction error.

Distinction between calibration and validation samples for DSM

Project proponents may collect soil samples for DSM model calibration, validation, or both. In many cases, a single dataset is used for both purposes, with validation implemented through cross-validation to maintain conditional independence between calibration and validation subsets. Alternatively, proponents may combine opportunistic or legacy data with intentionally collected samples for calibration and use a separate dataset exclusively for validation.

VT0014 does not impose requirements on the calibration data set. However, the data used for validation must meet defined criteria: validation data must be located exclusively within the project area, representative of project conditions, obtained during the same season as the target date, and collected in accordance with the requirements of the applied methodology (see Table 1 in *VT0014*). The guidance below focuses primarily on sampling requirements and sample size considerations for model validation.

7.2.2. Sampling for DSM: model calibration and validation

The objective of sampling under *VT0014* is to support statistical model-based inference rather than to directly estimate SOC stock from field measurements alone. Samples are used to: (1) calibrate the DSM model used to predict SOC content, SOC stock, or BD; (2) validate model predictions; and (3) quantify prediction error for use in uncertainty propagation under the target methodology. Model calibration, validation, and uncertainty estimation are conducted separately for each monitoring time point and change over time is calculated from the difference between predicted stocks. Consequently, uncertainty in SOC stock change depends on the uncertainty associated with predictions at each time point (see Equation 5 in *VT0014*).

7.2.3. Variables influencing the sample size for DSM

Sample-size calculations must result in a conservative estimate of the variance of the SOC stock change. In practice, the appropriate sample size may be driven by one or more of the following considerations: model calibration and validation needs, reliable estimation of model prediction error, characterization of spatial covariance of model prediction error, and conservative estimation of the resulting variance in SOC stock change. When a variogram is used to quantify spatial covariance in model prediction error, the fitting requirements specified in *VT0014* may increase the required sample size beyond what is needed for model calibration and validation alone (see Section 7.2.5).

These considerations are further influenced by several interacting project-specific factors, including the sampling design of field measurements, the type of statistical or machine learning procedure used, the choice and quality of covariates, the magnitude and spatial distribution of SOC stocks within the project area, and the overall project size. The geometry of the project area (e.g., compact versus elongated configurations, fragmentation, and spatial heterogeneity of management units) can also affect the variance of the prediction error of the mean of model predictions and, consequently, the variance of the estimate of the mean SOC stock change. Together, these technical and contextual factors determine the uncertainty of DSM-based estimates at the project scale.

Because determining a sample size that adequately supports the calibration, validation, and uncertainty objectives is non-trivial, project proponents are encouraged to consult qualified experts (e.g., DSM model experts) to determine an appropriate sampling design and sample size that balances sampling costs with uncertainty.

7.2.4. Empirical determination of the sample size for DSM

To determine an appropriate sample size for validation and uncertainty estimation for a given project, the project proponent should identify a performance metric and select an acceptable level of performance (the "performance threshold"), then work backward to identify the sampling intensity needed to meet that target. The relevant metric depends on how DSM is used in the project (Use Case 1 versus Use Case 2) and, for Use Case 1, how the biogeochemical model is implemented (i.e., point-based versus area-based):

1. Point-based implementation of Use Case 1:
   The performance metric is the prediction error for prediction location i at time t when the biogeochemical model is initialized using samples from the predictive distribution for prediction location i at time t (i.e., point-based biogeochemical model; see Section 5.1.2 of *VT0014*).
2. Area-based implementation of Use Case 1:
   The performance metric is the variance of the prediction error of the mean of model predictions at time t when the biogeochemical model is initialized using the predictive distribution of SOC stock for an area at time *t* (i.e., area-based biogeochemical model; see Section 5.1.2 of *VT0014*).
3. Use Case 2:
   The performance metric is the variance of the estimate of the mean SOC stock change (see Section 5.4 of *VT0014*).

The value of the performance threshold is dependent on context. It depends on the magnitude of total emission reductions and carbon dioxide removals (SOC stock change), the frequency of project verification, and overall project economics. For example, a project with large reductions/removals and relatively low MRV costs may be able to tolerate a larger variance in the estimate of the mean stock change – and therefore a larger uncertainty deduction – than a project with small reductions/removals and higher MRV costs.

Working backward from a target performance threshold to determine sample size is only feasible after information about model performance and uncertainty is available. Because this information is obtained through calibration and validation, it is not available in the absence of a calibrated DSM model. A previously calibrated model may be used, or a new model may be developed that depends on initial soil sampling. In order to conduct an initial sampling campaign, projects could adopt one of the following approaches:

1. Follow sample size guidance for design-based measure and re-measure approaches (see Section 3.3), which are assumed to be conservative (i.e., the sample size required for a design-

based approach is conservatively assumed to be greater than or equal to that required for DSM).

2. Use previously collected data, or conduct limited soil sampling, to obtain preliminary estimates of model performance and uncertainty before implementing a more comprehensive sampling campaign.

Once the project has collected the initial calibration and validation dataset needed to characterize model performance and uncertainty, the following procedure may be used to evaluate and, if necessary, adjust sampling intensity for future DSM sampling campaigns:

1. Assemble the calibration and validation dataset (*X*) to be used for prediction at time t and complete all steps in Section 5.1 of VT0014. This results in a calibrated and validated DSM model.
2. Resample from dataset (*X*) with replacement to generate a new instance of (*X*). If *X* contains n observations, randomly sample rows < n from X with replacement to generate new instances of *X, Xj,n*. For example, if *X* contains *n* = 5,000 rows, randomly sample instances of *Xj,n* with incrementally increasing sample sizes (e.g., 100, 200, 300, … 4,900 rows, corresponding to instances of *Xj*,100, *Xj*,200, *Xj*,300, … *Xj*,4900, respectively).
3. Repeat all steps in Section 5.1 of VT0014 for every instance of *Xj,n*.
4. Quantify the relationship between the value of the performance metric and the sample size.
5. The minimum acceptable sample size is the smallest number of samples for which the performance metric meets the predefined performance threshold. A simulated example is shown in Figure 8.

Sampling for VT0014 under VM0042

Under VM0042, sampling must follow a stratified random design. Stratified random sampling requires that sampling locations are selected randomly within predefined strata. In a DSM context, strata should be defined in a way that supports effective model calibration and validation. One key principle is to ensure that sampling locations are well distributed across the model’s feature space.

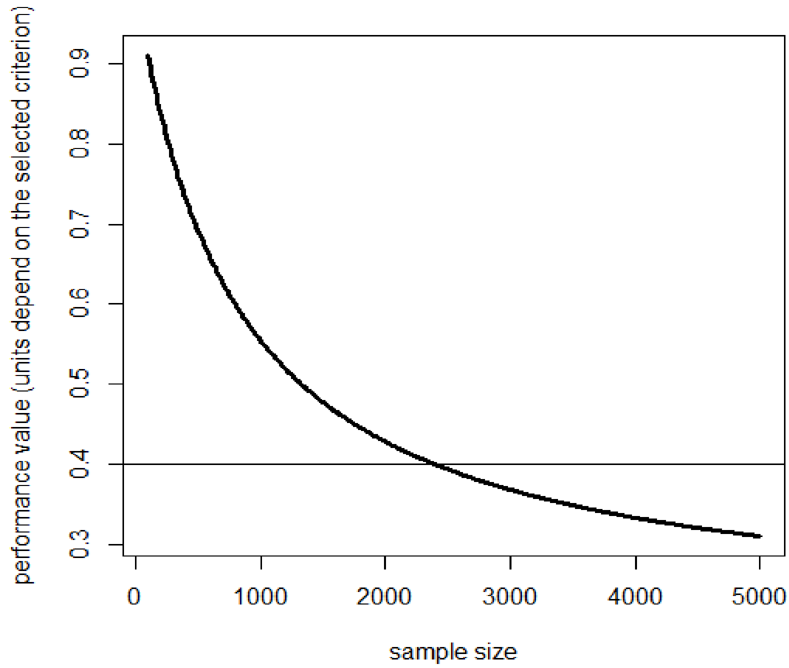


Figure 8. Determination of the minimum sample size from simulated data.
The Y axis is a performance metric (e.g., the standardized prediction error under Use Case 1), and the X axis is the sample size. Assume that the project proponent has identified a performance threshold of 0.4 units. The point at which the performance threshold intersects the curve is the minimum acceptable sample size. Here this number is about 2,400 samples.

The feature space is composed of the covariates used by the model (e.g., terrain attributes, climate variables, remote sensing indices). Although a spatially representative sample often provides representative coverage of feature space, this is not guaranteed. In some cases, important regions of the covariate space may be underrepresented even when geographic coverage appears adequate. Strata can therefore be defined in feature space rather than exclusively in geographic space. An efficient stratification improves model performance by ensuring that samples adequately cover the range, combinations, and structure of the covariates used for prediction, thereby reducing extrapolation.

Methods for stratifying the feature space include k-means clustering of covariate features and conditioned Latin hypercube sampling, among other options. Importantly, strata defined in feature space do not need to be spatially contiguous. Random sampling can still be applied within these strata to meet the stratified random sampling requirements of *VM0042* (or any other methodology that requires stratified random sampling), while simultaneously supporting model calibration and validation consistent with *VT0014*.

7.2.5. Variogram fitting requirements and additional sampling needs

In addition to sampling for model calibration and validation, the need to estimate the spatial correlation structure of the standardized prediction error may impose further constraints on the sample size and design of sampling units. When the method selected to account for spatial covariance requires a variogram, project proponents should ensure that the sampling design and total number of samples are adequate for variogram estimation (See Section 5.5 of *VT0014*).

VT0014 provides guidance on variogram fitting, including recommendations related to bin width, number of bins, and the minimum number of point pairs per bin (See Section 5.5.1 of *VT0014*). To evaluate whether a proposed or existing sampling design satisfies these recommendations, users can calculate a pairwise distance matrix for all sample locations and examine the resulting distribution of inter-point distances. Using this distance matrix, users can assess how alternative bin widths interact with the guidelines in Section 5.5.1 of VT0014. If an existing set of samples is inconsistent with the guidelines in Section 5.5.1 of VT0014, additional candidate sampling locations can be generated at random and evaluated. Project proponents can implement this process iteratively to identify a set of additional locations that more closely matches the guidance without explicitly targeting specific spatial configurations (i.e., while avoiding non-random sample placement).

7.3 Practical Considerations for Paired Sampling: Addressing Spatial sampling precision and Small-Distance SOC Variability

Paired sampling aims to detect changes in SOC stock over time by measuring the same sampling location at two different time points (t and t+Δt). However, due to the destructive nature of soil sampling, where the core sampled at time t is destroyed during lab analysis, it is impossible to sample the exact same core at t+Δt. Instead, a core should be sampled at a location as close as possible to the original location, assuming that the SOC stock at t at this new location sampled at t+Δt is equivalent to the SOC stock of the sample analyzed at time t. A critical aspect of paired sampling is ensuring that differences in SOC measurements between t and t+Δt are genuinely due to changes in SOC stock rather than small spatial distance variability of SOC. This raises two key questions:

What are the feasible spatial precisions for paired sampling?

How significant is the SOC stock variability within such small spatial distances?

1. Feasibility of Spatial Precision for Paired Resampling Campaigns

Practically, the challenge lies in revisiting the exact same location with a high degree of precision. GPS accuracy typically ranges from 5 to 10 meters, which means that during paired sampling, the sampling locations at *t* and *t+Δt* could well be 20 meters apart while still being recorded as the same coordinate (see Figure 9). Additional uncertainties in locating the exact sampling point arise from operator imprecision and potential inaccessibility, such as areas blocked by tramlines or other obstacles.

Figure 9. Visualization of GPS precision and its impact on the practical feasibility of sampling the exact same geolocation.

Given these practical constraints, it becomes important to evaluate how much variability in SOC stock can be expected within the typical spatial distance between t and t+Δt when trying to sample the same location.

2. Small-Distance Variability of SOC

SOC stocks exhibit considerable variability over small spatial distances. Research by Poeplau, Prietz, and Don (2022), Bradford et al. (2023), and Potash et al. (2025) has demonstrated this phenomenon, often referred to as plot-scale or **small-distance variability.** In their studies, Poeplau et al. found significant SOC stock differences (5.1 and 7.6 tC/ha on 0-30 cm cropland and grassland) even over distances as small as 40 cm. This difference is larger than the expected Net SOC stock changeover 5 years. These large variations in SOC stock underscore the challenge of ensuring that observed changes are genuinely due to temporal variations and not spatial heterogeneity.

Conclusion: Small-distance SOC stock variability needs to be addressed during sampling design

Paired sampling is a practical and often preferred method for detecting SOC stock changes over time, particularly given its advantages in reducing sample size requirements and improving model calibration. However, its limitations, especially the influence of small-distance SOC variability and limited spatial precision, must be addressed during the sampling design phase to ensure that the goals of the campaign are met (see Table 1).

In practice, small-distance SOC variability introduces noise into estimates of SOC stock change derived from paired samples. While paired sampling may still outperform independent sampling in some contexts, its design must be considered the following:

Individual samples: Variance in SOC stock change is likely overestimated when relying on individual samples which are paired, due to added spatial noise from small-distance SOC Stock variability. Increasing the number of samples helps average out this noise. It is therefore advisable to use conservative ex-ante estimates for variance of change when calculating sample sizes.

Statistical aggregates or physical composites (e.g., within a 10 m radius): Small-distance variability contributes uncertainty to the area-level SOC estimate. This can be mitigated by increasing the number of sub-samples, though at higher cost.

For QA1: biogeochemical models should consider small-distance variability and its impact on obtaining measurements of change as it may impact model calibration and validation.

In summary, while paired sampling remains a key requirement for QA1 and a valuable method for tracking SOC stock changes in QA2, its practical limitations must be carefully managed. Project proponents should critically assess the impacts of spatial precision and small distance SOC variability when creating sampling designs. Simulating the impacts of small-distance variability can optimize sampling design and improve the robustness of SOC change detection.

7.4 Model-assisted estimations to improve precision and reduce uncertainty deductions or sample size

Introduction

Model-assisted estimators use a statistical model to integrate ancillary (e.g., remote sensing) data with ground-based measurements. They provide a statistically sound way to improve the precision of SOC stock and SOC stock change estimates — without sacrificing objectivity and strict validity of design-based estimators (Brus, 1997). These estimators build on probability sampling designs (e.g., stratified simple random sampling) and are supported by decades of theory and applications (Cochran, 1977; Särndal et al., 1992; Breidt & Opsomer, 2017).

Statistical Background

Model-assisted estimators for mean SOC stocks are asymptotically design-unbiased, meaning their expected value converges to the true mean SOC stocks as the number of samples increases. This property holds even if the underlying statistical model is incorrect. This is in contrast to model-based estimators (i.e: DSMs or biogeochemical models), which rely on the model alone and may introduce bias and invalid uncertainty estimates if the model is not well-calibrated and validated properly.

In the model-assisted approach, a “working model” is fit using the sample data and ancillary variables. The final estimate for the mean SOC stocks consists of:

- A prediction for the mean SOC stocks based on the working model,
- A bias correction term based on model residuals computed from the probability sample.

This combination ensures objectivity and strict validity: even if the working model is wrong and produces biased estimates, the estimator remains unbiased as it takes that bias into account. If the model is good, the uncertainty (sampling variance) of the estimator is reduced.

Overall, model-assisted estimation is accurate, as the estimators do not extrapolate beyond the probability sample. It is also conservative, as working models do not reduce the uncertainty compared to classical design-based estimation. In addition, no model validation is required, because the probability sample still determines the estimation result.

Table 8. Comparison of different inference approaches

Question	Inference approach		
	Design-Based (Sampling)	Model-Assisted (Sampling)	Model-Based (DSMs and biogeochemical models)
Where does the **estimate** (SOC sequestration) come from?	From **samples**	From **samples**, supported by a working model	From **model**, trained on samples
What gives us **confidence** in the result?	Unbiased by design for every project area	Unbiased by design for every project area	Strong model must be empirically validated for a specific project area
What happens if the **model** is wrong?	N/A	• **SOC stock (change) estimate:** still unbiased • **Uncertainty deductions:** No improvements compared to design-based approach → no integrity risk and project proponent can always fall back to design-based approach	• **SOC stock (change) estimate:** can be biased • **Uncertainty deductions:** can be under- or overestimated
Requires **probability sampling**?	Yes	Yes	Yes, but only for model *validation* only. Not for model calibration.
Model validation required?	N/A	**No** as estimate is unbiased by design and uncertainty deductions cannot be underestimated	Yes

7.5 Assessment of Background Information

Background information about each project area will be required to select and execute the appropriate sampling design processes described. This information can come from a variety of sources such as pre-sampling campaigns (also known as reconnaissance sampling), DSMs, relevant soil datasets (e.g. national soil inventories and surveys) and third-party datasets, locally applicable academic studies, or a combination thereof. While all these sources may be used background information, the following must be reported in the sampling plan for every source of data used:

1) **Source details** such as name/title of the dataset or publication, author or organization, date of data collection and publication, and a link or citation (if applicable).
2) **Data characteristics** meaning the geographic location (boundaries) of the data, resolution and scale, sampling/collection methods, and units.
3) **Relevance and comparability.** Justification that the data matches the project's context and conditions.
4) **Any limitations or caveats** such as missing data or known biases.
5) **Pre-processing or processing** that the project or technical service providers performed on the data for use in the project.

The main sources of background information are given below, including examples and important considerations for the sampling design.

7.5.1 Academic Studies

Academic studies published in peer-reviewed journals can provide useful background information for a project, from defining target carbon sequestration, to providing estimates of SOC stock variance in a region. Project proponents must demonstrate the suitability of the academic studies used and make sure they provide evidence for compatibility with the project.

7.5.2 Third-Party (Private) Data Sets

Third-party data (e.g. commercial, unpublished) that matches the project conditions, scale, and scope, can be useful for the sampling design. If projects use third-party data, it is important to demonstrate that the data matches project characteristics, for example bioclimatic region, geography, land use, crop type and other factors. **Important Depth Consideration:** External data sources often do not provide SOC stock estimates to depths of 30 cm or beyond. When using such datasets, it is important to recognize the lack of information on SOC distribution in the subsoil (or lower topsoil) as a critical limitation. Notably, SOC stock variance tends to increase significantly with depth. Therefore, datasets limited to shallow depths should not be used to derive SOC stock estimates for deeper profiles unless appropriate depth corrections are applied.

7.5.3 Digital Soil Mapping and Remote Sensing-Derived Variables

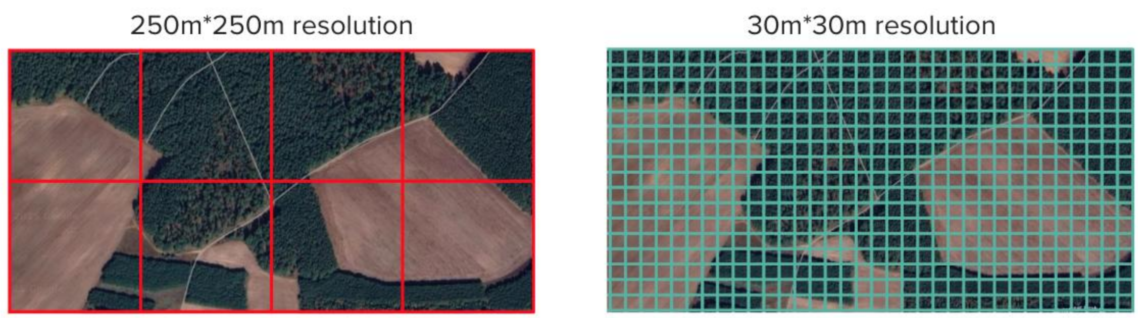
Global and local digital soil products can be useful in the Stratification and sample size estimation process by providing information on relevant soil properties (e.g., texture, SOC content, bulk density, SOC variance). Useful resources include:

- **FAO Soils Portal**, including the Harmonized World Soil Database (HWSD)
- **SoilGrids by ISRIC**, which offers coarse-resolution (250m*250m) predictions of SOC and other soil variables
- **Local or national digital soil maps**, which may offer finer-scale or more field-validated data e.g. USDA Web Soil Survey; Ireland National Soil Survey; UK Soil Observatory
- **DSMs bespoke for a project area** derived from direct measurement or remote-sensing variables.
- **Indices derived from remote sensing variables** such as normalized difference vegetation index (NDVI) or topographic wetness index (TWI), land surface temperature, or terrain derivatives (e.g., topographic position, slope, aspect, elevation) derived from Sentinel-2, Landsat, or digital elevation model data (e.g., SRTM)
- **Historical satellite imagery**, optionally combined with expert knowledge or participatory mapping, can indicate areas with similar land cover and land use patterns

Important considerations when using these data sources in the sampling design are:

1) **Spatial resolution**: The spatial resolution of available data is often limited and can introduce bias—particularly when surrounding areas are not representative of the actual quantification unit (see visual below).

Figure 10. Higher spatial resolution (right side) reduces the risk of biased insights by minimizing the influence of data from outside the target area.



2) **Depth Considerations:** SOC stock variance increases significantly with depth, making accurate depth representation critical. However, external data sources used in SOC stock modelling often lack estimates beyond shallow depths and often are limited to <30cm. If this limitation is not properly addressed during model development, predictions will be biased and may misrepresent SOC stocks for the intended depth interval (e.g., 0–30 cm). This can lead to significantly under-estimating the SOC Stock variance and hence the required sample size.

3) **Measurement error.** The inherent measurement uncertainty of the data source must be accounted for when interpreting results or making decisions based on the data.
4) **Incomplete spatial and / or temporal coverage.** Data from remote-sensing sources can be subject to limitations regarding its spatial or temporal coverage which may result in gaps in the data. If interpolation or imputation is performed to obtain a complete dataset, the sampling plan must justify the purpose of this process, and the methodology used.
5) **Smoothing.** Often, spatial models use Gaussian frameworks and target the mean of the target variables. While this is a useful working assumption, it can lead to underestimation of the value at the tails, particularly for soil properties, and must be treated carefully. If these models are used to determine SOC stock variance, project proponents must be careful that sources do not provide a lower (smoother) estimate of variability. If this happens then sample sizes may be underestimated.

7.5.4 Reconnaissance Surveys and Pre-Sampling

Projects may want to carry out a **reconnaissance survey** where there is insufficient background information or existing data to support sampling design. Surveying by local experts could provide useful information on land use, soil type distributions and other local characteristics to aid design.

Pre-sampling is a sampling campaign that could be carried out before the main sampling event at t=0. It may provide additional background information and could help refine sampling efficiency. It is particularly useful in the case that information on the project conditions is poor or nonexistent. Important considerations are:

1) Pre-sampling entails additional fieldwork and can consequently increase project costs and outweigh the benefit of optimizing the sampling design.
2) The aims and purpose of pre-sampling must be carefully considered, e.g. providing a preliminary estimate of SOC stock variance or aiding in Stratification through preliminary covariate data collection. The sampling methodology chosen for pre-sampling must therefore reflect the aims.
3) Pre-sampling must match project scope and scale if it is then used to inform soil variability across the project and influence justification of sample size assessment, allocation and placement.

7.5.5 Ex-ante estimating variance of the target variable

To calculate sample sizes, an ex-ante estimate of the variance of the target variable is required (see Section 3.2). It is strongly recommended to use conservative estimates, as the variance has a direct and significant impact on sample size calculations. Underestimating this variance can result in underestimating the sample size and hence lead to higher uncertainty deductions and potentially even applicability of crediting in general.

Estimating the variance of SOC Stock at a *snapshot*:

For independent sampling with independent statistical tests, as well as for BIOGEOCHEMICAL MODEL model initialization, sample size should be based on the variance of SOC stock measured at a single point in time (i.e., a snapshot).

While the SOC stock snapshot variance at the project start can be estimated using DSMs, finding representative datasets, or prior studies to estimate the variance at the time of remeasurement is more challenging. A common and practical assumption which may be appropriate is that the SOC Stock variance remains constant over time: SOC Stock snapshot variance (s^2): $s_t^2 = s_{t+\Delta t}^2$. Under this assumption, the variance of the SOC Stock change between t and t+Δt is:

$$s_{\Delta t}^2 = s_t^2 + s_{t+\Delta t}^2 = 2s_t^2$$

Estimating the variance of SOC Stock *change*:

For paired sampling and paired statistical tests, sample size must be based on the variance of SOC stock change between times t and t+Δt. However, estimating future changes is challenging and till today lacks a scientifically robust framework. This difficulty arises due to two key sources of variability:

1. The heterogeneity of practice changes across the landscape, which is often not known ex-ante by project proponents.
2. The variable impact of those changes across different sub-domains within the quantification unit.

A significant practical complication is that small-distance SOC variance can introduce noise into paired measurements, obscuring the true SOC stock variance of change (see Appendix 7.2). To mitigate these challenges, it is advisable to adopt conservative estimates for the variance of change. One practical approach is to proxy the variance of change estimate using the estimate of the variance of SOC Stock snapshot. This also helps address the issues posed by small distance SOC stock variability as discussed in Appendix 7.2.

Estimating the spatial variance of model prediction error:

For QA1, once the biogeochemical model has passed true-up and has been validated using project-specific data (so: no earlier than the third sampling campaign), sample size estimation should be based on the spatial variance of the model prediction error of the biogeochemical model.

This variance should be derived from the observed prediction errors when evaluating the model on paired samples from times t and t+Δt from within the quantification unit. Importantly, it is strongly advised to not use external validation datasets (e.g. from long-term experiments) to estimate the spatial variance of the model prediction error, as they may not reflect the conditions and variability within the actual project area and may lead to underestimating the sample size requirements.

7.6 Multistage sampling deepdive

As explained in Section 3.5, Multistage Sampling Design is permitted under *VM0042* but is often misunderstood or misapplied. Each additional sampling stage introduces extra uncertainty, which increases overall uncertainty unless offset by collecting more soil samples.

Sample Size Considerations:

Determining the appropriate sample size in a multistage design requires expert analysis to account for uncertainty at every stage. Generally, multistage sampling requires more samples than stratified simple random sampling.

Below is a real-world example illustrating the increase in sample size for a two-stage sampling design compared to simple random sampling:

<u>Project specifics:</u>

- #fields: 23.109 fields
- Combines size: 115.531ha
- Expected change: 1.96tC/ha/5-years
- Spatial Variance: 431.76tC^2/ha^2
- Type of sampling: Independent sampling

Sampling requirements following 1-sided MDD sample size calculation approach (section 3.2.2.2) with an MDD of 1.96tC/ha/5-years, 5% significance and 90% power.

In this example the fields present the primary sampling unit and the individual sample cores present the secondary sampling unit. Table 9 shows how sample sizes are impacted when comparing simple random sampling and a 2-staged sampling design. The main driver for the added variance of the estimate in multi-stage sampling is the additional between cluster variance component, as described by Brus (2023).

Comparison: Multistage Sampling vs. Stratified Random Sampling

While stratified random sampling can be applied within each stage of a multistage sampling design, the two concepts differ significantly:

1) Sampling units: In multistage sampling, only the randomly selected units are sampled or subdivided further. In contrast, stratified random sampling requires sampling every stratum in full.
2) Subdivision criteria: Multistage sampling allows for a variety of subdivision criteria, such as farm type or provincial boundaries. Stratification, however, groups units based on homogeneous SOC stocks (see Section 3.2).

Table 9. Example of the effect of multistage sampling on sample size

Parameter affected	Baseline: Simple random sampling	2-stage sampling with fields being PSUs			
		Scenario1: 2 samples per field[11]		Scenario2: 9 samples per field[12]	
		Absolute values	Delta to baseline	Absolute values	Delta to baseline
Sample size	1931	2240	+ 309 (+16%)	2897	+966 (+50%)
#fields to be visited	1756 (7.6% of all fields)	1124 (4.9% of all fields)	-632 (-36%)	316 (1.4% of all fields)	-1440 (-82%)
Average sampling density per field	1.01 samples per field	2 samples per field	+0.9 samples per field	9 samples per field	+7.9 samples per field

Additional Requirements for Multistage Sampling

Multistaged sampling introduces additional requirements for the validation of a project's sampling design:

1) Definition of sampling units: Clear justification for criteria defining sampling units at every stage.
2) Proof of random selection: Transparent evidence that sampling units at each stage were selected randomly without bias.
3) Quantification of uncertainties: Explicit quantification and propagation of uncertainties introduced by multiple sampling stages into overall project uncertainty.

Further guidance on implementing multistage sampling is provided in de Gruijter et al. (2006), Potash et al. (2023, 2025), and Brus (2023).

[11] Two samples per field is the absolute minimum to allow for staged sampling. It would mean that no stratification is done at the second to last stage anymore and that simple random sampling is done at that last stage.

[12] Nine samples would allow for a stratification of the second to last sampling unit. One could for example allow for three strata with three samples each.

7.7 Optional Additional Quantification Units at the Project Proponent's Discretion

Project proponents (PPs) are not required to define more than one Quantification Unit (QU) per enrollment year. However, additional QUs may be created at the PD's discretion for practical or strategic purposes. For example, a carbon credit buyer (offtaker) may request VM0042-compliant emission reduction and/or removal (ERR) statements specific to the region they are purchasing credits from, rather than for the project as a whole.

Important: ERR Claims that are formally validated against VM0042 requirements **must be made at the QU level.** VM0042 does not support validated ERR claims at finer spatial scales (e.g., individual fields or farms) unless those areas are defined as separate QUs.

That said, **more granular tracking and reporting is still possible** using alternative quantification approaches, including:

- **QA1:** Modeling-based methods
- **QA2:** Estimators which are Design-based (i.e: sampling), model-assisted (see appendix 7.3) , or model-based (i.e: Digital Soil Mapping)

These approaches can be used for internal purposes, such as:

- Incentivizing farmers based on their individual ERR outcomes
- Providing detailed, non-VM0042-compliant reporting to partners or stakeholders

Such internal or third-party reporting **does not require VM0042 compliance**, as long as the validated ERR claims for credit issuance remain aligned with the methodology at the QU level.

8 REFERENCES

Awad, A., & Scharwächter, E. (2025). Model-assisted estimation for MRV: How to boost the economics of SOC sequestration projects without compromising on scientific integrity. arXiv. https://doi.org/10.48550/arXiv.2510.1369

Aynekulu, E., Vågen, T. G., Shepherd, K., & Winowiecki, L. (2011). A protocol for modeling, measurement and monitoring soil carbon stocks in agricultural landscapes (Version 1.1). World Agroforestry Centre.

Beem-Miller, J. P., Kong, A. Y. Y., Ogle, S., & Wolfe, D. (2016). Sampling for soil carbon stock assessment in rocky agricultural soils. Soil Science Society of America Journal, 80(5), 1411–1423. https://doi.org/10.2136/sssaj2015.11.0405

Bisutti, I., Hilke, I., & Raessler, M. (2004). Determination of total organic carbon—An overview of current methods. TrAC Trends in Analytical Chemistry, 23(10), 716–726. https://doi.org/10.1016/j.trac.2004.09.003

Bradford, M. A., Eash, L., Polussa, A., Jevon, F. V., Kuebbing, S. E., Hammac, W. A., Rosenzweig, S., & Oldfield, E. E. (2023). Testing the feasibility of quantifying change in agricultural soil carbon stocks through empirical sampling. Geoderma, 440, Article 116587. https://doi.org/10.1016/j.geoderma.2023.116719

Brus, D. J. (2023). Spatial sampling with R. CRC Press.

Cochran, W. G. (1977). Sampling techniques (3rd ed.). Wiley.

Croft, H., Kuhn, N. J., & Anderson, K. (2012). On the use of remote sensing techniques for monitoring spatio-temporal soil organic carbon dynamics in agricultural systems. Catena, 94, 64–74. https://doi.org/10.1016/j.catena.2012.01.001

de Gruijter, J. J., Brus, D. J., Bierkens, M. F. P., & Knotters, M. (2006). Sampling for natural resource monitoring. Springer.

de Gruijter, J. J., McBratney, A. B., Minasny, B., Wheeler, I., Malone, B. P., & Stockmann, U. (2016). Farm-scale soil carbon auditing. Geoderma, 265, 120–130. https://doi.org/10.1016/j.geoderma.2015.11.010

Dietz, C. L., Jackson, R. D., Ruark, M. D., & Sanford, G. R. (2024). Soil carbon maintained by perennial grasslands over 30 years but lost in field crop systems in a temperate Mollisol. Communications Earth & Environment, 5(1), Article 1. https://doi.org/10.1038/s43247-024-01500-w

Ellert, B. H., & Bettany, J. R. (1995). Calculation of organic matter and nutrients stored in soils under contrasting management regimes. Canadian Journal of Soil Science, 75(4), 529–538. https://doi.org/10.4141/cjss95-075

England, J. R., & Viscarra Rossel, R. A. (2018). Proximal sensing for soil carbon accounting. SOIL, 4, 101–122. https://doi.org/10.5194/soil-4-101-2018

Even, R. J., Machmuller, M. B., Lavallee, J. M., Zelikova, T. J., & Cotrufo, M. F. (2025). Large errors in soil carbon measurements attributed to inconsistent sample processing. SOIL, 11(1), 17–34. https://doi.org/10.5194/soil-11-17-2025

Food and Agriculture Organization of the United Nations. (2019). Measuring and modelling soil carbon stocks and stock changes in livestock production systems: Guidelines for assessment (Version 1). FAO.

Food and Agriculture Organization of the United Nations. (2023). Standard operating procedure for soil bulk density, cylinder method. FAO. https://doi.org/10.4060/cc7568en

Fowler, A. F., Basso, B., Millar, N., & Brinton, W. F. (2023). A simple soil mass correction for a more accurate determination of soil carbon stock changes. Scientific Reports, 13(1), Article 2242. https://doi.org/10.1038/s41598-023-29289-2

Gifford, R. M., & Roderick, M. L. (2003). Soil carbon stocks and bulk density: Spatial or cumulative mass coordinates as a basis of expression? Global Change Biology, 9(11), 1507–1514. https://doi.org/10.1046/j.1365-2486.2003.00677.x

Grossman, R. B., & Reinsch, T. G. (2002). Bulk density and linear extensibility. In J. H. Dane & G. C. Topp (Eds.), Methods of soil analysis (pp. 201–228). Soil Science Society of America. https://doi.org/10.2136/sssabookser5.4.c9

Hedley, C. B., Payton, I. J., Lynn, I. H., Carrick, S. T., Webb, T. H., & McNeill, S. (2012). Random sampling of stony and non-stony soils for testing a national soil carbon monitoring system. Soil Research, 50(1), 18–29. https://doi.org/10.1071/SR11171

Hengl, T., Rossiter, D. G., & Stein, A. (2003). Soil sampling strategies for spatial prediction by correlation with auxiliary maps. Australian Journal of Soil Research, 41(8), 1403–1422. https://doi.org/10.1071/SR03005

ISO. (2018). Soil quality—Sampling—Part 104: Strategies (ISO 18400-104:2018). International Organization for Standardization.

Kish, L. (1965). Survey sampling. Wiley.

Loeppert, R. H., & Suarez, D. L. (1996). Carbonate and gypsum. In Methods of soil analysis (pp. 437–474). Soil Science Society of America. https://doi.org/10.2136/sssabookser5.3.c15

Lohr, S. L. (2010). Sampling: Design and analysis (2nd ed.). Brooks/Cole.

Maillard, É., McConkey, B. G., & Angers, D. A. (2017). Increased uncertainty in soil carbon stock measurement with spatial scale and sampling profile depth in world grasslands. Agriculture, Ecosystems & Environment, 236, 268–276. https://doi.org/10.1016/j.agee.2016.11.024

McBratney, A. B., & Minasny, B. (2010). Comment on "Determining soil carbon stock changes: Simple bulk density corrections fail." Agriculture, Ecosystems & Environment, 136(1), 185–186. https://doi.org/10.1016/j.agee.2009.12.010

Mudge, P., McNeill, S., Hedley, C., Roudier, P., Poggio, M., Malone, B., Baldock, J., Smith, P., McNally, S., Beare, M., & Schipper, L. (2020). Design of an on-farm soil carbon benchmarking and monitoring approach for individual pastoral farms (Technical Paper No. 2020/02). Ministry for Primary Industries.

Nelson, D. W., & Sommers, L. E. (1996). Total carbon, organic carbon, and organic matter. In Methods of soil analysis (pp. 961–1010). Soil Science Society of America. https://doi.org/10.2136/sssabookser5.3.c34

Peng, Y., Chahal, I., Hooker, D. C., & Van Eerd, L. L. (2024). Comparison of equivalent soil mass approaches to estimate soil organic carbon stocks under long-term tillage. Soil and Tillage Research, 238, Article 106021. https://doi.org/10.1016/j.still.2024.106021

Potash, E., Bradford, M., Oldfield, E. E., & Guan, K. (2025). Measure-and-remeasure as an economically feasible approach to crediting soil organic carbon at scale. Environmental Research Letters, 20(2), Article 024025. https://doi.org/10.1088/1748-9326/ada16c

Potash, E., Guan, K., Margenot, A. J., Lee, D., Boe, A., Douglass, M., Heaton, E., Jin, V., Li, N., Mitchell, R., Schmer, M., Wang, S., & Zumpf, C. (2023). Multi-site evaluation of stratified and balanced sampling of soil organic carbon stocks in agricultural fields. Geoderma, 438, Article 116587. https://doi.org/10.1016/j.geoderma.2023.116587

Prihar, S. S., & Hundal, S. S. (1971). Determination of bulk density of soil clod by saturation. Geoderma, 5(4), 283–286. https://doi.org/10.1016/0016-7061(71)90039-5

Rovira, P., Sauras-Yera, T., & Romanyà, J. (2022). Equivalent-mass versus fixed-depth as criteria for quantifying soil carbon sequestration. Catena, 214, Article 106283. https://doi.org/10.1016/j.catena.2022.106283

Särndal, C.-E., Swensson, B., & Wretman, J. H. (1992). Model assisted survey sampling. Springer.

Sharma, S., Wilson, T., Ochsner, T., & Warren, J. G. (2020). Sampling probes affect bulk density and soil organic carbon measurements. Agricultural & Environmental Letters, 5(1), Article e20005. https://doi.org/10.1002/ael2.20005

Sherrod, L. A., Dunn, G., Peterson, G. A., & Kolberg, R. L. (2002). Inorganic carbon analysis by modified pressure-calcimeter method. Soil Science Society of America Journal, 66(1), 299–305. https://doi.org/10.2136/sssaj2002.2990

Thompson, S. K. (2012). Sampling (3rd ed.). Wiley.

Vågen, T., & Winowiecki, L. (2023). The LDSF field manual: Land and soil health assessments using the Land Degradation Surveillance Framework (LDSF). https://ldsf.thegrit.earth/assets/LDSF_Field_Manual_31Jan2023.pdf

Vanguelova, E. I., Bonifacio, E., De Vos, B., Hoosbeek, M. R., Berger, T. W., Vesterdal, L., Armolaitis, K., Celi, L., Dinca, L., Kjønaas, O. J., Pavlenda, P., Pumpanen, J., Püttsepp, Ü., Reidy, B., Simončič, P., Tobin, B., & Zhiyanski, M. (2016). Sources of errors and uncertainties in the assessment of forest soil carbon stocks at different scales. Environmental Monitoring and Assessment, 188, Article 630. https://doi.org/10.1007/s10661-016-5608-5

von Haden, A. C., Yang, W. H., & DeLucia, E. H. (2020). Soils' dirty little secret: Depth-based comparisons can be inadequate for quantifying changes in soil organic carbon. Global Change Biology, 26(7), 3759–3770. https://doi.org/10.1111/gcb.15124

Wagner, S. W., Hanson, J. D., Olness, A., & Voorhees, W. B. (1998). A volumetric inorganic carbon analysis system. Soil Science Society of America Journal, 62(3), 690–693. https://doi.org/10.2136/sssaj1998.03615995006200030021x

Wendt, J. W., & Hauser, S. (2013). An equivalent soil mass procedure for monitoring soil organic carbon in multiple soil layers. European Journal of Soil Science, 64(1), 58–65. https://doi.org/10.1111/ejss.12002

Wiesmeier, M., et al. (2019). Soil organic carbon storage as a key function of soils: A review of drivers and indicators at various scales. Geoderma, 333, 149–162. https://doi.org/10.1016/j.geoderma.2018.07.043