Draft VCS Tool

CN0137

QUANTIFYING ORGANIC CARBON STOCKS USING DIGITAL SOIL MAPPING: CALIBRATION, VALIDATION, AND UNCERTAINTY ESTIMATION

Draft Version 1.0

20 February 2025

Sectoral Scope 14: AFOLU

This draft tool was developed by Perennial Climate Inc. and TerraCarbon LLC.

CONTENTS

1 SOURCES

This tool is based on and uses the most recent versions of the following VCS Program methodologies and modules:

- *VM0032 Methodology for the Adoption of Sustainable Grasslands through Adjustment of Fire and Grazing*
- *VM0042 Methodology for Improved Agricultural Land Management*
- *VMD0053 Model Calibration, Validation, and Uncertainty Guidance for the Methodology for Improved Agricultural Land*

2 SUMMARY DESCRIPTION

Digital soil mapping (DSM) is the use of spatially explicit computer models to predict soil properties using gridded ancillary variables (McBratney et al. 2003). Numerous studies have demonstrated the ability to quantify soil organic carbon (SOC) using DSM techniques (Castaldi et al. 2019; Fu et al. 2024; Gomez et al. 2008; Ratnayake et al. 2016; Sothe et al. 2022; Szatmári et al. 2021; Venter et al. 2021; Wadoux and Heuvelink 2023; Zhou et al. 2020). Most focus on SOC content, measured as a proportion or percentage of oven-dry soil mass. Others have quantified SOC stock in units of total mass or mass density (e.g., mass of SOC per unit area). These studies encompass a wide range of geographic locations, statistical procedures, and measurement sources, including airborne and satellite remote sensing. However, specific guidance on appropriate calibration, validation, and uncertainty estimation is needed to support the use of DSM in agricultural land management (ALM) carbon projects to ensure robust and verifiable quantification

This tool contains protocols for using digital soil mapping (DSM) as a quantification technology when applying existing VCS methodologies. The tool specifically addresses:

- DSM model calibration, model validation,[1] and uncertainty estimation. The calibration and validation of a DSM model to predict SOC content in units of proportion or

[1] The *VCS Standard, v4.7* defines validation as "… the independent assessment of the project by a validation/verification body that determines whether the project and its GHG statement conforms with the VCS Program rules and evaluates the reasonableness of assumptions, limitations, and methods that support a claim about the outcome of future activities." This tool distinguishes between project validation under the VCS Standard and model validation. Model validation, as defined in VCS module *VMD0053 Model Calibration, Validation, and Uncertainty Guidance for the Methodology for Improved Agricultural Land Management*, *v2.0*, is "the process of evaluating model performance relative to measured values." *VMD0053, v2.0* states that a validated model demonstrates "satisfactory performance in terms of goodness of fit and characterization of model prediction error."

percentage by soil mass, bulk density (BD), SOC stock, and the change in SOC stock over time during the project crediting period.

- collection of physical soil samples, including guidelines on sampling design and stratification, depth increments, and the selection of laboratory methods.
- selection and processing of remote sensing and other environmental covariates.
- use of DSM to initialize and/or true-up a biogeochemical process-based model (BGCM) (e.g., Quantification Approach 1: Measure and Model in VCS methodology *VM0042 Improved Agricultural Land Management*).
- use of DSM as a primary measurement technology (e.g., Quantification Approach 2: Measure and Re-Measure in *VM0042*, *v2.1* and the Measured or Modeled quantification approaches identified in *VM0032 Methodology for the Adoption of Sustainable Grasslands through Adjustment of Fire and Grazing*).

By providing a generalized approach to the quantification of SOC stocks in vegetated or bare agricultural soils, this tool enables model calibration, model validation, and uncertainty estimation to be conducted in compliance with the guidelines established under the applied VCS methodology.

2.1 Key Concepts and Rationale: Validation with Safeguards

The procedures documented in this tool provide a robust framework for using DSM to quantify SOC stocks and stock changes in agricultural lands. The approach relies on detailed soil sampling, model validation procedures, and safeguards – including “true-up” procedures and equations for the calculations of cumulative carbon stock change – to ensure long-term accuracy and flexibility for real-world projects. The tool supports two primary applications:

Use Case 1: DSM to initialize and/or true-up a BGCM

In Use Case 1, methods described in the tool provide SOC stock values for initializing and/or truing-up a BGCM. In this case, the DSM model is validated (and optionally recalibrated) at project start and at least once every 5 years.

Use Case 2: DSM as a primary measurement approach

In Use Case 2, DSM is used to directly quantify SOC stocks and stock changes. DSM model validation is required at the start of the project and at least once every 5 years. When verification is sought between model validation events, the DSM model must be recalibrated subject to sampling and other requirements described in Section 5.1.6.

2.1.1 Validation-Driven Approach with Safeguards

The tool requires both time-specific and project-specific model validation using representative in-situ soil samples. The model applied at the start of the verification period (t_0) must be validated against in-situ samples collected at t_0 in accordance with the requirements outlined here. DSM model validation by an approved independent modeling expert (IME)[2] is required by the time of the project's first verification and at least once every five years thereafter (Appendix 1). Between these model validations, projects may undergo verification and issue credits under Use Case 2, but doing so requires the model to be recalibrated at each verification event. When credits have been issued in the absence of model validation, a cumulative carbon stock change adjustment ensures the accuracy of credits over the project's lifetime by reconciling previously generated credits against validated measurements. A schematic description of this process is illustrated in Figure 1 and APPENDIX 2. This approach was designed to enable real-world use of the tool, including rolling enrollments, without compromising rigor.

2.1.2 Model Validation Requirements

To be considered valid, a DSM model must pass three tests. These tests are conducted at a specific point in time using independent in-situ soil samples collected through a representative sampling design of the project area.

1) Coverage: at least 90% of validation observations must fall within 90% prediction intervals.
2) Goodness of fit: $R^2 > 0$, indicating that the model provides a more precise estimate than the mean of validation data[3]
3) Lack of bias: the model prediction error is not significantly different from 0 at the 0.05 level under a two-tailed one-sample t-test.

These tests are necessary and sufficient to demonstrate that the DSM model is accurate, precise, and unbiased across the project area and that its uncertainty is well-quantified. Repeating these tests through time demonstrates that the DSM model maintains these properties throughout the project's lifetime.

2.1.3 Flexibility in Model Development

The tool's approach to validation and safeguards allows for flexibility. State-of-the-art environmental and remote sensing covariates are permitted by the tool along with any statistical or machine-learning procedure that has been demonstrated by a publication appearing in the Web of Science: Science Citation Index (SCI; available at https://mjl.clarivate.com). The tool's approach to validation and safeguards requires representative sampling and is outcome-based, rather than prescriptive. This flexibility encourages innovation and adaptability to real-world conditions while upholding verifiable scientific integrity.

2.1.4 Similarities to Existing Approaches

Projects using the methods documented in this tool are procedurally similar to traditional soil sampling or BGCM methods already in use in the VCS Program. In particular:

---

[2] See Appendix 1 Assessment by Independent Modeling Expert (IME)

[3] See Janssen and Heuberger (1995) and Wadoux et al. (2022) for justification of the appropriateness of the $R^2 > 0$ threshold.

- in-situ soil measurements form the basis of calibration and validation.
- uncertainty (including accounting for spatial covariance of model prediction errors) is quantified and accounted for using approaches similar to existing VCS methodologies.
- periodic resampling and validation within project and baseline sites mirror measure-remeasure approaches.
- the use of a true-up and cumulative carbon stock change adjustments provides protections equivalent to those used for BGCM crediting under existing VCS methodologies.
- rules for baseline control sites, uncertainty propagation, and how the uncertainty deduction is calculated are not changed by the tool – these guidelines follow the applied methodology in all cases.

The procedures outlined in this tool draw from extensive peer-reviewed literature and establish a scientifically rigorous and flexible framework for employing DSM in ALM projects. By focusing on thorough validation, key safeguards, and verifiable best practices, the tool provides a roadmap for accurate SOC quantification at a single point in time and through time while allowing project proponents to leverage the efficiency and scalability of DSM approaches.

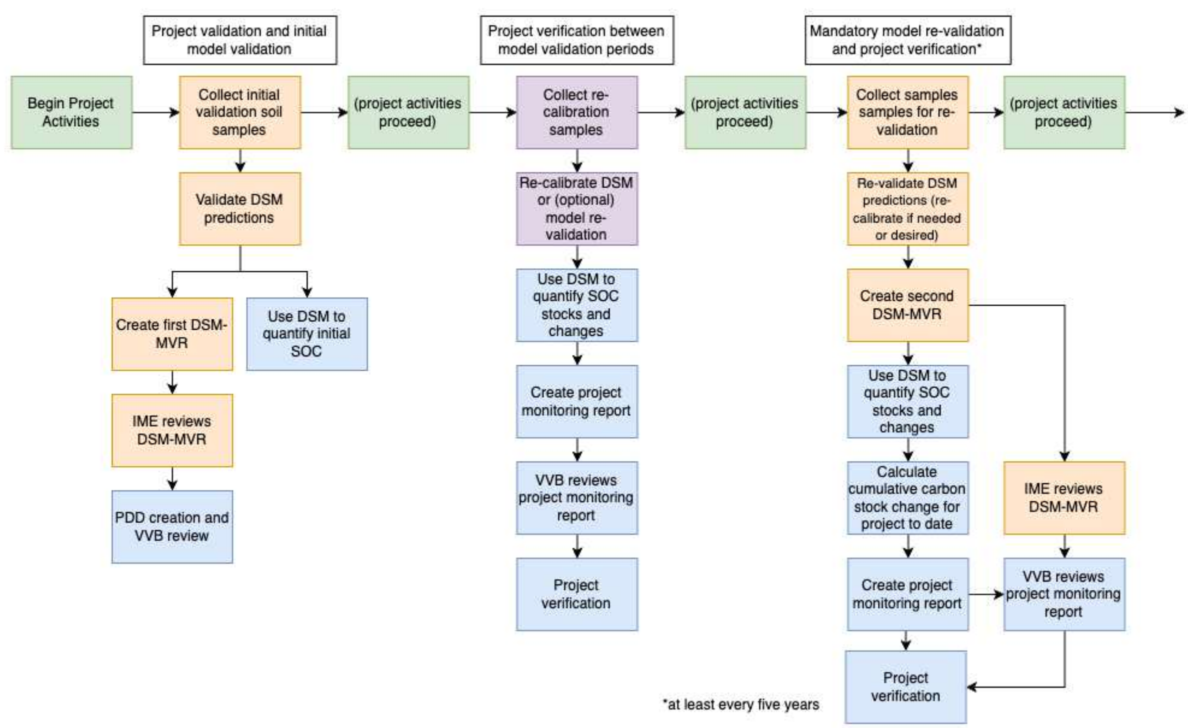
Figure 1. Example quantification life cycle for a project with rolling enrollment of new project areas (additional project lifecycle scenarios are provided in APPENDIX 2)

Example project quantification lifecycle

SCENARIO 1: > Not Grouped > Model Validation at Project Validation > Verification Every 2-3 Years	Y0	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Project Validation											
Model Validation Soil Samples Collected	1000**					1000					1000
Model Validation (optional re-calibration) (year validation soil data was collected)	(Y0 val. data)					(Y5 val. data)					(Y10 val. data)
Cumulative carbon stock change adjustment											
Model Re-Calibration Soil Samples Collected				100				100			
Model Re-Calibration (year re-calibration soil data was collected)				(Y3 cal. data)				(Y7 cal. data)			
Documentation Required	PDD DSM-MVR			MR*		MR DSM-MVR		MR*			MR DSM-MVR
VVB Audit											
IME Assessment											

* Monitoring Reports (MRs) must report re-calibration statistics in re-calibration years.
** Sample numbers are illustrative. Actual number of soil samples collected should follow the guidance in Section 5.

Process flow for the above scenario:



3 DEFINITIONS

In addition to the definitions set out in the *VCS Program Definitions*, the following definitions apply to this tool.

Biogeochemical process-based model (BGCM)
A computational tool that applies biogeochemical principles to simulate soil processes, such as variations in soil organic carbon stocks.

Calibrated model
A workflow that uses digital soil mapping to produce an estimate of soil organic carbon stocks, along with associated uncertainty estimates for all prediction support units within a project area.

Calibration
The process by which a digital soil mapping model learns from data to minimize prediction error, and can be closed-form (e.g., least-squares) or iterative (e.g., machine learning and Bayesian methods). Calibration results in a set of parameters or settings that can be used to generate predictions from the model.

Calibration and validation dataset
A dataset that contains the target variable and covariates used to calibrate and validate the digital soil mapping model. The calibration and validation dataset may contain direct measurements of soil organic carbon stock and numerous corresponding environmental and remote sensing variables associated with each soil organic carbon stock measurement.

Cluster
A group within a staged sampling design that is chosen for sampling with a known, non-zero probability. Cluster sampling is employed when direct sampling of individual elements is challenging or impractical.

Covariate
Inputs to a model, used to predict the target variable during and after calibration. Covariates can be measured or modeled and may be generated using feature engineering. See APPENDIX 3 for additional resources related to covariate selection.

Coverage
The percentage of independent measurements that are correctly predicted by a model. Coverage is assessed using a prediction interval by counting the number of validation observations within a given prediction interval. Nominal coverage occurs where the percentage of observations within the prediction interval matches the prediction interval-width (e.g., 90% of observations are within the 90% prediction interval at 90% coverage).

Cumulative carbon stock change adjustment
A carbon crediting process that awards credit based on cumulative performance adjusted for systematic error (if any) at prior verification events.

Digital soil mapping (DSM)
The use of spatially explicit computer models to predict soil properties using gridded ancillary variables.

Expected average project effect (EAPE)
An estimate from at least one peer-reviewed journal article (or proprietary data where no peer-reviewed journal articles are available) of the expected mean change in soil organic carbon (SOC) stocks in the next five years of the project, expressed in units of SOC stock. EAPE is used to guide sample size requirements for model validation.

Feature engineering
The process of selecting, transforming, manipulating, and combining raw data into the covariates used in a model. This includes developing new variables, such as biophysical or vegetation indices, conducting quality assessment and data screening (e.g., excluding or processing satellite images with cloud cover and snow), and standardizing data values over time to ensure consistency. Outputs from feature engineering can be measured, modeled, or both.

Hyperparameter
A configuration setting that dictates the behavior and learning process of a machine learning or statistical model. Examples include the depth of a decision tree, the learning rate in gradient descent methods, and sampling characteristics of a Markov chain Monte Carlo procedure. In the context of this tool, hyperparameters also include weights applied to different sets of calibration data during calibration.

Hyperparameter tuning
The process of choosing values for hyperparameters prior to model calibration. When hyperparameter tuning is performed by evaluating model performance, the data used for hyperparameter tuning must be independent of the calibration and validation sets.

Independent modeling expert (IME)
An individual or organization with demonstrated competence in digital soil mapping, especially with respect to soil organic carbon stocks and error propagation, and independence from the project proponent.

Independent sample data
Sample data that are not used for calibration procedures under a given instance of a model. Independent sample data are used for model validation.

Lowest-level sampling unit (LLSU)
A stage within the project area that is homogeneous with respect to management practices. The LLSU must be sampled with simple random or stratified random sampling methods, in accordance with the requirements of the applied methodology.

Missing at random
Used to describe an incomplete dataset. Where data are missing at random, the probability that an observation is missing does not depend on the value of the missing observation.

Model architecture
Determines how the model processes input data and generates predictions. Examples include neural networks, gradient-boosted regression trees, and multiple regression. The components of model architecture include the number of layers or nodes in a neural network; activation functions; whether the model is linear or nonlinear, parametric or non-parametric, or frequentist or Bayesian; values for hyperparameters; and the loss function.

Mapped area
The area within which the digital soil mapping model is used to make predictions in the project area and/or baseline control sites.

Model instance
A calibrated version of a model generated during cross-validation, indexed in the tool using the subscript *k*.

Model prediction error
The difference between a model prediction and an independent measured value at a given location and prediction support unit.

Model validation
The process of evaluating a model's performance by comparing its predictions to measured values. In this tool, model validation requires passing three tests that address coverage, goodness of fit, and bias. Model validation ensures independence of calibration and validation data, achieved by using an independent validation set or cross-validation. Independence of calibration and validation data avoids overfitting, which occurs when a model learns specific details in calibration that fail to generalize, leading to poor performance under new conditions.

Model validation report for digital soil mapping (DSM-MVR)
A document describing calibration and validation procedures and outcomes for the digital soil mapping model used in the project (see APPENDIX 4). The DSM-MVR confirms that requirements related to coverage, goodness of fit, and bias have been met and includes the following components, i.e., the calibration and validation dataset, a description of all covariates and how they were generated, the architecture and components of the model.

Multi-stage sampling design
A hierarchical method used to generate a representative sample using a series of stages. The primary stage is the highest level in the hierarchy. Subsequent stages are smaller units within

larger stages in the hierarchy. For example, the primary stage could be a political or administrative boundary that contains smaller regions and fields.

Prediction support unit
The land area and volume of soil for which model predictions are calibrated and validated. The PSU refers to individual points or composite samples at specific depth intervals. For example, the PSU could be individual soil cores or composite samples from collections of soil cores over a specified depth range. Where the prediction support unit is the individual soil core, there must be a set of coordinates that define the collection location for the physical soil sample.

Probability sampling
Selecting samples from the population based on randomization. In a probability sample, every unit in the population has a known, non-zero probability of being included in the sample. This approach ensures that the sample is representative of the population, allowing for valid statistical inferences to be made. Probability sampling methods include simple random sampling, stratified sampling, systematic sampling, and cluster sampling, each of which maintains the principle of randomization.

Standardized prediction error
The prediction error at a given prediction support unit divided by the predictive standard deviation.

Stratum
A homogeneous subset of a quantification unit. Stratification in a sampling design involves dividing a larger unit into homogeneous segments, all of which are sampled. Strata are created to ensure that each subgroup within the population is adequately represented in the sample. For example, fields might be grouped into geographic clusters, some of which are selected for sampling. Where all fields within a selected cluster are sampled independently, these fields represent strata. Individual fields can also be stratified based on one or more variables. Stratification improves the precision of sampling by minimizing variability within each stratum.

Target date
The date of a given prediction from the calibrated model. For example, a calibrated model could generate a prediction of SOC stock in a specific prediction support unit for 1 July 2025. The initial target date (t_0) for a prediction support unit, used in validation and prediction, is set based on the date of the first validation sampling campaign that includes that unit in the sampling design.

Target variable
The dependent variable being predicted or explained by a digital soil mapping model (e.g., soil organic carbon (SOC) as a percentage by mass, SOC stock, bulk density).

4 Applicability Conditions

This tool is applicable globally under the following conditions:

1) A calibration and validation dataset is used to calibrate and validate a model of SOC stock.

1) The project activity is agricultural land management (e.g., an improved cropland management project type on agricultural land, grasslands, or rangelands).

2) All calculations are auditable by a validation/verification body (VVB) and an independent modeling expert (IME).

3) The mapped area contains less than 30% tree canopy cover.

The tool is not applicable under the following conditions:

4) The project area has permanent flooding.

5 Procedures

This workflow is a systematic approach for calibrating and validating a DSM model and for estimating the uncertainty in predictions of SOC stocks and stock changes under the applied VCS methodology. Model validation occurs on every project using independent sample data. A graphical illustration of the workflow is provided in Figure 2. For simplicity, this process is described for a single time point, *t*, and a single DSM model that generates predictions of SOC stock. In practice, this process may be applied at multiple time points.

The property being validated is always the prediction of SOC stock, even where separate components of SOC stock (SOC content and BD) are independently predicted. The variance of the estimate of the mean change in SOC stock over the duration of the project relative to baseline conditions is used to compute uncertainty and the associated uncertainty deduction under the applied methodology (see Sections 5.35.4 and 5.4).

Multiple localized DSM models may be used for different subsets of the project area. Where multiple localized models are used, the model validation procedure must be conducted in aggregate across the outputs of the localized models for the entirety of the project area. In grouped projects, every time a project activity instance is added to the project, an updated model validation report for digital soil mapping (DSM-MVR) must be created to demonstrate that the DSM model is valid in the new project areas represented by the new project instances.

All software, computer code, data and other dependencies must be documented, archived, version-controlled and available on request.

Estimates of SOC stock generated under the procedures outlined in this tool may be used in two ways:

- **Use Case 1:** DSM is used to predict SOC stocks to initialize and/or true-up a BGCM (e.g., Quantification Approach 1 in *VM0042*). Where DSM is used to true-up a BGCM, the DSM model is used to quantify SOC stocks across the project area at the time of the BGCM true-up. The DSM is validated prior to or at verification using the procedures given here, and the BGCM is validated according to the requirements of the applied methodology. Uncertainty of the DSM predictions of SOC is propagated through the BGCM based on the requirements of the applied methodology. Where Monte Carlo error propagation is permitted, project proponents must use the procedure in APPENDIX 5.

- **Use Case 2:** DSM is used to predict SOC stocks at a single point in time and changes in SOC stocks over time (e.g., Quantification Approach 2 in *VM0042*). The model requires validation at the start of the project and at least once every five years.[4] Verification may occur between years 1 and 5 at the discretion of the project proponent, but model calibration or model validation is required at every verification event (Figure 1, APPENDIX 2). Recalibration requirements are described in Section 5.1.6.

5.1 Model Development

Use the steps below to develop the model, as illustrated in Figure 2.

1) Assemble the calibration and validation data set (*X*) to be used for prediction at time *t*. Guidance on the collection of covariate and sample data are detailed in Sections 5.1.3 and 5.2, respectively.

2) Split *X* into *K* calibration and validation sets. The calibration set is used to estimate model parameters and the validation set is used to test model performance.

    a) Multiple procedures for identifying calibration and validation sets are applicable (e.g., completely independent validation set, leave one sample out, *k*-folds, geographically dependent cross-validation).

    b) Data from outside the project area may be used for calibration, but all data in the validation set must come from within the project area.

    c) Calibration and validation sets must remain independent (or where cross-validation is being used, conditionally independent). All parameter estimation, including hyperparameter tuning, where used, must not be exposed to observations in the validation set prior to validation. For multi-stage sampling designs, care must be

[4] Requiring calibration or validation at every verification event and requiring validation at least once every five years accommodates project proponents that prefer to fix model parameters after calibration and those who prefer to update the calibration over time. Regardless of the frequency of model calibration, all models used must pass the three validation tests described in Section 5.1 .

taken to account for data dependencies in the sampling design and avoid downward bias in model uncertainty estimation (see De Bruin et al., 2022)

3) Using the calibration set, calibrate an instance (*k*) of the model.

4) Generate predictions and estimate 90% prediction intervals for the predicted value of SOC stock for prediction support unit *i* at time *t* for every observation in the validation set.

    a) Prediction intervals for the validation points must be generated in a way that ensures independence of training and validation data.

    b) Any procedure that has been documented in at least one peer-reviewed publication in a journal indexed in the Web of Science: Science Citation Index may be used to estimate prediction intervals. Users must select one method and provide a justification for its use.

5) Repeat steps 3–4 *K* times to generate a new model instance for each iteration *k* and assemble a performance dataset for validation. The performance dataset contains pairs of predicted and observed values of SOC stock and the associated 90% prediction interval for each prediction.

    a) The appropriate number of iterations *K* depends on Step 2(a). Where a completely independent validation set is used, *K* = 1. Where cross-validation is used, *K* is the number of folds in the cross-validation procedure.

6) For every prediction support unit in the validation set, determine whether the prediction interval for SOC stock at time *t* contains the measured value. Where the validation sample falls within the prediction interval, the prediction passes. Otherwise, the prediction fails.

    a) Determine the percentage of validation observations that pass. The model must generate predictions that are within 90% prediction intervals at least 90% of the time, such that at least 90% of tests pass. If the model passes, proceed to step 7.

    b) Recalibration may require additional sampling, hyperparameter tuning, or other adjustments, but users must avoid fitting the model to the validation set after the properties of the validation set are known. Validation data must remain independent (or where cross-validation is used, conditionally independent).

7) Compute the prediction error for all paired observations in the validation set according to the following equation:

$$\epsilon_{i,t} = \widehat{SOC}_{i,t} - SOC_{obs,i,t} \tag{1}$$

Where:

$\epsilon_{i,t}$	=	Model error in prediction support unit *i* at time *t* (Mg C/ha)
$\widehat{SOC}_{i,t}$	=	Predicted SOC stock in prediction support unit *i* at time *t* (Mg C/ha)
$SOC_{obs,i,t}$	=	Observed SOC stock in prediction support unit *i* at time *t* (Mg C/ha)

8) Determine whether the mean model prediction error is significantly different from 0 using a one-sample *t*-test with α = 0.05.

9) Using all paired observations in the test set, compute the amount of variance explained (R^2) using:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} {\epsilon_{i,t}}^2}{\sum_{i=1}^{n} \left(SOC_{obs,i,t} - \overline{SOC}_{obs,t}\right)^2} \tag{2}$$

Where:

R^2	=	Fraction of variance in SOC stocks that is explained by the model (dimensionless)
$\overline{SOC}_{obs,t}$	=	Mean of observed values of SOC stock at time *t* (Mg C/ha)
n	=	Number of prediction support units in the project area (unitless)

The amount of variance explained (R^2) must be greater than zero.[5]

10) The model must pass all three of the following validation tests:

   a) Coverage is at least 90%.

   b) R^2 is greater than 0.

   c) Model prediction error is not significantly different from 0 under a two-tailed, one-sample *t*-test.

For a model that passes these three validation tests, predictions should then be generated over the project area within every prediction support unit and aggregated to determine the arithmetic mean SOC stock at time *t* using the following equation:

$$\overline{\widehat{SOC}}_t = \frac{1}{\sum_{i=1}^{n} A_i} \cdot \sum_{i=1}^{n} \widehat{SOC}_{i,t} \cdot A_i \tag{3}$$

Where:

$\overline{\widehat{SOC}}_t$	=	Mean of predicted SOC stock values at time *t* (Mg C/ha)
A_i	=	Area of prediction support unit *i* within the project area[6] (ha)

When new project activity instances are added to the project in different years, the time (*t*) of initial stock measurement will not be equivalent. The target date for the map of initial stock values is defined for each prediction support unit using the first year in which the unit is included in the sampling design (see APPENDIX 2).

[5] $R^2 > 0$ indicates that the model provides a more precise estimate than the mean of validation data. See Janssen and Heuberger (1995) and Wadoux et al. (2022) for further justification.

[6] Accounts for the possibility that prediction support units may be of different size.

11) Estimate the variance of the mean SOC stock at time *t* using any valid procedure that explicitly accounts for the covariance of model prediction errors between prediction support units (see Section 5.3 and APPENDIX 6). Further guidance on valid procedures to account for covariance of model prediction errors is provided in APPENDIX 7.

12) Repeat Steps 1–11 at every model validation event.[7] The mean stock change between times *t* and *t* + Δ*t* is calculated using the following equation:

$$\Delta\overline{\widehat{SOC}}_{t,\,\Delta t} = \overline{\widehat{SOC}}_{t,\Delta t} - \overline{\widehat{SOC}}_{t} \tag{4}$$

Where:

$\Delta\overline{\widehat{SOC}}_{t,\,\Delta t}$	=	Mean predicted SOC stock change over all prediction support units between time *t* and time *t* + Δ*t* (Mg C/ha)
$\overline{\widehat{SOC}}_{t,\Delta t}$	=	Mean of model predictions of SOC stock over all prediction support units at time *t* + Δ*t* (Mg C/ha)
$\overline{\widehat{SOC}}_{t}$	=	Mean of model predictions of SOC stock over all prediction support units at time *t* (Mg C/ha)

Mean predicted SOC stock change in the project area between time *t* and time *t* + Δ*t* ($\Delta\overline{\widehat{SOC}}_{t,\,\Delta t}$) is greater than zero when SOC stock increases between *t* and *t* + Δ*t*, and ≤0 otherwise.

The variance of the mean stock change is calculated from the sample variances and standard deviations using the following equation:

$$var\left(\Delta\overline{\widehat{SOC}}_{t,\,\Delta t}\right) = var\left(\overline{\widehat{SOC}}_{t+\Delta t}\right) + var\left(\overline{\widehat{SOC}}_{t}\right) - 2\rho \cdot \sqrt{var\left(\overline{\widehat{SOC}}_{t+\Delta t}\right)} \cdot \sqrt{var\left(\overline{\widehat{SOC}}_{t}\right)} \tag{5}$$

Where:

$var\left(\Delta\overline{\widehat{SOC}}_{t,\,\Delta t}\right)$	=	Variance of the mean of predicted SOC stock change across the project area between times *t* and *t* + Δ*t* (Mg C/ha)^2
$var\left(\overline{\widehat{SOC}}_{t+\Delta t}\right)$	=	Variance of the mean of predicted SOC stock at time *t* + Δ*t* (Mg C/ha)^2
$var\left(\overline{\widehat{SOC}}_{t}\right)$	=	Variance of the mean of predicted SOC stock at time *t* (Mg C/ha)^2
ρ	=	Correlation between the standard deviations of SOC stock at times *t* and *t* + Δ*t*

Each of the *var* terms is derived using the procedure from Step 11. The term $2\rho \cdot \sqrt{var\left(\overline{\widehat{SOC}}_{t+\Delta t}\right)} \cdot \sqrt{var\left(\overline{\widehat{SOC}}_{t}\right)}$ is the covariance of the error terms over time.

[7] For example, model validation is required every time when a new project instance is added to a project area.

The variance of the estimate of the mean SOC stock can be used to propagate uncertainty through BGCM simulations using the Monte Carlo method (see APPENDIX 5), as described in *VM0042*.

13) The mean carbon dioxide removal estimate in soil is calculated using the following equation:

$$CO2_{soil,\,t,\,\Delta t} = \left(\overline{\Delta\widehat{SOC}}_{t,\,\Delta t} - \overline{\Delta\widehat{SOC}}_{bsl,\,t,\,\Delta t}\right) \times \frac{44}{12} \tag{6}$$

Where:

$CO2_{soil,t,\Delta t}$	=	Mean estimate of carbon dioxide removal in SOC stocks between times t and $t + \Delta t$ (Mg C/ha)
$\Delta\overline{\widehat{SOC}}_{bsl,\,t,\,\Delta t}$	=	Mean predicted SOC stock change in the baseline control area between time t and time $t + \Delta t$; equal to zero where baseline control sites are not used (Mg C/ha)
$44/12$	=	Ratio of molecular weight of carbon dioxide to carbon

The variance of the mean emissions removal estimate in soil is calculated using the following equation:

$$var\left(CO2_{soil,t,\,\Delta t}\right) = \left(var\left(\overline{\Delta\widehat{SOC}}_{t,\,\Delta t}\right) + var\left(\overline{\Delta\widehat{SOC}}_{bsl,\,t,\,\Delta t}\right)\right) \times \left(\frac{44}{12}\right)^2 \tag{7}$$

Where:

$var(CO2_{soil,t,\Delta t})$	=	Variance of mean soil removals estimate $(\text{Mg C/ha})^2$
$var\left(\overline{\Delta\widehat{SOC}}_{bsl,\,t,\,\Delta t}\right)$	=	Variance of mean predicted SOC stock change across the baseline control sites between times t and $t + \Delta t$ $(\text{Mg C/ha})^2$

Where the applied methodology requires baseline control sites, $\overline{\Delta\widehat{SOC}}_{bsl,\,t,\,\Delta t}$ and its variance are calculated using Steps 11–12 on the baseline control sites rather than the project area.

The variance of the estimate of the mean removals is used to compute the uncertainty deduction in compliance with the applied methodology.[8]

For a detailed life cycle outline over the course of a project under Use Case 2, see Figure 1 and APPENDIX 2. The initial (t_0) model validation and associated DSM-MVR is reviewed by an IME at

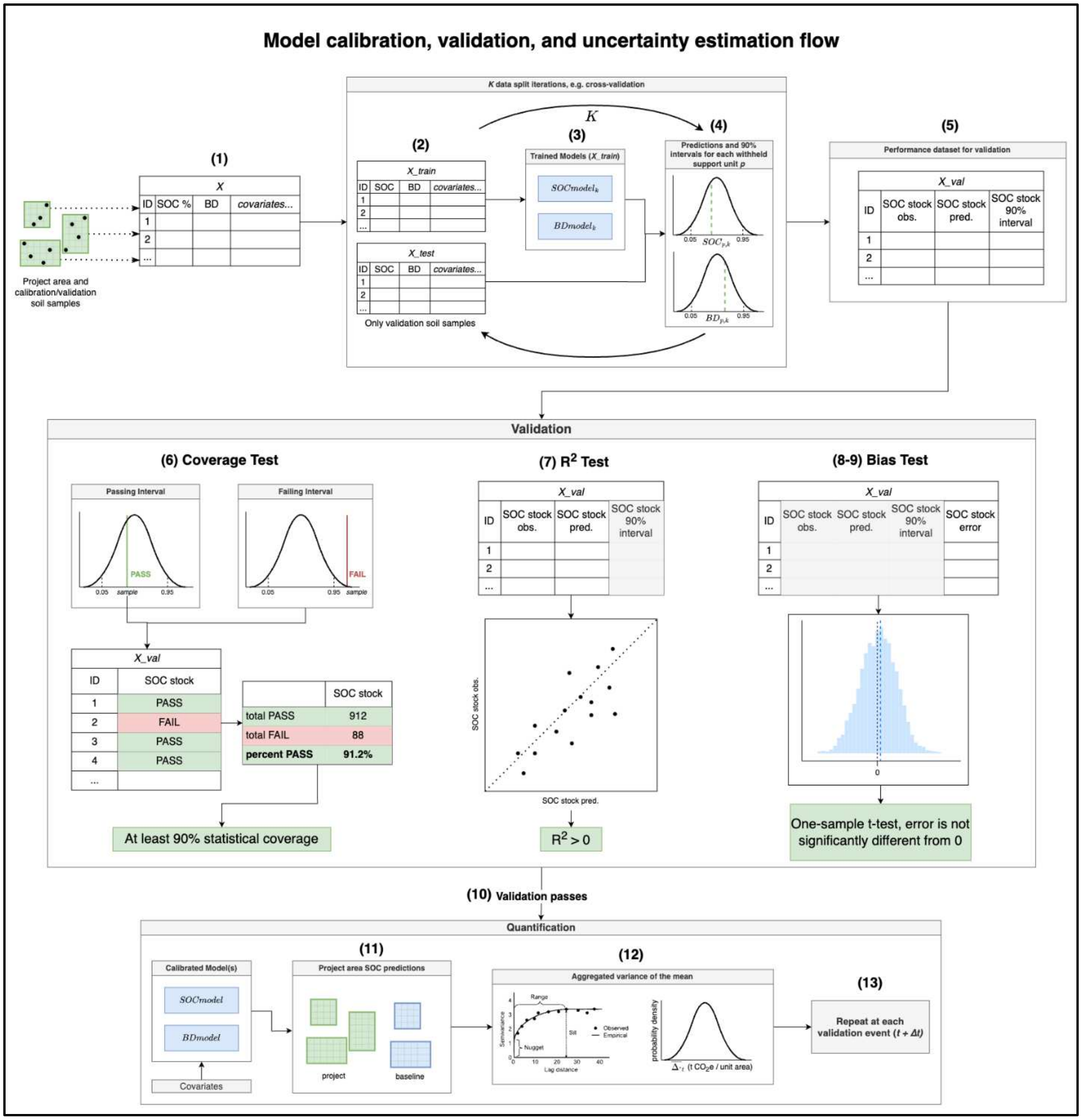
[8] For example, the probability of exceedance can be calculated directly from the variance of the estimate of mean removals and can be used with Equation (74) in *VM0042, v2.1*, or equivalent equation in the most recent version of *VM0042*.

project validation or at the first verification. Re-validations occur at least once every 5 years thereafter.

The model validation procedure described in this section must follow the sampling design of the project and any sampling requirements (e.g., for baseline control sites or permanent sampling stations) specified by the applied methodology. Where baseline control areas are required by the applied methodology, model validation must be conducted separately on the project and baseline control sites at the second and all subsequent model validation events.[9]

[9] Project and baseline scenarios are expected to be similar at the start of the project (first model validation), but after project activities commence, agricultural management practices will diverge. Therefore, each subsequent model validation and true-up must treat project and baseline control areas separately.

Figure 2. Model calibration, validation, and uncertainty estimation flow



5.1.1 Definition of the Mapped Area

The mapped area is the portion of the project area within which DSM estimates of SOC stock and stock changes are generated. This area may encompass the entire project area or a subset (e.g., if other quantification methods are applied to some locations within the project area). All prediction support units within the mapped area must be completely contained within the project boundary and free of extraneous features. For example, pixels on edges, such as field boundaries, are not contained within the mapped area. Pixels that partially or completely contain buildings, trees, waterways, or roads must be excluded from the mapped area. These exclusions do not change the project area. For example, a project area might be 10,000 hectares. After excluding edge pixels, roads, trees, buildings, and waterways, the summed area of all prediction support units in the mapped area might be 9950 hectares. The mean SOC stock density within the summation of all prediction support units is assumed to represent the 10,000 hectare project.

5.1.2 Model Architecture

Project proponents may use any statistical or machine-learning procedure to predict SOC stocks, including but not limited to traditional regression, ensemble-based regression trees, neural networks, and other methods in machine learning. Frequentist, Bayesian, parametric, and non-parametric methods are permitted. The selected model architecture may include one or more components. Examples of multicomponent models include but are not limited to:

- separately generating estimates of SOC as a percentage by mass and BD and combining these predictions to arrive at an estimate of SOC stock on an ESM basis.
- averaging the outputs of different procedures to arrive at a weighted mean estimate.
- using one method to give a mean point-prediction and another to quantify its uncertainty.
- using localized versions of the model with differing calibration weights to generate predictions for different prediction support units.
- using one method to generate predictions of SOC stock, then using the predicted value of SOC stock as a single predictor in a simple linear regression.

The selected model architecture must be described in at least one peer-reviewed journal article in any field and justified by the user in the DSM-MVR (APPENDIX 3). The DSM-MVR must describe each component and the relationships between them and must demonstrate that the model passes the validation tests in Section 5.1(10) (APPENDIX 4).

5.1.3 Covariate Selection / Feature Engineering

Peer-reviewed studies demonstrate that numerous covariates can predict SOC content, BD, and SOC stock in soils. Since more covariates are likely to emerge as appropriate predictors in the future, there is no fixed positive list of acceptable covariates, and no restrictions on their types. However, users must ensure that covariates are generated in accordance with the recommendations in Sections 5.1.3.1–5.1.3.3. A comprehensive set of peer-reviewed publications describing a wide range of covariate features for predicting SOC content, BD, or SOC stocks, including the appropriate steps to generate them and ensure their accuracy, is provided in APPENDIX 3.

A complete list of covariates, their data sources, and procedures used for data processing and feature engineering must be provided in the DSM-MVR. Examples of covariate raster data used for prediction must be made available as supplements to the DSM-MVR for VVB and/or IME review upon request.

5.1.3.1 Quality Control for Covariates Derived from Remote Sensing

Covariates from remote sensing must be processed carefully to remove features unrelated to ecosystem dynamics (e.g., cloud cover, shadows, snow, roads, trees, and waterways). The same covariate processing procedures must be used consistently during calibration, validation, and prediction.[10] Where optical remote sensing data depend on physical interpretations of reflectance, data must undergo atmospheric correction before further processing.

5.1.3.2 Time-invariant Covariates

Time-invariant (or static) covariates are those for which the data source serves as a fixed representation of these properties over the project crediting period (e.g., digital elevation models, long-term climate normals, and prior soil property and class maps).

5.1.3.3 Time-varying Covariates

Time-varying covariates depend on the target date of model calibration and prediction (e.g., summaries from optical remote sensing, such as remote sensing indices indicative of vegetation or soil characteristics, weather measurements, including temperature, precipitation, vapor pressure and humidity, and farm-practice data, such as cover cropping, reduced tillage, and other changes to land management practices). These covariates must maintain the same time interval and temporal relationship with respect to the target prediction date during calibration and prediction phases. If the model was calibrated using data from a specific period (e.g., days, weeks, or months) leading up to the target date, the same time structure and methodology must be applied when making predictions. For example, if a covariate represents the three-year mean surface temperature prior to the prediction date, users cannot change this to a two-year mean during prediction unless the model is recalibrated with the new temporal relationship.[11]

5.1.4 Calibration Data

The target variable is typically SOC content, BD, or SOC stock measured using individual soil cores or composite samples, but the tool permits use of augmented or synthetic calibration data[12] provided that all validation data follow the guidance in Section 5.2. Calibration samples may be collected within or outside the project area subject to the following constraints:

1) Samples may be collected prior to the start date of the project. All samples must be matched with covariate features that coincide in space and time with the location and sample date. Samples used for validation must be from exclusively within the project area.

---

[10] For instance, if a cloud masking method is applied, it must be used consistently whenever the covariate is used.
[11] Maintaining this consistency ensures the model correctly applies the patterns learned during calibration to new predictions.
[12] For instance, Xie et al. (2022) showed that creating synthetic calibration data using predictions from a BGCM can improve the temporal stability of DSM predictions.

2) There must be a known month and year in which the sample was collected.

3) Where the prediction support unit is the individual soil core, there must be a set of coordinates that define the collection location for the physical soil sample. These coordinates define the location of the individual sample, not the site or general location where the sample was acquired. Where the prediction support unit is an area, such as a smallholder farm, the coordinates define the area centroid.

4) All individual or composite samples must have an associated soil depth or depth range. Samples from any soil depth or depth range may be used during calibration, but only those samples at the specific soil depth or depth ranges identified by the applied methodology are used for validation. Accounting for gravel particles in the estimate of BD must follow guidance in the applied methodology.

5) Samples collected as soil cores may be depth-aligned[13] to target depths using a method described in at least one peer-reviewed publication appearing in the Web of Science: Science Citation Index (e.g., mass-preserving splines or linear interpolation; Bishop et al. 1999).

5.1.5 Treatment of Depth in the Model

The use of a surface-to-subsurface relationship from ancillary depth-profile data is not permitted. The calibrated model must treat soil depth in one of two ways:

1) The calibrated model may treat soil depth as a continuous covariate feature to predict subsurface SOC content, BD, or SOC stock (e.g., Fu et al. 2024; Ma et al. 2021; Sanderman et al. 2018).

2) Alternatively, a project proponent may choose to use independent models to generate predictions at specific soil depths or depth ranges. For example, a project proponent could choose to develop one calibrated model that predicts SOC or BD over the 0–5 cm depth range, and another calibrated model that predicts these properties over the 5–30 cm depth range. The project proponent could then combine model outputs to represent SOC and BD over the 0–30 cm depth range. Where separate models are used, prediction uncertainty from each model must be propagated through all calculations of SOC stock.

5.1.6 Quantification and Calibration at Intermediate Verification Events

Under Use Case 2, the DSM model must be recalibrated[14] (see Section 5.2.3 for sampling requirements for recalibration) when project proponents pursue verification between model

[13] Depth alignment is the process of standardizing soil core measurements to the same depths or depth increments.

[14] Recalibration reduces dependency on cumulative carbon stock change for correcting systematic errors, if any. Consider a project that sequesters 0.2 t C/ha/year above the baseline scenario. After five years, the potentially creditable quantity (ignoring uncertainty) is 1 t C/ha. Imagine that credits are issued in year 3 using a biased model that predicted only 0.4 t C/ha by year 3, which is 0.2 t C/ha less than the true value of 0.6. At year 5, when model validation and cumulative adjustments occur, the model correctly predicts 1 t C/ha, but the issuance at year 5 is 0.6 t to make up for the underestimation in year 3 (i.e., 0.4 t C/ha was issued in year 3 and an additional 0.6 t C/ha was issued in year 5). Recalibration at verification events between model validation events will reduce

validation and true-up. Model performance metrics (coverage, R^2, and bias) for the recalibration must be reported in an amendment to the initial DSM-MVR, but the DSM-MVR is not reviewed by the VVB until model validation. The project-specific recalibration data must pass the model validation tests of coverage and R^2 (see Section 5.1(10)). The model validation bias test is not required between model validation events. For detailed illustrations of various scenarios, see APPENDIX 2.

5.2 Soil Sampling

Soil samples must be collected in the project area for validation. Best practices for the collection and analysis of soil samples for VCS projects are described in applicable VCS methodologies, including Section 9 of *VM0042*. Collection of soil samples must follow procedures outlined in the applied methodology, subject to the following constraints:

1) Soil samples collected to any depth or depth range may be used for model calibration, but validation samples must match the depth or depth range in the applied methodology.
2) Soil samples collected prior to the project start date may be used for model calibration, but validation samples must be collected within six weeks[15] of the target date used for validation.
3) Soil samples collected outside the project area may be used for model calibration, including samples from publicly available sample archives, such as those generated under national or international soil inventories, but validation samples must be collected exclusively within the project boundary.
4) Composite samples may be used (e.g., for smallholder farms or replicate samples within a single prediction support unit). Where composite samples are acquired, the prediction support unit is the area within which the composite sample was acquired and the number of samples in the composite (e.g., a 1 ha field with 5 cores, or a single 10 × 10 meter pixel with 3 cores). The prediction support unit represented by a target variable must be equivalent among all calibration, validation, and prediction observations.

5.2.1 Sampling Design

Validation data must be drawn from a representative probability sample within the project area, but there are no spatial sampling design requirements for calibration data. Where a representative sample is unavailable but other data exist, the project proponent must justify treating unsampled units as “missing at random” in the DSM-MVR. The validation data must represent the spatial distribution of the project area, with the justification provided in the DSM-MVR.[16] Sampling procedures must comply with the target ALM methodology. The most effective

the potential for swings in creditable carbon disconnected from real changes. These recalibration events provide safeguards against over and under-crediting.

[15] This is consistent with general requirements for soil sampling in *VM0042*, which states that sampling must be conducted during the same season. For example, if the target date is 6 June, validation samples must be collected within the 25 April–18 July interval.

[16] The justification in the DSM-MVR should explain the sampling process used to generate the validation data, and why that process will generate representative sample.

methods for achieving a representative sample with sufficient density will depend on factors such as the project's geographic scope, environmental domains, and model performance.

1) Where new project instances are enrolled across multiple years in a grouped project design, the year in which a prediction support unit was included in the sampling design must be the top-level stratum. This requirement indicates that rolling enrollments cannot be added to existing strata.

2) Where the method selected to account for spatial covariance in the variance of mean SOC stock requires a variogram, the project proponent must ensure that a variety of inter-point distances in the 0–500 m range are adequately represented among the sample locations. Where this is not achieved by the initial probability sample, additional locations should be selected (see guidance for the use of geostatistical methods in Section 5.5).

3) The number of strata is at the discretion of the project proponent but may be defined using estimates of SOC stock, covariates, potential or realized SOC stock change, or ancillary variables not included in the model.

4) Each prediction support unit must belong to only one stratum.

5) Under Use Case 2, where baseline control sites are required by the applied methodology they must be sampled according to the requirements of the applied methodology.

6) The sample size required for model validation depends on the model prediction error and the expected average project effect (EAPE). Project proponents may use the sample size formula for a one-sample t-test to estimate the approximate sample size:

$$n_t = \left(\frac{\left(t_{\beta} + t_{\frac{\alpha}{2}}\right) \times \sigma}{EAPE} \right)^2 \tag{8}$$

Where:

n_t	=	Estimated sample size required to conduct the bias test (unitless)
t_β	=	Critical value of the t distribution for the desired type II error probability, β = 0.8
$t_{\alpha/2}$	=	Critical value of the t distribution for type I error probability, α = 0.05
σ	=	Standard deviation of the model prediction error
$EAPE$	=	Expected average project effect, an estimate of the expected mean change in SOC stocks in the first five years after project initiation; should be obtained from at least one peer-reviewed journal article or proprietary data where no peer-reviewed journal articles are available

Beyond these rules, project proponents have wide flexibility to tailor the sampling design to the conditions of the project, provided the sampling design is compliant with the applied methodology. For example, a rangeland project might constitute a contiguous area on a single soil type with little to no variation in temperature or precipitation. In this scenario, it might be appropriate to use a single level stratified random sample based on environmental covariates within the project area. Where a project area is very large or contains strong gradients in environmental variables or ALM conditions, stratified random sampling applied to the whole area is likely to be inefficient. A variety of methods to improve sampling efficiency are available, including those described in Cochran (1977) and Som (1995), such as multi-stage or hierarchical sampling.

5.2.2 Sampling Requirements under Use Case 1

All sampling and validation requirements for initial SOC stocks described above apply under Use Case 1. DSM can be used to initialize a BGCM, as outlined in Quantification Approach 1 in *VM0042*. The BGCM must be validated according to the applied methodology. Validation of the DSM estimate of SOC stocks follows the procedures specified in Section 5.1. The target date for initialization must be within six weeks of the date represented by the stock measurements, which is determined by the sampling campaign associated with enrollment for each prediction support unit.[17] Since estimates of changes in greenhouse gases (GHGs) under this approach depend on the BGCM, there are no additional sampling requirements beyond those necessary to validate the SOC stock estimate at the specified time, as well as any other sampling requirements under the applied methodology.

5.2.3 Sampling Requirements under Use Case 2

The model must be validated, by the first project verification and at least once every five years, against observations of SOC stock prior to implementation of project activities (Figure 1, APPENDIX 2). Issuance of verified carbon units (VCUs) may occur more frequently, but model calibration or model validation is required at every verification event, subject to the following constraints:

1) There are no spatial sampling design requirements for model calibration.[18]
2) The number of samples used to recalibrate the model must be sufficient to update the calibration for the current target date; at least 10% of the number of samples used in the previous model validation must be used for recalibration.
3) Model validation locations should be the same as the locations used at the prior model validation (i.e., the same prediction support units and the same locations within those

[17] The six-week time window is consistent with *VM0042*, which states that sampling and re-sampling campaigns must be conducted during the same season over time.

[18] This is because the tool evaluates model performance on an outcome basis (model validation) by comparing model predictions to independent measured values within the mapped area.

prediction support units) to minimize the impact of the spatial component of sampling error on model validation.

4) Where the project proponent does not resample all previous validation locations, the subset selected from the original locations for resampling must be identified at random.

5.3 Computing the Variance of the Average SOC Stock

Errors in DSM predictions are linked to unseen factors. Where map errors are spatially structured, estimates of the variability in mean estimates such as the standard error will significantly underestimate the precision of the mean. Geostatistical methods address this issue by quantifying and accounting for spatial dependence in the errors.

The contribution of spatial correlation in the variance of the SOC stock must be addressed. Project proponents may implement the methods described by Wadoux and Heuvelink (2023) using the steps below (see also APPENDIX 6 and APPENDIX 7), which uses a Monte Carlo approximation to estimate the variance of the prediction error of the spatial average, or any valid method described in APPENDIX 7 that accounts for spatial covariance.

1) Compute the standardized prediction error at every validation prediction support unit using the following equation:

$$\varepsilon_{i,standardized} = \frac{\varepsilon_i}{\sigma_i} \quad (9)$$

Where:

$\varepsilon_{i,standardized}$	=	Standardized prediction error for prediction support unit *i*
ε_i	=	Prediction error for prediction support unit *i*
σ_i	=	Predictive standard deviation of the model for prediction support unit *i*; must be generated by the same process used to evaluate the coverage test described in Section 5.1

2) Compute the spatial correlation function by fitting a variogram to the standardized prediction errors and transforming the variogram's predictions into a correlation function. Section 5.5 provides guidance on fitting variograms.

$$\rho(h) = \frac{sill - \gamma(h)}{sill} \quad (10)$$

Where:

$\rho(h)$	=	Correlation function of the standardized model prediction error at lag distance *h*
$\gamma(h)$	=	Semivariance for a pair of points separated by lag distance *h*
sill	=	Value of the semivariance at the range in the dataset

3) For each Monte Carlo draw, randomly select a pair of prediction locations s and u, and compute the covariance between them using the following equation:

$$cov(\sigma_s, \sigma_u) = \sigma_s \times \sigma_u \times \rho(|s-u|) \tag{11}$$

Where:

σ_s	=	Predictive standard deviation of the model at location s
σ_u	=	Predictive standard deviation of the model at location u
$\rho(\|s-u\|)$	=	Correlation of model prediction error at the lag distance separating points s and u

4) Compute the variance of the mean SOC stock using the following equation:

$$var\left(\overline{\widehat{SOC_t}}\right) = \frac{1}{L}\sum_{l=1}^{L} \sigma_{s,l} \cdot \sigma_{u,l} \cdot \rho(|s_l, u_l|) \tag{12}$$

1) Where:

$var\left(\overline{\widehat{SOC_t}}\right)$	=	Variance of the prediction error of the spatial average
$\sigma_{s,l}$	=	Predictive standard deviation of the model at location s selected in sample l
$\sigma_{u,l}$	=	Predictive standard deviation of the model at location u selected in sample l
$\rho(\|s_l,u_l\|)$	=	Correlation of the model prediction errors at the lag distances separating points s and u in sample l
L	=	1, 2, …, L Monte Carlo samples

5) Confirm that the Monte Carlo simulation generated a precise estimate of the variance. The precision of the variance will increase with the number of Monte Carlo samples, L. Project proponents must demonstrate that the variance of the mean GHG emission removal estimate in soil has been calculated with sufficient precision such that imprecision in estimates of the terms in Equation (5) impacts the uncertainty deduction by less than ±1 percentage point. This can be achieved by repeating the Monte Carlo process a large number of times and showing that the uncertainty deduction is fluctuating less than ±1 percentage point (see APPENDIX 6).

5.4 Computing the Variance of the Change in SOC Stock

The variance of the mean stock change estimate is calculated using Equation (5). This calculation accounts for the covariance term explicitly using an estimate of ρ. This parameter can be estimated in one of two ways.

1) Where measurements of the standardized prediction error are available at the same model validation locations at more than one time (i.e. there are repeated validation measurements at the same locations), use the following cross variogram procedure:[19]

$$\gamma_{01}(h) = \frac{1}{2}E\left[\left(\varepsilon_0(s+h) - \varepsilon_0(s)\right)\cdot\left(\varepsilon_1(s+h) - \varepsilon_1(s)\right)\right] \quad (13)$$

Where:

$\gamma_{01}(h)$	=	Semivariance for the erros, ε_0 and ε_1, separated by distance h
ε_0	=	Model prediction error at time 0
ε_1	=	Model prediction error at time 1

2) Where measurements of the standardized prediction error are available at more than one time, but not at the same validation locations, use the following pseudo cross variogram procedure:[20]

$$\pi_{0,1}(h) = \frac{1}{2}E\left[\left(\varepsilon_0(\mathrm{s}+\mathrm{h}) - \varepsilon_1(s)\right)^2\right] \quad (14)$$

Where:

$\pi_{0,1}(h)$	=	Semivariance obtained with the pseudo cross variogram for in the errors ε_0 and ε_1, separated by distance h
$\varepsilon_0(s + h)$	=	Model prediction error at location s at time 0
$\varepsilon_1(s)$	=	Model prediction error at location s at time 1

After the cross variogram or pseudo cross variogram is estimated, ρ can be obtained using Equation (10).

5.5 Guidance on Variogram Selection and Fitting

Project proponents must compare multiple variogram functions to ensure that the variogram has been correctly estimated (e.g., spherical, Gaussian, and nugget-effect). Publicly available software is available to fit variograms. Project proponents must describe the software and version used in the DSM-MVR. A candidate variogram model should be selected using visual interpretation of the fitted variogram and may include model selection criteria (e.g., Akaike's information criterion, Bayesian information criterion).

The DSM-MVR must include a plot of the selected variogram and several candidate alternatives, and must include numerical values of model selection criteria, where used. Nested variogram models are permitted, where different functions are used to estimate the spatial autocorrelation of standardized prediction errors at different lag distances. Nested variograms

[19] Adapted from Equation 20.10 in Wackernagel (2003, p. 147)
[20] Adapted from Equation 20.18 in Wackernagel (2003, p. 149)

may be geographically stratified, such that short-distance lags are handled differently in different parts of the project area.

5.5.1 Sampling Guidelines for Variogram Calculations

Sampling guidelines must ensure that distances in the 0–500 m range are adequately represented. Webster and Oliver (1992) and Kerry and Oliver (2007) provide guidance on sample sizes and the spatial proximity of samples that must be followed. At least 100 points are needed for variogram estimation, but more points will generally improve the robustness of the variogram.

Fitting a variogram involves calculating a distance matrix containing the distances between all pairs of points. The variogram is then fitted to these distances within specific bins, such as 0–10 m, 10–20 m, and so on. A key consideration is determining the number of bins and the width of each bin. This can be approached in various ways, including using bins with an even number of pairs, bins of uniform width, or bins that minimize variation in distance.

If bin intervals are too small, semivariance estimates will be based on few point pairs, leading to imprecise estimates of the variogram. Conversely, if bin intervals are too wide, the reduced number of intervals may limit the ability to accurately estimate the range of spatial correlation. Use the following principles to guide the selection of appropriate bin numbers and widths:

1) Each bin should contain at least 50 pairs of locations (Schabenberger and Gotway 2017).
2) The distance represented by the center of the first bin must be smaller than the estimated range of the variogram.
3) The distance represented by the center of the last bin should be half the maximum distance among all possible locations (Cressie 1985).

For additional guidance related to bin width and number, see Section 3.2.3 in Oliver and Webster (2015).

6 Data and Parameters

6.1 Data and Parameters Available at Validation

Data/Parameter	$\widehat{SOC}_{i,t}$
Data unit	Mg C/ha
Description	Predicted SOC stock in prediction support unit *i* at time *t*

Equations	(1), (3)
Source of data	Calibrated DSM model
Value applied	N/A
Justification of choice of data or description of measurement methods and procedures applied	This is the predicted value from the calibrated DSM model applied to covariates at prediction support unit i at time t.
Purpose of data	Calculation of project emissions
Comments	N/A

Data/Parameter	A_i
Data unit	Area (e.g., hectares or acres)
Description	Area of prediction support unit i within the project area
Equations	(3)
Source of data	Prediction support unit i
Value applied	N/A
Justification of choice of data or description of measurement methods and procedures applied	The area of each prediction support unit is used to account for the possibility that not all prediction support units have the same area.
Purpose of data	Calculation of project emissions
Comments	N/A

Data/Parameter	$\overline{SOC}_t$
Data unit	Mg C/ha
Description	Mean of model predictions of SOC stock over all prediction support units at time *t*
Equations	(4)
Source of data	Calibrated DSM model
Value applied	N/A
Justification of choice of data or description of measurement methods and procedures applied	This is the predicted mean value from the calibrated DSM model applied to covariates at time *t*.
Purpose of data	Calculation of project emissions
Comments	N/A

Data/Parameter	$\overline{SOC}_{t,\Delta t}$
Data unit	Mg C/ha
Description	Mean of model predictions of SOC stock over all prediction support units at time $t + \Delta t$
Equations	(4), (6)
Source of data	Calibrated DSM model
Value applied	N/A

Justification of choice of data or description of measurement methods and procedures applied	This is the predicted mean value from the calibrated DSM model applied to covariates at time t + Δt.
Purpose of data	Estimation of SOC stock and changes over time
Comments	N/A

Data/Parameter	$\overline{\Delta SOC}_{bsl,t,\Delta t}$
Data unit	Mg C/ha
Description	Mean predicted SOC stock change in the baseline control area between time t and time t + Δt
Equations	(6)
Source of data	Calibrated DSM model
Value applied	N/A
Justification of choice of data or description of measurement methods and procedures applied	This is the mean emissions removal estimate in the baseline control area time t and time t + Δt.
Purpose of data	Calculation of project emissions
Comments	Equal to zero where baseline control sites are not used.

Data/Parameter	$CO2_{soil,t,\Delta t}$
Data unit	CO_2e
Description	The mean emissions removal estimate for soil.

Equations	(6)
Source of data	Calibrated DSM model
Value applied	N/A
Justification of choice of data or description of measurement methods and procedures applied	This is the mean emissions removal estimate in soil in units of CO_2e. It is the change in CO2e in the project area between time t and t + Δt minus the corresponding change in the baseline control area.
Purpose of data	Calculation of project emissions
Comments	N/A

Data/Parameter	EAPE
Data unit	Mg C/ha
Description	Estimated average project effect
Equations	(8)
Source of data	Peer-reviewed literature or proprietary data
Value applied	Where no estimate is available, a value of 1.5 may be used.
Justification of choice of data or description of measurement methods and procedures applied	An estimate from at least one peer-reviewed journal article (or proprietary data where no peer-reviewed journal articles are available) of the mean change in SOC stocks in the first five years after project initiation
Purpose of data	Guidance on sample size for DSM model validation
Comments	DSM model validation must occur by the first project verification and at least once every five years.

Data/Parameter	σ_s
Data unit	unitless
Description	Predictive standard deviation of the model at location s
Equations	(11)
Source of data	Any valid method to compute the predictive standard deviation
Value applied	N/A
Justification of choice of data or description of measurement methods and procedures applied	The predictive standard deviation is used in combination with the spatial covariance of the standardized prediction error to estimate the variance of the mean using geostatistical methods.
Purpose of data	Calculation of project emissions
Comments	N/A

Data/Parameter	σ_u
Data unit	unitless
Description	Predictive standard deviation of the model at location s
Equations	(11)
Source of data	Any valid method to compute the predictive standard deviation
Value applied	N/A

Justification of choice of data or description of measurement methods and procedures applied	The predictive standard deviation is used in combination with the spatial covariance of the standardized prediction error to estimate the variance of the mean using geostatistical methods.
Purpose of data	Calculation of project emissions
Comments	N/A

6.2 Data and Parameters Monitored

Data/Parameter	$\widehat{SOC}_{i,t}$
Data unit	Mg C/ha
Description	Predicted SOC stock in prediction support unit i at time t
Equations	(1), (2)
Source of data	Calibrated DSM model
Value applied	N/A
Justification of choice of data or description of measurement methods and procedures applied	This is the predicted value from the calibrated DSM model applied to covariates at prediction support unit i at time t.
Purpose of data	Calculation of project emissions
Comments	NA

Data/Parameter	$SOC_{obs,i,t}$
Data unit	Mg C/ha
Description	Observed SOC stock in prediction support unit *i* at time *t*
Equations	(1), (2)

Source of data	Measurement of SOC content and BD from a soil core or composite sample
Value applied	N/A
Justification of choice of data or description of measurement methods and procedures applied	Guidance on measurement techniques must follow the applied methodology.
Purpose of data	Calibration and validation of the DSM model
Comments	DSM model validation must occur by the first project verification and subsequently at least once every five years.

7 References

Adhikari, K., Hartemink, A.E., Minasny, B., Bou Kheir, R., Greve, M.B., Greve, M.H., 2014. Digital Mapping of Soil Organic Carbon Contents and Stocks in Denmark. PLoS ONE 9, e105519. https://doi.org/10.1371/journal.pone.0105519

Bishop, T.F.A., McBratney, A.B., Laslett, G.M., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91, 27–45. https://doi.org/10.1016/S0016-7061(99)00003-8

Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B., 2019. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. ISPRS Journal of Photogrammetry and Remote Sensing 147, 267–282. https://doi.org/10.1016/j.isprsjprs.2018.11.026

Cochran, W.G., 1977. Sampling Techniques, 3rd Edition, 3rd edition. ed. John Wiley & Sons, New York.

Cressie, N., 1985. Fitting variogram models by weighted least squares. Mathematical Geology 17, 563–586. https://doi.org/10.1007/BF01032109

De Bruin, S., Brus, D.J., Heuvelink, G.B.M., Van Ebbenhorst Tengbergen, T., Wadoux, A.M.J.-C., 2022. Dealing with clustered samples for assessing map accuracy by cross-validation. Ecological Informatics 69, 101665. https://doi.org/10.1016/j.ecoinf.2022.101665

Dvorakova, K., Heiden, U., van Wesemael, B., 2021. Sentinel-2 Exposed Soil Composite for Soil Organic Carbon Prediction. Remote Sensing 13, 1791. https://doi.org/10.3390/rs13091791

Dvorakova, K., Shi, P., Limbourg, Q., Van Wesemael, B., 2020. Soil Organic Carbon Mapping from Remote Sensing: The Effect of Crop Residues. Remote Sensing 12, 1913. https://doi.org/10.3390/rs12121913

Fu, P., Clanton, C., Demuth, K.M., Goodman, V., Griffith, L., Khim-Young, M., Maddalena, J., LaMarca, K., Wright, L.A., Schurman, D.W., Kellner, J.R., 2024. Accurate Quantification of 0–30 cm Soil Organic Carbon in Croplands over the Continental United States Using Machine Learning. Remote Sensing 16, 2217. https://doi.org/10.3390/rs16122217

Gomes, L.C., Faria, R.M., De Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Filho, E.I.F., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350. https://doi.org/10.1016/j.geoderma.2019.01.007

Gomez, C., Viscarra Rossel, R.A., McBratney, A.B., 2008. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. Geoderma 146, 403–411. https://doi.org/10.1016/j.geoderma.2008.06.011

Hengl, T., Jesus, J.M. de, Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLOS ONE 12, e0169748. https://doi.org/10.1371/journal.pone.0169748

Heuvelink, G.B.M., Angelini, M.E., Poggio, L., Bai, Z., Batjes, N.H., Van Den Bosch, R., Bossio, D., Estella, S., Lehmann, J., Olmedo, G.F., Sanderman, J., 2021. Machine learning in space and time for modelling soil organic carbon change. European J Soil Science 72, 1607–1623. https://doi.org/10.1111/ejss.12998

Janssen, P.H.M., Heuberger, P.S.C., 1995. Calibration of process-oriented models. Ecological Modelling 83, 55–66. https://doi.org/10.1016/0304-3800(95)00084-9

Jenny, H., 1941. Factors of Soil Formation, A System of Quantitative Pedology. McGraw-Hill, New York.

Kerry, R., Oliver, M.A., 2007. Comparing sampling needs for variograms of soil properties computed by the method of moments and residual maximum likelihood. Geoderma 140, 383–396. https://doi.org/10.1016/j.geoderma.2007.04.019

Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., Walter, C., 2014. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma 213, 296–311. https://doi.org/10.1016/j.geoderma.2013.07.002

Ma, Y., Minasny, B., McBratney, A., Poggio, L., Fajardo, M., 2021. Predicting soil properties in 3D: Should depth be a covariate? Geoderma 383, 114794. https://doi.org/10.1016/j.geoderma.2020.114794

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

McRoberts, R.E., Næsset, E., Saatchi, S., Quegan, S., 2022. Statistically rigorous, model-based inferences from maps. Remote Sensing of Environment 279, 113028. https://doi.org/10.1016/j.rse.2022.113028

Oliver, M.A., Webster, R., 2015. Basic Steps in Geostatistics: The Variogram and Kriging, SpringerBriefs in Agriculture. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-15865-5

Poggio, L., de Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL 7, 217–240. https://doi.org/10.5194/soil-7-217-2021

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., Thompson, J., 2018. Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution. Soil Sci. Soc. Am. j. 82, 186–201. https://doi.org/10.2136/sssaj2017.04.0122

Ratnayake, R.R., Karunaratne, S.B., Lessels, J.S., Yogenthiran, N., Rajapaksha, R.P.S.K., Gnanavelrajah, N., 2016. Digital soil mapping of organic carbon concentration in paddy growing soils of Northern Sri Lanka. Geoderma Regional 7, 167–176. https://doi.org/10.1016/j.geodrs.2016.03.002

Sanderman, J., Hengl, T., Fiske, G., Solvik, K., Adame, M.F., Benson, L., Bukoski, J.J., Carnell, P., Cifuentes-Jara, M., Donato, D., Duncan, C., Eid, E.M., Ermgassen, P.Z., Lewis, C.J.E., Macreadie, P.I., Glass, L., Gress, S., Jardine, S.L., Jones, T.G., Nsombo, E.N., Rahman, M.M., Sanders, C.J., Spalding, M., Landis, E., 2018. A global map of mangrove forest soil carbon at 30 m spatial resolution. Environ. Res. Lett. 13, 055002. https://doi.org/10.1088/1748-9326/aabe1c

Schabenberger, O., Gotway, C.A., 2017. Statistical Methods for Spatial Data Analysis: Texts in Statistical Science, 1st ed. Chapman and Hall/CRC. https://doi.org/10.1201/9781315275086

Som, R.K., 1995. Practical Sampling Techniques, 0 ed. CRC Press. https://doi.org/10.1201/9781482273465

Sothe, C., Gonsamo, A., Arabian, J., Snider, J., 2022. Large scale mapping of soil organic carbon concentration with 3D machine learning and satellite observations. Geoderma 405, 115402. https://doi.org/10.1016/j.geoderma.2021.115402

Szatmári, G., Pásztor, L., Heuvelink, G.B.M., 2021. Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics. Geoderma 403, 115356. https://doi.org/10.1016/j.geoderma.2021.115356

Venter, Z.S., Hawkins, H.-J., Cramer, M.D., Mills, A.J., 2021. Mapping soil organic carbon stocks and trends with satellite-driven high resolution maps over South Africa. Science of The Total Environment 771, 145384. https://doi.org/10.1016/j.scitotenv.2021.145384

Wackernagel, H., 2003. Multivariate Geostatistics. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-05294-5

Wadoux, A.M.J. -C., Heuvelink, G.B.M., 2023. Uncertainty of spatial averages and totals of natural resource maps. Methods Ecol Evol 14, 1320–1332. https://doi.org/10.1111/2041-210X.14106

Wadoux, A.M.J.-C., Walvoort, D.J.J., Brus, D.J., 2022. An integrated approach for the evaluation of quantitative soil maps through Taylor and solar diagrams. Geoderma 405, 115332. https://doi.org/10.1016/j.geoderma.2021.115332

Webster, R., Oliver, M.A., 1992. Sample adequately to estimate variograms of soil properties. Journal of Soil Science 43, 177–192. https://doi.org/10.1111/j.1365-2389.1992.tb00128.x

Xie, E., Zhang, X., Lu, F., Peng, Y., Chen, J., Zhao, Y., 2022. Integration of a process-based model into the digital soil mapping improves the space-time soil organic carbon modelling in intensively human-impacted area. Geoderma 409, 115599. https://doi.org/10.1016/j.geoderma.2021.115599

Zhou, T., Geng, Y., Chen, J., Liu, M., Haase, D., Lausch, A., 2020. Mapping soil organic carbon content using multi-source remote sensing variables in the Heihe River Basin in China. Ecological Indicators 114, 106288. https://doi.org/10.1016/j.ecolind.2020.106288

Zhou, Y., Chartin, C., Van Oost, K., Van Wesemael, B., 2022. High-resolution soil organic carbon mapping at the field scale in Southern Belgium (Wallonia). Geoderma 422, 115929. https://doi.org/10.1016/j.geoderma.2022.115929

APPENDIX 1: ASSESSMENT BY INDEPENDENT MODELING EXPERT (IME)

This appendix describes the step-wise process for the assessment of a DSM model by an IME according to the following steps:

2) The project proponent must generate a DSM-MVR to demonstrate the validity and use of the DSM model consistent with the guidance and requirements of the CN0137 tool (Appendix 4).

3) A VVB must contract an IME to review the DSM-MVR. The VVB may select an IME approved by Verra or contract a new IME. New IMEs must fulfill minimum qualifications defined by Verra (see "Minimum IME qualifications" below).

4) The IME must assess the DSM-MVR and generate an IME assessment report that:

    a. Assesses the quality of calibration data (e.g., soil core measurements, environmental covariates, remote sensing data) and the overall measurement uncertainty;

    b. Confirms that the calibration procedure meets the requirements stated in Section 5.1;

    c. Confirms that the samples used for validation follow the guidelines in Sections 5.1 and 5.2. The assessment report must explicitly confirm that validation samples were acquired exclusively within the project area, that the depth or depth range of validation samples matches the depth or depth range of the target methodology, and that the validation samples are a representative probability sample of the project area or that unsampled units can be treated as “missing at random” as defined in Section 5.2.1.

    d. Confirms that the model passes the three validation tests: coverage, goodness of fit, and bias, as described in Section 5.1. Where DSM has been used to initialize and/or true-up a biogeochemical model (BGCM), the IME must confirm that the uncertainty propagation procedures in APPENDIX 5 have been correctly implemented.

    Project proponents must promptly respond to inquiries and requests for consultation from the IME, including submission of additional documentation. The burden of proof in the IME assessment process rests with the project proponent.

5) The IME assessment report must be submitted to the VVB for approval alongside other project documentation. The IME must keep the VVB apprised of questions and resolved findings related to the DSM-MVR, and should provide documentation to the VVB justifying its recommendation. The VVB has ultimate responsibility for approval of the DSM-MVR.

All DSM-MVRs and IME assessment reports will be made public alongside project documentation in the Verra registry.

Minimum Qualifications of IMEs

Verra defines minimum qualifications that IMEs must fulfill to perform an evaluation of the use of DSM models following CN0137 tool guidance. To provide an assessment report, the IME – an individual – must meet the following criteria:

1) Demonstrated competence in quantifying SOC stocks and/or stock changes using DSM. Specialization in certain practices, land uses, and regional/country expertise may be relevant. The IME must have at least five years of relevant work experience.
2) Stated ability to assess DSM model types based on demonstrated use of statistical and/or machine learning procedures for DSM. Prospective IMEs may demonstrate expertise through peer-reviewed scientific publication(s) appearing in the Web of Science: Science Citation Index, or by submitting relevant project reports.
3) Demonstrated ability to assess uncertainty in DSM predictions, including methods to account for spatial covariance in model prediction errors.
4) Demonstrated freedom from conflict of interest. This must be established by disclosing all relevant organizational and financial affiliations that could potentially undermine the integrity of the IME review process.
5) Recommendation by two references, preferably research scientists with public, private, or government affiliations, including but not limited to academia.

The IME Qualification Form must be used to provide evidence demonstrating that the IME meets the above criteria. *The IME Qualification Form will be provided on the Verra webpage if this tool is approved and becomes active in the VCS program.*

APPENDIX 2: ILLUSTRATIVE SCENARIOS FOR PROJECT QUANTIFICATION LIFECYCLE

In addition to the example shown in Figure 1, the following illustrative scenarios are based on common ALM project structures. These scenarios indicate a variety of non-exhaustive possible options for how VCS program guidelines (project validation, project verification) integrate with guidelines in the tool related to model validation, model revalidation and true-up, and optional intermediate verification. Note that the model re-calibration between model re-validation periods is only necessary when using DSM in a measure re-measure approach, for example *VM0042* Quantification Approach 2.

SCENARIO 1:
> Not Grouped
> Model Validation at Project Validation
> Verification Every 2-3 Years

	Y0	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Project Validation											
Model Validation Soil Samples Collected	1000**					1000					1000
Model Validation (optional re-calibration) (year validation soil data was collected)	(Y0 val. data)					(Y5 val. data)					(Y10 val. data)
Cumulative carbon stock change adjustment											
Model Re-Calibration Soil Samples Collected				100				100			
Model Re-Calibration (year re-calibration soil data was collected)				(Y3 cal. data)				(Y7 cal. data)			
Documentation Required	PDD DSM-MVR			MR*		MR DSM-MVR		MR*			MR DSM-MVR
VVB Audit											
IME Assessment											

* Monitoring Reports (MRs) must report re-calibration statistics in re-calibration years.
** Sample numbers are illustrative. Actual number of soil samples collected should follow the guidance in Section 5.

SCENARIO 2:
> Not Grouped
> Model Validation at First Verification
> Verification Every 2-3 Years

	Y0	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Project Validation											
Model Validation Soil Samples Collected											
Model Validation (optional re-calibration) (year validation soil data was collected)				(Y0 val. data)		(Y5 val. data)					(Y10 val. data)
Cumulative carbon stock change adjustment											
Model Re-Calibration Soil Samples Collected											
Model Re-Calibration (year re-calibration soil data was collected)				(Y3 cal. data)				(Y7 cal. data)			
Documentation Required	PDD			MR* DSM-MVR		MR DSM-MVR		MR*			MR DSM-MVR
VVB Audit											
IME Assessment											

* Monitoring Reports (MRs) must report re-calibration statistics in re-calibration years.

SCENARIO 3:
> Grouped Project
> Model Validation at Project Validation
> Verification Every 5 Years

	Y0	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Project Validation											
Project Instance 1 Start Date											
Project Instance 2 Start Date											
Project Instance 3 Start Date											
New Project Instances Added to Project (at next project verification event)						(project instances 2 & 3)					
Model Validation Soil Samples Collected (which project instance was sampled)	(project instance 1)	(project instance 2)	(project instance 3)			(all project instances)					(all project instanc
Model Validation (optional re-calibration) (year validation soil data was collected)	(project instance 1 start year val. data)					(project instance 2 & 3 start year data and all Y5 val. data)					(Y10 val. data)
Cumulative carbon stock change adjustment											
Model Re-Calibration Soil Samples Collected											
Model Re-Calibration (year re-calibration soil data was collected)											
Documentation Required	PDD DSM-MVR					MR* DSM-MVR					MR DSM-MVR
VVB Audit											
IME Assessment											

Model validation soil samples must be collected in the start year for every new project instance. Model performance must be validated against the start year soil data for each new project instance. This means that for any verification event during which the project is adding new instances, the model must go through a full model validation exercise and update the DSM-MVR to demonstrate the model is valid in the new areas added to the project. Note that because new project instances can only be added at a project verification event, therefore in this example the new project instances are all added at the Y5 project verification event.
* Monitoring Reports (MRs) must report re-calibration statistics in re-calibration years.

SCENARIO 4:
> Grouped Project
> Model Validation at Project Validation
> Verification Every 2-3 Years

	Y0	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Project Validation											
Project Instance 1 Start Date											
Project Instance 2 Start Date											
Project Instance 3 Start Date											
New Project Instances Added to Project (at next project verification event)				(project instances 2 & 3)							
Model Validation Soil Samples Collected (which project instance was sampled)	(project instance 1)	(project instance 2)	(project instance 3)			(all project instances)					(all project instanc
Model Validation (optional re-calibration) (year validation soil data was collected)	(project instance 1 start year val. data)			(project instance 2 & 3 start year val. data)		(Y5 val. data)					(Y10 val. data)
Cumulative carbon stock change adjustment											
Model Re-Calibration Soil Samples Collected				(all project instances)				(all project instances)			
Model Re-Calibration (year re-calibration soil data was collected)				(Y3 cal. data)				(Y7 cal. data)			
Documentation Required	PDD DSM-MVR			MR* DSM-MVR		MR DSM-MVR		MR*			MR DSM-MVR
VVB Audit											
IME Assessment											

Model validation soil samples must be collected in the start year for every new project instance. Model performance must be validated against the start year soil data for each new project instance. This means that for any verification event during which the project is adding new instances, the model must go through a full model validation exercise and update the DSM-MVR to demonstrate the model is valid in the new areas added to the project. Note that because new project instances can only be added at a project verification event, in this example the project would need to undergo verification in Y3, Y5, Y7, and Y10.
* Monitoring Reports (MRs) must report re-calibration statistics in re-calibration years.

SCENARIO 5:
> Grouped Project
> Model Validation at First Verification
> Verification Every 2-3 Years

	Y0	**Y1**	**Y2**	**Y3**	**Y4**	**Y5**	**Y6**	**Y7**	**Y8**	**Y9**	**Y10**
Project Validation											
Project Instance 1 Start Date											
Project Instance 2 Start Date											
Project Instance 3 Start Date											
New Project Instances Added to Project (at next project verification event)				(project instances 2 & 3)							
Model Validation Soil Samples Collected (which project instance was sampled)	(project instance 1)	(project instance 2)	(project instance 3)			(all project instances)					(all project instances)
Model Validation (optional re-calibration) (year validation soil data was collected)				(project instances 1, 2, & 3 start year val. data)		(Y5 val. data)					(Y10 val. data)
Cumulative carbon stock change adjustment											
Model Re-Calibration Soil Samples Collected				(all project instances)				(all project instances)			
Model Re-Calibration (year re-calibration soil data was collected)				(Y3 cal. data)				(Y7 cal. data)			
Documentation Required	PDD			MR* DSM-MVR		MR DSM-MVR		MR*			MR DSM-MVR
VVB Audit											
IME Assessment											

Model validation soil samples must be collected in the start year for every new project instances. Model performance must be validated against the start year soil data for each new project instance. This means that for any verification event during which the project is adding new instances, the model must go through a full model validation exercise and update the DSM-MVR to demonstrate the model is valid in the new areas added to the project. Note that because new project instances can only be added at a project verification event, in this example the project would need to undergo verification in Y3, Y5, Y7, and Y10.
* Monitoring Reports (MRs) must report re-calibration statistics in re-calibration years.

SCENARIO 6:
> Grouped Project with long enrollment period
> Model Validation at First Verification
> Verification Every 2-3 Years

	Y0	**Y1**	**Y2**	**Y3**	**Y4**	**Y5**	**Y6**	**Y7**	**Y8**	**Y9**	**Y10**
Project Validation											
Project Instance 1 Start Date											
Project Instance 2 Start Date											
Project Instance 3 Start Date											
Project Instance 4 Start Date											
Project Instance 5 Start Date											
New Project Instances Added to Project (at next project verification event)				(project instances 2 & 3)		(project instances 4 & 5)					
Model Validation Soil Samples Collected (which project instance was sampled)	(project instance 1)	(project instance 2)	(project instance 3)	(project instance 4)	(project instance 5)	(all project instances)					(all project instances)
Model Validation (optional re-calibration) (year validation soil data was collected)				(project instances 1, 2, & 3 start year val. data)		(project instances 4 & 5 start year val. data, and all Y5 val. data)					(Y10 val. data)
Cumulative carbon stock change adjustment											
Model Re-Calibration Soil Samples Collected				(project instances 1-3)				(all project instances)			
Model Re-Calibration (year re-calibration soil data was collected)				(Y3 cal. data)				(Y7 cal. data)			
Documentation Required	PDD			MR* DSM-MVR		MR DSM-MVR		MR*			MR DSM-MVR
VVB Audit											
IME Assessment											

Model validation soil samples must be collected in the start year for every new project instance. Model performance must be validated against the start year soil data for each new project instance. This means that for any verification event during which the project is adding new instances, the model must go through a full model validation exercise and update the DSM-MVR to demonstrate the model is valid in the new areas added to the project. Note that because new project instances can only be added at a project verification event, in this example the project would need to undergo verification in Y3, Y5, Y7, and Y10.
* Monitoring Reports (MRs) must report re-calibration statistics in re-calibration years.

APPENDIX 3: MODEL ARCHITECTURES AND COVARIATES

More than 1000 peer-reviewed academic journal articles have been written in the field of DSM with an explicit focus on SOC content, BD, or SOC stock, and how these quantities change over time. Sixteen influential publications that include a wide range of model architectures and covariates are summarized below.

McBratney, A.B., Mendonça Santos, M.L., Minasny, B., 2003. On digital soil mapping. Geoderma 117, 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

Proposes a framework for digital soil mapping. This review discusses a wide range of model architectures, including generalized linear models, classification and regression trees, neural networks, fuzzy systems, and geostatistical tools. The authors generalize the state factor framework originally developed by Jenny (1941) by discussing covariates related to soil, climate, organisms, parent material, age and spatial or geographic position.

Gomez, C., Viscarra Rossel, R.A., McBratney, A.B., 2008. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. Geoderma 146, 403–411. https://doi.org/10.1016/j.geoderma.2008.06.011

Combines partial least squares regression with visible and near infrared spectral data from proximal and spaceborne sensors to predict SOC as a percentage by mass in northwestern New South Wales, Australia. Spaceborne remote sensing was from the NASA Hyperion sensor on the EO-1 spacecraft. This sensor provided coverage of the 400 – 2,500 nm spectral range using a spectral sampling interval of approximately 10 nm.

Adhikari, K., Hartemink, A.E., Minasny, B., Bou Kheir, R., Greve, M.B., Greve, M.H., 2014. Digital Mapping of Soil Organic Carbon Contents and Stocks in Denmark. PLoS ONE 9, e105519. https://doi.org/10.1371/journal.pone.0105519

Uses regression kriging, a geostatistical technique, to predict SOC content, BD and SOC stock in Denmark. The authors used 18 continuous and categorical predictors related to land use and soils, hydrology, surface topography, climate and solar insolation.

Lacoste, M., Minasny, B., McBratney, A., Michot, D., Viaud, V., Walter, C., 2014. High resolution 3D mapping of soil organic carbon in a heterogeneous agricultural landscape. Geoderma 213, 296–311. https://doi.org/10.1016/j.geoderma.2013.07.002

Uses Cubist, a rule-based regression method, to predict SOC content and BD in cropland areas of France. Covariates included topographic properties derived from a high-resolution lidar digital elevation model, geological and land use data, and a map of A-horizon thickness.

Ratnayake, R.R., Karunaratne, S.B., Lessels, J.S., Yogenthiran, N., Rajapaksha, R.P.S.K., Gnanavelrajah, N., 2016. Digital soil mapping of organic carbon concentration in paddy growing soils of Northern Sri Lanka. Geoderma Regional 7, 167–176. https://doi.org/10.1016/j.geodrs.2016.03.002

Uses linear mixed models to predict SOC content in rice cropping in Sri Lanka. The study used topographic climatic, biological and spatial covariates, including satellite remote sensing from the Landsat sensor.

Hengl, T., Jesus, J.M. de, Heuvelink, G.B.M., Gonzalez, M.R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G.B., Ribeiro, E., Wheeler, I., Mantel, S., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLOS ONE 12, e0169748. https://doi.org/10.1371/journal.pone.0169748

Uses random forest, gradient boosting, and multinomial logistic regression, to map SOC content, bulk density, SOC stocks, and other soil properties globally up to a depth of 30 cm. Covariates included numerous climate variables, topographic measurements from a digital elevation model, and land cover information from global satellite products.

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S., Thompson, J., 2018. Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution. Soil Sci. Soc. Am. j. 82, 186–201. https://doi.org/10.2136/sssaj2017.04.0122

Uses random forest to predict SOC content, bulk density, and other soil properties in the United States. The model used a wide range of covariates, including a digital elevation model and derived topographic properties, long-term climate data, hydrological variables, optical remote sensing measurements from the NASA Landsat and MODIS sensors and previously generated SOC maps and soil data.

Castaldi, F., Hueni, A., Chabrillat, S., Ward, K., Buttafuoco, G., Bomans, B., Vreys, K., Brell, M., van Wesemael, B., 2019. Evaluating the capability of the Sentinel 2 data for soil organic carbon prediction in croplands. ISPRS Journal of Photogrammetry and Remote Sensing 147, 267–282. https://doi.org/10.1016/j.isprsjprs.2018.11.026

Uses Partial Least Squares Regression and Random Forest, an ensemble machine learning method that builds decision trees for classification and regression. The authors applied these architectures to multispectral satellite data from the European Space Agency Sentinel-2 sensor, and to airborne spectral measurements from two hyperspectral sensors: the Airborne Prism Experiment (APEX) and a commercial off-the-shelf sensor from the Norwegian company Norsk Elektro Optikk (NEO). These sensors provide coverage throughout the visible, near infrared, and shortwave infrared spectrum. The workflow was used to predict SOC as a percentage by mass in Germany, Belgium and Luxembourg.

Gomes, L.C., Faria, R.M., De Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Filho, E.I.F., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350. https://doi.org/10.1016/j.geoderma.2019.01.007

Uses Random Forest, Cubist (a rule-based regression method), Generalized Linear Model Boosting and Support Vector Machines to predict SOC content in Brazil. Covariates included 74 measurements of surface topography from a digital elevation model, vegetation indices and climate variables.

Dvorakova, K., Shi, P., Limbourg, Q., Van Wesemael, B., 2020. Soil Organic Carbon Mapping from Remote Sensing: The Effect of Crop Residues. Remote Sensing 12, 1913. https://doi.org/10.3390/rs12121913

Uses Partial Least Squares Regression to predict SOC content in Belgium. Covariates were from two remote sensing instruments: APEX and Sentinel-2. APEX is an airborne hyperspectral sensor with coverage throughout visible, near infrared and short-wave infrared regions. Sentinel-2 is a multispectral satellite sensor. The study used two spectral indices to examine the impact of crop residue on SOC prediction: the Cellulose Absorption Index (CAI) and the Normalized Burn Ratio 2 (NBR2). The study demonstrates that using the CAI to remove pixels with residue coverage can improve the performance of SOC prediction.

Dvorakova, K., Heiden, U., van Wesemael, B., 2021. Sentinel-2 Exposed Soil Composite for Soil Organic Carbon Prediction. Remote Sensing 13, 1791. https://doi.org/10.3390/rs13091791

Uses Partial Least Squares Regression to predict SOC content in croplands. The study used Sentinel-2 multispectral satellite data as covariates and created composite images to isolate exposed soil. Composites were assembled using thresholds applied to spectral indices, including the Normalized Difference Vegetation Index (NDVI), and NBR2. This study demonstrates the use of time series filtering based vegetation phenology to select data for SOC prediction. The authors argue that these methods minimize the influence of crop residues, surface roughness and soil moisture.

Heuvelink, G.B.M., Angelini, M.E., Poggio, L., Bai, Z., Batjes, N.H., Van Den Bosch, R., Bossio, D., Estella, S., Lehmann, J., Olmedo, G.F., Sanderman, J., 2021. Machine learning in space and time for modelling soil organic carbon change. European J Soil Science 72, 1607–1623. https://doi.org/10.1111/ejss.12998

Uses a quantile regression forest to predict SOC stock in Argentina. Covariates included topographic measurements and derived properties from a digital elevation model, land cover, long-term climate variables and geological data in addition to measurements from two NASA spaceborne instruments: the Moderate Resolution Imaging Spectroradiometer (MODIS) and Advanced Very High Resolution Radiometer (AVHRR).

Poggio, L., de Sousa, L.M., Batjes, N.H., Heuvelink, G.B.M., Kempen, B., Ribeiro, E., Rossiter, D., 2021. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL 7, 217–240. https://doi.org/10.5194/soil-7-217-2021

Employs a quantile regression forest to predict SOC content, BD and other soil properties using an approach similar to Hengl et al. (2017). Covariates included more than 400 environmental variables, including long-term climate proxies, bioclimatic regions, geological properties, land use and landcover data, topographic measurements from a digital elevation model, vegetation indices and optical measurements from spaceborne remote sensing and hydrological variables.

Sothe, C., Gonsamo, A., Arabian, J., Snider, J., 2022. Large scale mapping of soil organic carbon concentration with 3D machine learning and satellite observations. Geoderma 405, 115402. https://doi.org/10.1016/j.geoderma.2021.115402

Uses a quantile regression forest to predict SOC content in Canada. The analysis used 40 covariates that included long-term climate data, optical remote sensing summaries, soil properties, topographic measurements from a digital elevation model, and data from a spaceborne synthetic aperture radar (SAR).

Zhou, Y., Chartin, C., Van Oost, K., Van Wesemael, B., 2022. High-resolution soil organic carbon mapping at the field scale in Southern Belgium (Wallonia). Geoderma 422, 115929. https://doi.org/10.1016/j.geoderma.2022.115929

Uses gradient boosting to predict SOC content for agricultural fields in Belgium. Covariates were NDVI from optical remote sensing, elevation, clay content, precipitation and organic carbon input from crops.

Ugbemuna Ugbaje, S., Karunaratne, S., Bishop, T., Gregory, L., Searle, R., Coelli, K., Farrell, M., 2024. Space-time mapping of soil organic carbon stock and its local drivers: Potential for use in carbon accounting. Geoderma 441, 116771. https://doi.org/10.1016/j.geoderma.2023.116771

Uses a quantile regression forest to predict SOC stock at multiple points in time in Australia. Covariates included soil properties, topography, weather and climate variables, in addition to satellite-derived quarterly optical indices. This study applied a time-weighted term to increase the importance of recently acquired covariates and demonstrates how time series covariates can be integrated into a DSM workflow.

APPENDIX 4: GUIDANCE ON REQUIREMENTS FOR MODEL VALIDATION REPORTS

This appendix describes required components of the DSM-MVR. Additional details regarding the timing of submission of the DSM-MVR are in APPENDIX 2.

Model Architecture:

- Components, processing of input data, and generation of predictions (see Section 0). If localized models are used to make predictions in different parts of the project area, each localized model and the conditions that govern its use must be enumerated (e.g., decision rules).
- Justification and citation for the model architecture.
- Description and justification for how predictive standard deviations were generated.
- Model source code and/or executables and/or citation, under stable versioning (as a private appendix).
- Whether or not hyperparameter tuning was employed, and if so, a statement describing how tuning was conducted in a manner independent of validation data.

Covariates and Feature Engineering:

- List of model covariates and the data sources for their raw data.
- Description of feature engineering methods used to prepare the covariates.
- Source code for the feature engineering methods, if relevant, as a private appendix.

Soil Sampling for Validation

- Sampling design used for validation.
- Soil sampling data (locations using the appropriate EPSG code and observed values as a private appendix).

Calibration Data External to the Project, if any:

- Number of sample sites and dataset references, if applicable.
- Soil sampling data (locations using the appropriate EPSG code and observed values as a private appendix).
- Description of any harmonization procedures applied (e.g. depth alignment, harmonization across lab analysis procedures, sensor inter-calibration).

Model Validation:

- Attestation of the independence of calibration and validation data and how this independence was maintained.
- Validation design and justification within the chosen soil sampling design.
- Validation summary statistics for the three required tests (goodness of fit, coverage, and bias).

Uncertainty Propagation Methods:

- Evaluation and justification of the project variogram with candidate alternatives (if applicable).
- Description of and justification of the uncertainty propagation procedure.
- Computer code implementing the uncertainty propagation method (as a private appendix.)

APPENDIX 5: MONTE CARLO ERROR PROPAGATION WHEN USING DSM TO INITIALIZE AND/OR TRUE-UP A BGCM

Predicted SOC stock in prediction support unit *i* at time t ($\widehat{SOC}_{i,t}$) may be used to initialize and/or true-up at BGCM in accordance with the target methodology. Depending on the target methodology and version of the target methodology, there are different requirements for the lowest level unit of land that is represented by a single BGCM simulation. In some cases, this unit may be the point level (e.g. for a single point in space represented by a soil sample), and in others this may be an areal level (e.g. a single polygon representing a sample unit or stratum). Guidance is provided below for both cases.

When the BGCM is applied on a point basis, procedures for the integration of DSM with a BGCM under Use Case 1 follow those described in VM0042 applied to soil spectroscopy tools.

1) Initialize the SOC stock (and/or SOC percentage and bulk density, if required) input value to the BGCM by drawing a sample from the predictive distribution in prediction support unit *i* at time t. When the BGCM is initialized using SOC stock, this distribution has a mean equal to $\widehat{SOC}_{i,t}$ and a standard deviation equal to σ_s.
2) Run the BGCM in accordance with the target methodology.
3) When the target methodology permits Monte Carlo propagation of error, repeat steps 1 – 2 for each instance, *l*, of the Monte Carlo simulation. Compute the total estimate of uncertainty in accordance with the target methodology.

When the BGCM is applied on an areal basis, the BGCM is initialized using a predictive distribution of the mean SOC stock at time t ($\overline{\widehat{SOC}_t}$) for the target area (e.g. sample unit or stratum). This distribution has a mean given by Equation 3, and a variance computed using the methods in Section 6.3. Uncertainty of BGCM output is calculated in accordance with the target methodology.

APPENDIX 6: EXAMPLE UNCERTAINTY CALCULATION

This appendix is available as an HTML supplement (link).

APPENDIX 7: ACCOUNTING FOR THE COVARIANCE OF MODEL PREDICTION ERRORS

Project proponents must use a valid method to account for the covariance of model prediction errors when estimating the variance of spatial averages. In most cases, this will involve a spatially explicit method. The steps below describe how an IME can determine whether the method used by a project proponent is permissible under the tool. The method developed by Wadoux and Heuvelink (2023) is acceptable without further justification. However, all methods require:

- Validity of the method's implementation

  The project proponent must provide software or code that demonstrates the implementation of the method. This should be compared with the method description in a valid publication and confirmed to be accurate.

When the method used to account for spatial covariance in variance estimates of spatial averages is not the method developed by Wadoux and Heuvelink (2023), it must be confirmed that:

- The method itself is valid

  The method must be described in at least one peer-reviewed article published in a reputable, subject-specific journal where demonstrating the method is the primary focus of the article or in a reputable academic textbook. The publication or textbook must illustrate the method by applying it to the spatially aggregated total or average of some quantity from a map generated by a statistical model.

- The method must be valid within this application

The method must be appropriate for the project's prediction support unit. For example, a method designed for use over larger areas, like fields, may not be suitable for individual point data. If the publication or book does not explicitly account for the spatial covariance of model errors (e.g., McRoberts et al., 2022), the project proponent must demonstrate that the impact of spatial covariance is minimal. This can be shown using spatial covariance statistics of residual errors (e.g., spatial autocorrelation measures among fields or counties) or other statistical techniques (e.g., as discussed in McRoberts et al., 2022).

DOCUMENT HISTORY

Version	Date	Comment
v1.0	February 20, 2025	Draft version for public consultation