SOIL ORGANIC CARBON (SOC) MODEL REQUIREMENTS AND GUIDELINES

Sustainable Development Goal 13

Publication Date: 08/09/2025

Version: 1.0

Next Planned Update: 08/09/2028

SUMMARY

The Soil Organic Carbon (SOC) Model Guidelines provide a comprehensive methodological framework for quantifying SOC stock changes in agricultural soils through modelling. These guidelines are designed to support SOC sequestration project developers, auditors, and validation/verification bodies (VVBs), and are complementary to the SOC Framework Methodology (SOC FM).

The guidelines do not prescribe specific SOC models but instead focus on a standardized procedure applicable to a range of models. Models are treated as "black boxes"—emphasis is placed on outputs, inputs, and validation, rather than internal mechanics. The document outlines a seven-step workflow for applying SOC models: defining modelling objectives, selecting a model, collecting necessary data, calibrating the model, validating the model, predicting and estimating uncertainty, and verifying model predictions over time.

The most important advantage of the Guideline is a procedure to gradually minimizing measurements using statistically secured model assistance.

Key modelling objectives include:

- Estimating SOC stocks with ground-based measurements
- Estimating SOC stock variance to inform sampling strategies.
- Supporting model-assisted SOC estimation for improved accuracy.
- Enabling model-based estimation for calculating total SOC stock changes.

The project context—including land use, soil types, climate, and management practices—must be clearly defined to guide model selection and ensure relevance. Calibration and validation require robust, context-representative data, ideally derived from within the project boundaries. The guidelines provide detailed requirements for data quality, representativeness, statistical independence, and spatial/temporal consistency.

Model validation is crucial when models are used for direct quantification (model-based estimation) and must meet specific performance metrics: bias, root mean square error (RMSE), and R^2 score. Validation datasets must be independent from calibration datasets, and visual and statistical documentation is required for verification. Models must be revalidated as new field data becomes available.

The guidelines further emphasize uncertainty quantification and the need for transparency and reproducibility. They promote good scientific practices, including clear documentation of modelling decisions, data sources, assumptions, and limitations.

In essence, these SOC Model Guidelines aim to standardize and enhance the credibility of SOC modelling efforts across diverse agricultural contexts, enabling reliable carbon quantification that supports climate mitigation and sustainable land management practices.

Acknowledgements

This document was co-developed by Seqana, Sustainable AG and Gold Standard.

TABLE OF CONTENTS

1. INTRODUCTION

The growing significance of soil organic carbon (SOC) in climate change mitigation strategies requires accurate and scientifically robust SOC modelling approaches (Henry et al., 2022). To support the adoption of best practices in SOC modelling across the industry, Gold Standard for the Global Goals (GS4GG) has developed these guidelines to direct the process of SOC model selection, calibration, validation, and uncertainty handling. These guidelines provide a comprehensive framework for project developers and third-party auditors. They complement the GS4GG SOC Framework Methodology (SOC FM), bridging the gap between theoretical concepts and practical implementation of SOC models.

Given the diversity of SOC sequestration projects and corresponding practices, as well as a myriad of available models for quantifying SOC stock changes that are each fit for different modelling objectives, these guidelines do not prescribe the use of specific models to represent specific ecosystem interactions. Instead, they have a special focus on the common *outcomes* produced when applying and evaluating any SOC model in the context of a SOC sequestration project. The models themselves are treated as "black boxes" that require certain inputs and produce certain outputs; the internal mechanisms of the models are not discussed.

When planning to use a SOC model in a SOC sequestration project, it is key to first understand the specific context and objectives of the project, since local soil properties, climatic conditions, land use practices, and management strategies (e.g., conservation tillage) may critically influence the performance of the SOC model. Detailed knowledge of the project context and objectives ensures the selection of the most suitable SOC model and allows tailoring it to the unique environmental and operational conditions of the project.

After clarifying the project context and modelling objectives, these guidelines delve into the critical aspects of model selection, data acquisition, model calibration, and model validation, underlining the importance of using project-specific data for all these stages. The most important metrics for model validation are provided, along with relevant thresholds to assess the applicability of a SOC model for a specific SOC sequestration project. The guidelines also outline methods for effective and statistically sound uncertainty estimation, which is crucial for assessing the reliability of SOC model predictions. Finally, recurring verification of the model outputs is put forward as a means to monitor the accuracy of the model in the specific SOC sequestration project over time and to repeatedly confirm its applicability in the project context.

1.1 | Structure of SOC Model Guidelines

These guidelines are primarily directed at SOC sequestration project developers to help them in the process of using models in their SOC sequestration projects. Additionally, there are some pointers for Validation and Verification Bodies (VVBs) on how to properly verify the use of the model. Sections 5 to 7 of these guidelines outline the typical sequence of steps required when applying a model in a SOC sequestration project. These sections are generally agnostic to particular methodologies but do include specific explanations of how to apply SOC models in the context of the SOC FM. As a framework methodology, the SOC FM was developed with a modular approach, allowing for extensions with activity modules that describe the exact activities to which it applies. The SOC FM explicitly names “improved agricultural practices” (p. 3) as its core focus. Accordingly, these guidelines focus on SOC models dedicated to quantifying SOC stock (changes) on agricultural lands with improved (in the context of SOC sequestration) agricultural practices.

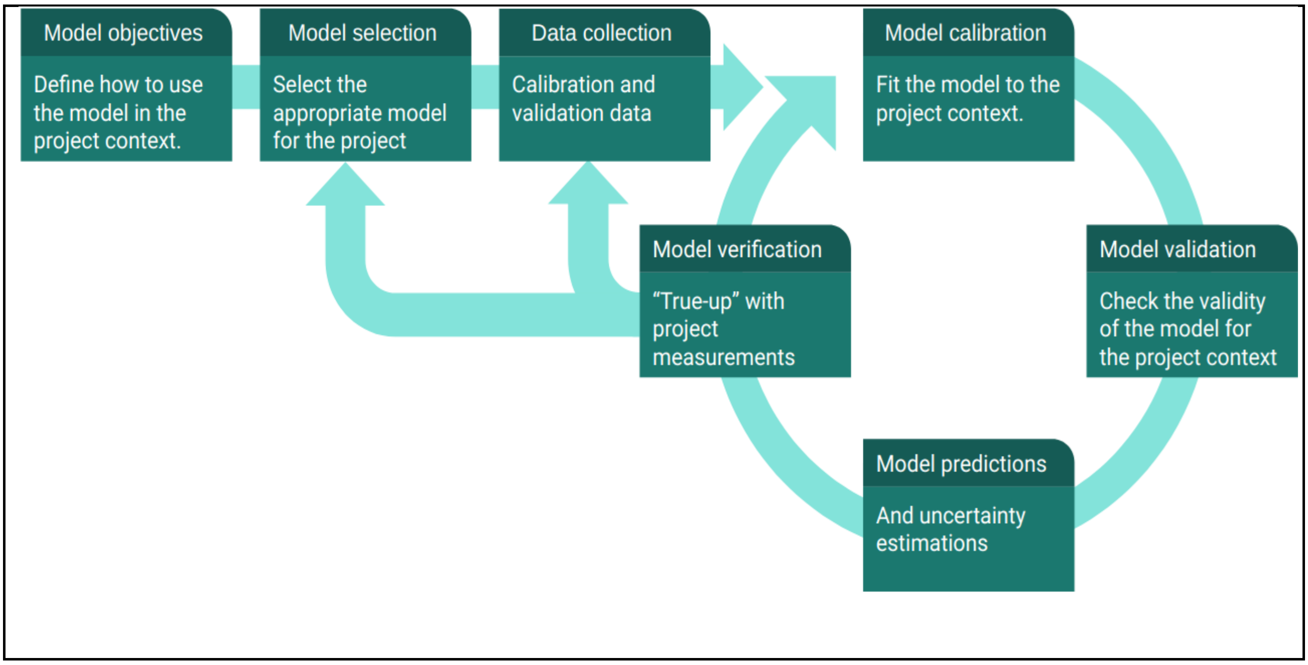


Figure 1. Modelling workflow in the context of a SOC sequestration project

The typical steps necessary to quantify SOC stock (changes) in a SOC sequestration project using any SOC model are ordered and presented in the following sections:

- Define the Modelling Objectives
- Select an Appropriate Model
- Collect the Required Data
- Calibrate the Model
- Validate the Model
- Make Predictions and Estimate the Uncertainty of the Predictions
- Verify the Model Predictions

As shown in Figure 1, after determining the modelling objectives, selecting an appropriate model, and collecting sufficient data for model calibration and validation, the subsequent process of model calibration, validation, prediction, and verification follows an iterative cycle. This cycle has repeated recalibration and validation prior to making predictions and reverifying these predictions with ground-measured observations. Occasionally, a verification event can highlight the need for additional data collection or even the selection of a different model.

Each section of these guidelines has roughly the same structure with the following subsections:

Subsection	Question addressed
• Rationale	Why is this needed?
• Requirements	What is required?
• How-to	How can the requirements be met?
• Required outputs	For the final outputs that need to be submitted to the Gold Standard Secretariat or the designated VVB, what shall they look like?

In terms of required outputs in general:

- The **modelling objectives** shall be communicated to the Gold Standard Secretariat before the start of the respective calculation period (i.e., before SOC stocks are quantified for the start of the calculation period).
- No model validation is required when a model is used exclusively for variance estimation, which allows for efficient design-based SOC stock (change) estimation based on ground-measured samples.
- When a model requires **validation** (i.e., when the model is used for Quantification Approach 2 in the SOC FM [pp.12-14]), the reproducible model calibration and validation report shall be submitted to the Gold Standard Secretariat before the start of the calculation period. Upon request from the responsible VVB, the full calibration and validation dataset shall be made available to the VVB for the purpose of verifying the reproducibility of the results.
- When a model is used for direct quantification (**prediction**) of mean or total SOC stocks (change) (model-based estimation) as part of Quantification Approach 2 in the SOC FM at any point in the project, the mean/total SOC stock (change) predictions from the model shall be made available by the project developer to the Gold Standard Secretariat and VVB before soil sampling for model validation takes place. Official model validation and verification by the VVB can be done only with ground-measured observations that were taken after the model predictions were submitted.

2. DEFINITIONS AND GLOSSARY

2.1. DEFINITIONS

Unless otherwise specified, the definitions from Section 2.1, pp. 6-8 of the SOC FM apply. In cases of doubt or contradictions, the definitions from the SOC FM overrule the definitions listed here.

Term	Definition

Baseline scenario	The activities that would occur in the absence of the proposed project (business as usual, or BAU). (SOC FM, Section 2.1, p. 6). Continuation of the historical land management practices that are being followed in [the] last five years before the project start date (BAU). (SOC FM, Section 6, p. 13)
Baseline SOC stock	Sum of [SOC] stocks in [the project's spatial boundaries]. (SOC FM, Section 6, p. 14) For the first calculation period, [the SOC stock at the beginning of the calculation period within the project's spatial boundaries] is equal to [the baseline SOC stock]. (SOC FM, Section 5, p. 12) SOCBL,y = [SOC stock] before the project start in stratum y [tCha-1]. (SOC FM, p. 15)
Diachronic data	Data measured from the same location unit (point, plot-aggregate, field-aggregate) at multiple points (dates) in time to establish a time series of paired data.
Model calibration/Model training	The process of adapting any (hyper)parameters of a model to a specific project context. For machine learning models, "model training" is a more commonly used term than "model calibration." Since process-based models are already established in the SOC domain and machine learning models present an innovation in the industry, the term "model calibration" is used for both types of models throughout these guidelines.
Model (hyper)parameters	Model internal mechanics that might be calibrated but not changed after calibration from model run to model run in contrast to model input data.
Model input data	All data that is fed into a model to run a simulation or prediction. In statistical modelling, this is often referred to as the independent variable(s) or covariate(s) and in machine learning as feature(s) or predictor(s). In the process-based modelling domain, this refers to any measured or assumed input data ranging from initial SOC measurements to climate and management data.
Model initialisation	Part of model input data but specific to process-based modelling: setting initial conditions for process-based models to run, including the initial SOC measurements.
Model selection	The process of selecting an adequate model (version) for the declared modelling objectives and context.
Model validation	The demonstration of the selected SOC model's suitability to achieve the declared modelling objectives. This shall be demonstrated by satisfying certain statistics thresholds when comparing the model's predictions against ground-measured observations.

	A demonstration that a model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model. (Rykiel, 1996)
Model verification	Evaluating the actual accuracy and uncertainty that the model achieved in the project area, using measured reference data collected in the project area at regular intervals after the project start.
Modelling unit	Refer to the "Definitions and References" section in the SOC FM.
Prediction unit	The smallest spatial unit on which a model produces predictions. Examples: individual points (in geographical space), pixels, parcels, farms.
Process-based model	A model that encodes the relationship between inputs and outputs purely or mostly based on scientific assumptions on the real-world processes that govern this relationship (antonym: statistical model). Examples: RothC [Rothamsted Carbon], DayCent [Daily Century]. These often include several levels of process detail, e.g., plant organ > plant > canopy or cell > plant organ > plant.
Project scenario	Refer to the "Definitions and References" section in the SOC FM.
Sampling unit	The spatial level of granularity at which SOC reference values are sampled. Examples: individual points (in geographical space), pixels, fields, farms. For model validation and use, this shall be hierarchically consistent with the prediction unit, i.e., the physical sampling unit shall be small enough so that it can be spatially contained by the prediction unit. For example, the prediction units point, pixel, parcel, or farm can contain the sampling unit point, but a sampling unit field (e.g., composite field sample) cannot be used for model validation with a prediction unit point or pixel. In this case, the prediction unit shall become the field as well and model predictions shall be aggregated to that prediction unit, as described in Section 5.4.2 of these guidelines.
Snapshot SOC stock	The SOC stock at a single particular point in time.
Statistical model	A model that fits ("learns") a relationship between input data and the target variable purely or mostly based on mathematical relationships of measured data, with little or no hard-coded scientific assumptions on the real-world processes that govern this relationship (antonym: process-based model). Examples: generalised linear models, random forest models, neural networks.

Target variable	The entity under investigation that shall be predicted/simulated by the model. In SOC sequestration projects, this is typically the SOC stock at one point in time or the SOC stock change over a period of time. In statistical modelling, the term's response or dependent variable is also commonly used.

3. SCOPE, APPLICABILITY, AND ENTRY INTO FORCE

3.1. Scope

These guidelines provide methods to calculate project and/or leakage emissions resulting from the combustion of fossil fuels.

3.2. Applicability

a) These guidelines can be applied where carbon dioxide (CO2), methane (CH4), and nitrous oxide (N2O) emissions from fossil fuel combustion are calculated based on the quantity and properties such as chemical composition or net calorific value (NCV) of the fuel combusted.

b) Methodologies and other methodological tools applying these guidelines shall specify the particular combustion process to which they are being applied.

3.3. Entry into Force

The date of entry into force of these guidelines is from the date of publication.

4. NORMATIVE REFERENCES

The following documents are to be referred:

- Land-use & Forests Activity Requirements
- Soil Organic Carbon Framework Methodology

5. METHODOLOGY PROCEDURE

5.1. Project Boundary

Refer to SOC FM, Section 4, pp. 9-11 for the definition of the project boundary.

5.2. Define the Modelling Objectives

First, the goals for SOC modelling shall be identified, and the specific (spatial and temporal) context within which the SOC modelling goals shall be attained. These steps are critical for selecting an appropriate SOC model and collecting adequate validation data.

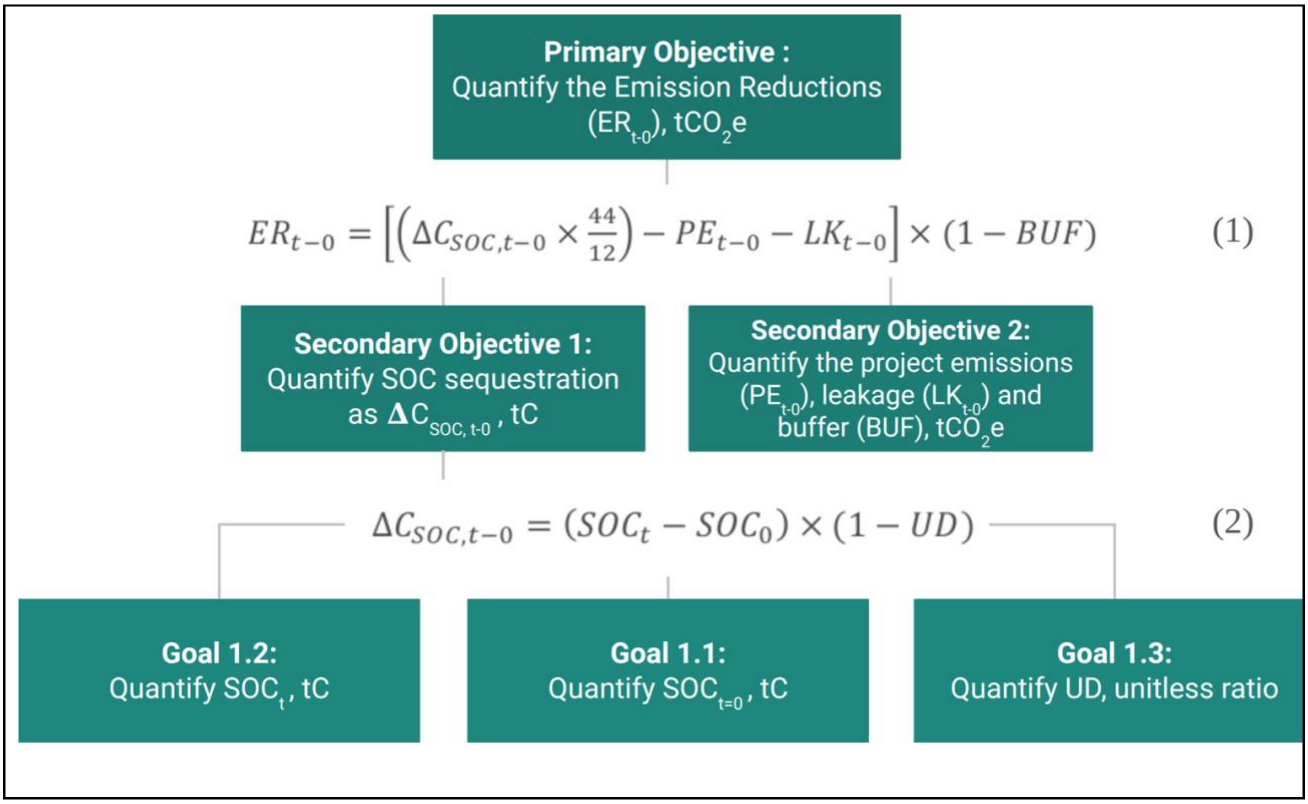


Figure 2. Quantification objectives and goals from the SOC FM with the corresponding Equations 1 and 2 (SOC FM, p. 11)

5.2.1 The Potential SOC Modelling Goals

Box 1. Synopsis of modelling goals for projects with the SOC FM
The ultimate goal of modelling within a SOC project is to quantify the SOC stock changes $\Delta C_{SOC,\,t\text{-}0}$ between two points in time, with t = end of calculation period, 0 = start of calculation period (the project). This change can be modelled directly or derived as the difference between the *SOC stock snapshot at t=t - SOC stock snapshot at t=0.* (See Box 3 in these guidelines.) SOC models, especially process-based models, have traditionally been developed, calibrated, and validated at small plot scales in research settings. Whether they can be applied to larger-scale farming systems needs to be validated and verified as these projects take place. This means that soil sampling and measurement are still required, at least in the first calculation periods of the project, to ensure that the selected SOC model is adequately calibrated and validated to quantify the SOC stock changes (via direct modelling of change or modelling of snapshots in time). This means that over the project life cycle, three different strategies can be employed for quantifying SOC stock (changes), with different roles for modelling: 1. **Design-based estimation:** Quantification of SOC stock via soil sampling and measurements. Modelling can be used prior to the sampling campaign to quantify the SOC stock variance over the project's spatial boundaries to delineate strata (modelling units) and determine the required sample size.

2. **Model-assisted estimation** quantification of SOC stock via modelling SOC stock, in addition to ground-measured samples, to increase the precision of a design-based estimate.
3. **Model-based estimation** quantification of mean and total SOC stocks over the project's spatial boundaries via a SOC model.

With any SOC sequestration project, the primary objective is to quantify the removed emissions in tons of CO_2e. This quantification is formalised in Equation 1 of the SOC FM (p. 11) and can be broken down into two secondary objectives:

1. Quantify SOC sequestration as $\Delta C_{SOC,\ t\text{-}0}$.

2. Quantify the project emissions ($PE_{t\text{-}0}$), the leakage ($LK_{t\text{-}0}$), and the buffer (BUF) emissions.

These guidelines are exclusively concerned with the first of the two: the quantification of the SOC sequestration as the total SOC stock change within the project's spatial and temporal boundaries and the corresponding uncertainty of this estimate. This is formalised in Equation 2 in the SOC FM (pp. 11-12, Figure 2), where $\Delta C_{SOC,\ t\text{-}0}$ stands for the change within the project's spatial and temporal boundaries, SOC_0 and SOC_t stand for the SOC stock at the beginning and end of the calculation period, and UD stands for the corresponding uncertainty deduction.

Box 2. The baseline in the SOC FM

In SOC sequestration projects in general, the change in the project needs to be compared to a baseline. In the SOC FM, the baseline scenario is defined as the "continuation of the historical land management practices ... in the last five years before the project start date (BAU)" (p. 13), and the baseline SOC stocks are calculated as "the sum of stocks in [the project's spatial boundaries in the year of the project start]" (p. 14). As a default, the SOC FM prescribes a static baseline with the assumption that SOC stock is at equilibrium (i.e., would not change) under the BAU baseline scenario. Thus, the baseline SOC stock is set equal to the initial SOC stock within the project's spatial boundaries in the project start year (p. 12): "SOC_0 is equal to SOC_{BL}."

Since the definition of the baseline might change in newer versions of the SOC FM, the latest version shall be the definitive source to clarify the definition of the baseline scenario and the quantification of the baseline SOC stock.

This means the following quantities can generally be modelled within the context of the SOC FM:

- Mean, total, and variance of SOC stock change over the project's spatial boundaries (or individual modelling units) and over calculation/crediting period(s) ($\Delta C_{SOC,\ t\text{-}0}$)

- Mean, total, and variance of SOC stock over the project's spatial boundaries (or individual modelling units) for a snapshot (project start [i.e., SOC_0/SOC_{BL}], middle, or end [SOC_t])

These guidelines focus on the latter (modelling snapshots) with the goal of applying Equation 2 in the SOC FM to derive $\Delta C_{SOC,\ t\text{-}0}$. This is in line with the SOC FM, and there are also validation and calibration data limitations to modelling the change directly, as explained here in Box 3.

Box 3. Modelling change directly versus subtracting modelled snapshots
While the SOC stock change can be modelled directly with temporally explicit dynamic process-based models, it is quite challenging to validate (and calibrate) such model predictions because of the lack of appropriate validation studies and reference data (Garsia et al., 2023; Le Noë et al., 2023). (Also see Section 5.4 of these guidelines, "Collect the Required Data.") Overall, there is a shortage of SOC stock time series data at the spatial scale (extent and density) of typical SOC sequestration projects (Bradford et al., 2023; Lavallee et al., 2024). The projects themselves can generate this data when repeatedly measuring SOC data in situ for calculation of the sequestered SOC (see Table 3 in these guidelines for the calculation instructions), as described in Approach 1 for baseline and project scenario calculations in the SOC FM (Sections 6-7, pp. 14-21). However, creating a time series of paired soil data points requires careful planning in both sampling and measurement. The full requirements are listed in Box 5 in these guidelines. Because of these limitations, and in correspondence with the SOC FM, these guidelines focus on modelling the mean, total, and/or variance of the SOC_t stock at temporal snapshots and determining the corresponding uncertainty for calculating the SOC stock change over time as the difference between two snapshot predictions.

To quantify the SOC stock over the spatial boundaries of the project, the SOC FM allows for two distinct quantification approaches:[1]

> "Approach 1: ... On-site measurements to directly document baseline and project SOC stocks.
>
> Approach 2: ... Models from peer-reviewed publications to estimate baseline and project SOC stocks. Project developers need to prove that [the models] are conservative and applicable to the project site and management practice. ... Models derived locally may be applied only if validated by direct measurements in the project area (Approach 1). Generally, project developers shall select the most specific approach possible with the data available, giving preference to local data sources and models." (SOC FM, p. 12)

1 Approach 3: Default factors to estimate SOC changes, relating to the general Tier 1/2 model described in the Intergovernmental Panel on Climate Change (IPCC) Guidelines for National Greenhouse Gas Inventories (IPCC 2019) is not relevant to the types of models covered in these guidelines.

In estimation statistics, these two approaches are coined **design-based** (Approach 1) versus **model-based** (Approach 2) estimation respectively (Särndal et al., 2003). From here on, we will use these terms instead of Approach 1 or 2 respectively, as they are more generally applicable beyond the scope of the SOC FM.

While purely model-based estimation is generally permitted under the SOC FM, it is clearly stated that the model shall be previously validated against direct measurements in the project area. When direct measurements are not available before the start of the project, such measurements can be taken and then used for dual use: for validating the model for the project and for design-based estimation. Also prior to or during design-based estimation, a model can be used as follows:

- To determine where and how many soil samples shall be taken in the prescribed stratified random sampling design as noted in Section 16.2 of the SOC FM (pp. 35-36). To determine how many samples (soil pits) are required, an estimate of the variance of SOC is needed as described in Section 5.4.2 of these guidelines, "Requirements for Data Collection." This estimate of the variance can be obtained through modelling.
- In addition to the sampled ground-measured observations in a model-assisted estimation process to increase the precision of the estimate. (Brus, 2000)

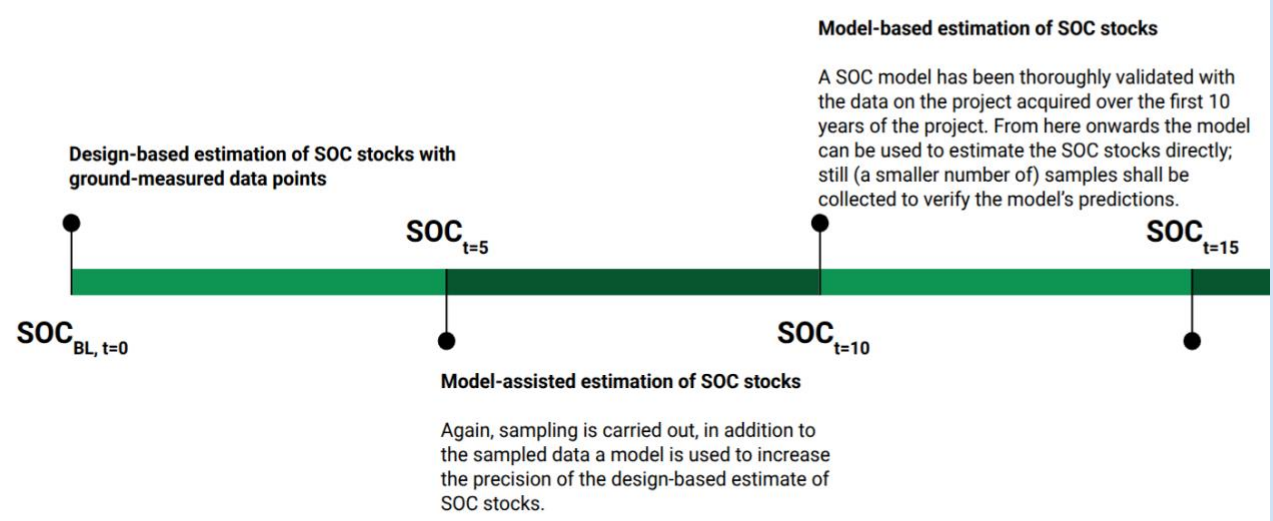
Thus, the modelling goals can be threefold in the SOC FM (also see Box 1 in these guidelines):

1. **Variance estimation for informing design-based estimation of SOC stocks:** Modelling the variance of SOC stock within the project's spatial boundaries to determine the number of required soil samples to take and draw the stratum boundaries (modelling units) for taking soil samples.
2. **Model-assisted estimation of SOC stocks:** Modelling a regression function in addition to the ground-measured samples for design-based estimation to increase the overall precision of the estimate.
3. **Model-based estimation of SOC stocks:** Modelling the mean or total SOC stock as a snapshot at the start of the project or prior to any following verification event to quantify the creditable SOC sequestration for the respective calculation period.

Throughout the life cycle of a SOC FM sequestration project, one, two, or all three of these modelling goals might be applied as shown in Box 4 below. This box presents a fictitious example and is not intended as a strict prescription. In particular, model-assisted design-unbiased estimates can already be applied at t=0 without explicit model validation. Model-based estimates can also be used at t=0 if the models can be validated in accordance with Section 6.1 of these guidelines, "Validate the Model," for the respective target variable on Data Sources B, C, or D (see Section 5.4.3 of these guidelines, "How to Acquire Data") with the respective safety discounts in Table 4 of these guidelines.

Box 4. Fictitious example of modelling SOC within a SOC FM project

When there is not sufficient validation data at the start of a project, the first quantification rounds (i.e., for the baseline quantification at t=0 and for the first calculation period, e.g., at t=5 years) can be conducted via design-based estimation, where in parallel a model is calibrated and predictions are generated solely for the purpose of validating these predictions against the ground-measured observations as described in Section 6.1 of these guidelines, "Validate the Model." After successful validation of the model for the baseline and the first calculation period, the model might be used for quantification in later calculation/crediting periods if permitted by the Gold Standard Secretariat and/or the responsible VVB in accordance with Section 8 of the SOC FM (pp. 23-24).



Only the purely model-based estimation of the mean or total SOC stocks over the project's spatial boundaries requires strict model validation, as the validity and objectivity of the approach are not guaranteed without ground-measured observations to compare with the model predictions.

5.2.2 Defining the Modelling Context

In addition to determining whether the goal of the model is to assist with design-based estimation or conduct model-based estimation, the modelling context shall be specified with respect to the project's spatial and temporal extent and the intended land management changes.

SOC models are often developed and validated with data from specific ecosystems and land management practices. A model developed and validated with data from a specific spatial and temporal extent, climate zone and soil texture class, and/or for a specific type of agricultural land management might be valid and applicable in the same or similar context—but not necessarily in a different climate zone or soil texture class and/or for a different practice.

The SOC FM prescribes that the project area shall be stratified into multiple modelling units with similar soil types, climate zones, and land management practices (further differentiated in the SOC FM, Section 7, p. 20). As a consequence, the modelling objectives in different modelling units of the same project may be different. Projects with multiple modelling objectives may require selection, calibration, and validation of separate models for the individual modelling units since a single model may not be suitable for all modelling objectives.

5.2.3 Requirements for Defining the Modelling Objectives and Context

To effectively define the modelling objective(s), the following requirements shall be met:

1. Project definition: Define the spatial and temporal boundaries of the project, activities, and land management changes in line with Chapter 4 of the SOC FM.
2. Data collection: Collect and record project data and parameters as specified in Section 16.3 of the SOC FM, which includes:
   a) Project activity/land management change
   b) Project start and end dates
   c) Details of each farm/land parcel within the project boundaries, including unique identifiers, owner information, physical address, global positioning system (GPS) coordinates/geographic shape, land area and use, agrochemical and fertiliser usage, and fossil fuel and electricity consumption for baseline activities
3. Contextual data: Gather data to establish representativeness of validation data to the project context, including climate zones, soil texture classes, vegetation types, and management history (for every modelling unit).

5.2.4 How to Define the Modelling Objectives

Defining the project and modelling objectives involves several key steps:

a) Defining project boundaries and activities: Clearly delineate the spatial and temporal boundaries of the project and define the specific activities undertaken.

b) Data collection and analysis: Collect detailed data on the farms and land parcels. The exact requirements are listed in Section 7.1 of these guidelines. Most of the relevant data shall be provided by the project developer. If the data is not readily available from the project developer, then do the following:
   - Search for the data in public data records or use published maps.
   - Implement soil sampling and measurement activities to obtain the required data.

a) Stratify the project area into modelling units in line with Chapter 6 of the SOC FM.

b) Formulate the modelling objective(s) for the project as a whole or the individual modelling units as precisely as possible, including at least the following contextual information:

- Target variable to be quantified (e.g., change in SOC stock)
- Soil depth of interest
- Project activity
- Climate zone
- Soil type

Formulating the modelling objective in this detailed way will help in selecting the most suitable model and finding representative data for model calibration and validation. If any of the contextual information listed above differs between modelling units, the modelling objectives shall be stated individually for each modelling unit. In this case, different SOC models, different calibration data, and/or different validation data might be required for the different modelling units.

5.2.5 Outputs of the Modelling Objective and Context Definition

The project context shall be defined in line with Chapter 4 of the SOC FM. The project's spatial and temporal boundaries shall be clearly delineated, and the project activity/activities shall be clearly defined. In addition to the requirements for data to be collected as listed in Sections 16.3 and 16.4 of the SOC FM, the data in this Table 1 shall be collected and submitted.

Table 1. Expected output: geospatial file(s) (geojson, geopackage, or any other open standard geospatial file format) with the following data and metadata on every farm/parcel within the project's spatial boundaries (one row per farm or parcel as appropriate)

Attribute	Data Type
Unique identifier	Numeric/Hash
Descriptive name (name of the farm/parcel, address, landowner/project developer)	Text
Exact spatial boundaries of the total and eligible area (expected spatial accuracy within single-digit metres; if otherwise, report the spatial accuracy)	Text (i.e, well-known text representation of geometry)
Start date and end date of the crediting period	Text (yyyy-mm-dd)
Modelling objective	Text
Modelling unit to which the farm/parcel belongs	Text

Climate zone as defined in the 2019 Refinements to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories (Reddy et al., 2019)	Text
Soil texture class and specifically the soil's clay content in %	Text/Numeric (clay %)
Vegetation type (e.g., grasses, legumes, non-legume broadleaf species)	Text
Current land use	Text
Management history (land use changes), historic vegetation types, tillage techniques, and fertilisation	Text

In addition, the modelling goal shall be clearly communicated when the project context definition is submitted to the Gold Standard Secretariat as one of the goals noted in Section 5.2.1 of these guidelines, "The Potential SOC Modelling Goals."

5.3 Select an Appropriate Model

5.3.1 Model Selection Rationale

This section provides general guidance on how to select appropriate SOC models to achieve the modelling objectives of a SOC sequestration project. It does not provide definite recommendations to use a specific SOC model in a specific project context. However, it is important that the chosen SOC model can be independently validated. To that end, the model's predictions and simulations shall be fully reproducible and accessible, meaning a proper versioning of the model and associated data (including input data and random seeds for stochastic models) shall be available.

5.3.2 Understanding SOC Models

Before selecting a model, it's essential to understand the different types of SOC models and what they do. Process-based SOC models simulate the accumulation and/or loss of organic carbon in the soil. Statistical models use regression or machine learning algorithms to derive a quantification formula to determine the SOC stock (change) from input data and associated weights. Both types of models vary in complexity, scale, and the factors they incorporate, such as climate variables, land use, and management practices.

The selection of an appropriate SOC model is critical for the success of a SOC sequestration project. The right model balances accuracy, complexity, data requirements, and resource availability. Carefully considering these factors and thoroughly evaluating available models ensure that a SOC sequestration project is based on reliable and relevant modelling outcomes.

5.3.3 Model Selection Requirements

Key Considerations for Model Selection:

a) Project context: The SOC model shall be appropriate for the project context (i.e., project activity, climate zone, soil type). Different models are designed for various scales, such as field level, regional, or continental assessments or changes on a daily, monthly, yearly, or decadal basis. Whether a model is appropriate can be judged only through model validation. When a model meets the requirements noted in Section 6.2.1 of these guidelines, it is deemed appropriate to the project context.

b) Data availability and quality: In addition to the aforementioned validation data (Section 5.4 of these guidelines), the availability of input data (e.g., predictors, features, parameter sets, independent variables) shall be assessed. Some models need extensive input data, while others are less data-intensive. The choice of model may depend on the data that can be realistically obtained.

c) Model complexity and usability: Models range from being simple and statistical (e.g., linear regression, one target/predictor variable, predicting single point in time) to being complex and process-based (e.g., DayCent, multiple internal variables, predicting at daily time steps). While complex models may provide more detailed insights, they require more data and computational resources. The team's expertise and resources in model operation shall be considered.

d) Local calibration/training and validation: Check if the model can be calibrated/trained and validated with local data. A model’s performance greatly depends on its applicability to the local context of the project. Reserve some local data for validation that shall not be used for calibration/training purposes.

e) Temporal and spatial resolution: Consider the temporal and spatial resolution needed for the project. Some models provide fine-scale, high-resolution outputs, while others work at coarser resolutions.

f) Integration with other tools: Determine if the model needs to be integrated with other tools or models (e.g., geographic information system software, climate models). Some SOC models are standalone, while others are designed as part of larger environmental modelling suites.

g) Model support and community: Consider the support available for the model, including documentation, user communities, and technical support. A well-supported model can significantly ease the implementation process.

5.3.4 How to Select a Model

Steps in Selecting a SOC Model

a) Define the project context as described in Section 5.2 of these guidelines, “Define the Modelling Objectives.”

b) Research available models: Conduct thorough research on existing SOC models, including their capabilities, limitations, and data requirements. (See the resources below.)

c) Check prior validation: If one of the available models is already validated on the project context (e.g., same type of practice change, prior land use, soil texture class, and climate zone), it is possible to reuse the model.

d) Evaluate data availability: Match the available data with the data needs of the models, taking into account both input and output data for calibration and validation.

e) Shortlist suitable models: Based on the prior steps, create a shortlist of potential models.

f) Compare the models: If possible, test the shortlisted models with a subset of the data to compare their performance as described in Section 5.3.5 of these guidelines, "Expected Model Selection Outputs."

g) Final selection: Choose the model(s) that best fit the project needs, considering performance, data availability, and resource constraints.

Resources

Garsia et al (2023) have compiled several resources on process-based SOC model selection and validation. Additional resources are available at the Model Portal of the International Soil Modelling consortium (ISMC, 2023).[2] Remote sensing–based SOC models also have become popular in recent years (Dvorakova et al., 2023; Hengl et al., 2021; Heuvelink et al., 2021; Poggio et al., 2021; Szatmári et al., 2021; Zepp et al., 2021) and might be suitable. However, take care during the validation process to test these models on the right validation data.

5.3.5 Expected Model Selection Outputs

The output of the model selection process is one model per modelling unit of the project. The same model can be selected for multiple modelling units if the model can be validated for all of them. Any high-level model class (e.g., DayCent, RothC, Random Forest) that was considered and dismissed during the model selection process shall be recorded in a list, along with a short explanation of why it was dismissed.

5.4 Collect the Required Data

5.4.1 Rationale of Data Collection

Any SOC model intended for use in a SOC sequestration project shall demonstrate reliable predictions for SOC stocks and corresponding uncertainty estimates. The only

2 https://www.soil-modeling.org/resources-links/model-portal

way to robustly carry out this analysis is by validating the model outputs against measured reference data. Data used for model validation is called **validation data**. Measured reference data can also be used during model calibration to fit a model to a specific project context or to select the best model from a set of candidate models. Data used for these purposes is called **calibration data**. This section provides the key requirements to consider when gathering validation and calibration data for SOC stock modelling.

The SOC stock change data from the project site is required to ultimately validate the whole quantification procedure. However, such data rarely exists within the project spatial boundaries before or at the project start as it would require at least two SOC stock measurements over time. Ideally, the SOC stock change data should be obtained from diachronic (i.e., paired-at-same-location time series) data (Le Noë et al., 2023), i.e., it should be derived from time series of SOC stock data measured recurrently from the same location(s) at time t0 before a land management change was initiated and at times t1 (t2, t3, …, tn) some time (1-20 years) after the specific change was implemented. Since this data hardly exists in practice, the capacity to validate change predictions of the model is limited at the start of the project. Instead, what can be measured and validated is the spatial variation of SOC stocks at one point in time and the capacity of the model to predict the parameters (e.g., mean, variance) of this spatial distribution of SOC stocks at a single point in time. From the precision of this spatial estimation of a single point in time, the uncertainty for the calculation of the change as the difference between two single points in time can be calculated. (See Annex A.1 Simulated Example)

Over the course of the project, sufficient validation data shall be sampled and measured over time within the project's spatial boundaries to allow for independent diachronic validation of the model, as Le Noë et al. (2023) suggest (see Figure 3 in these guidelines). This collected validation data can then be used to validate a model and opt for purely model-based estimation of the SOC stock in later calculation periods (as explained in Box 6 in these guidelines).

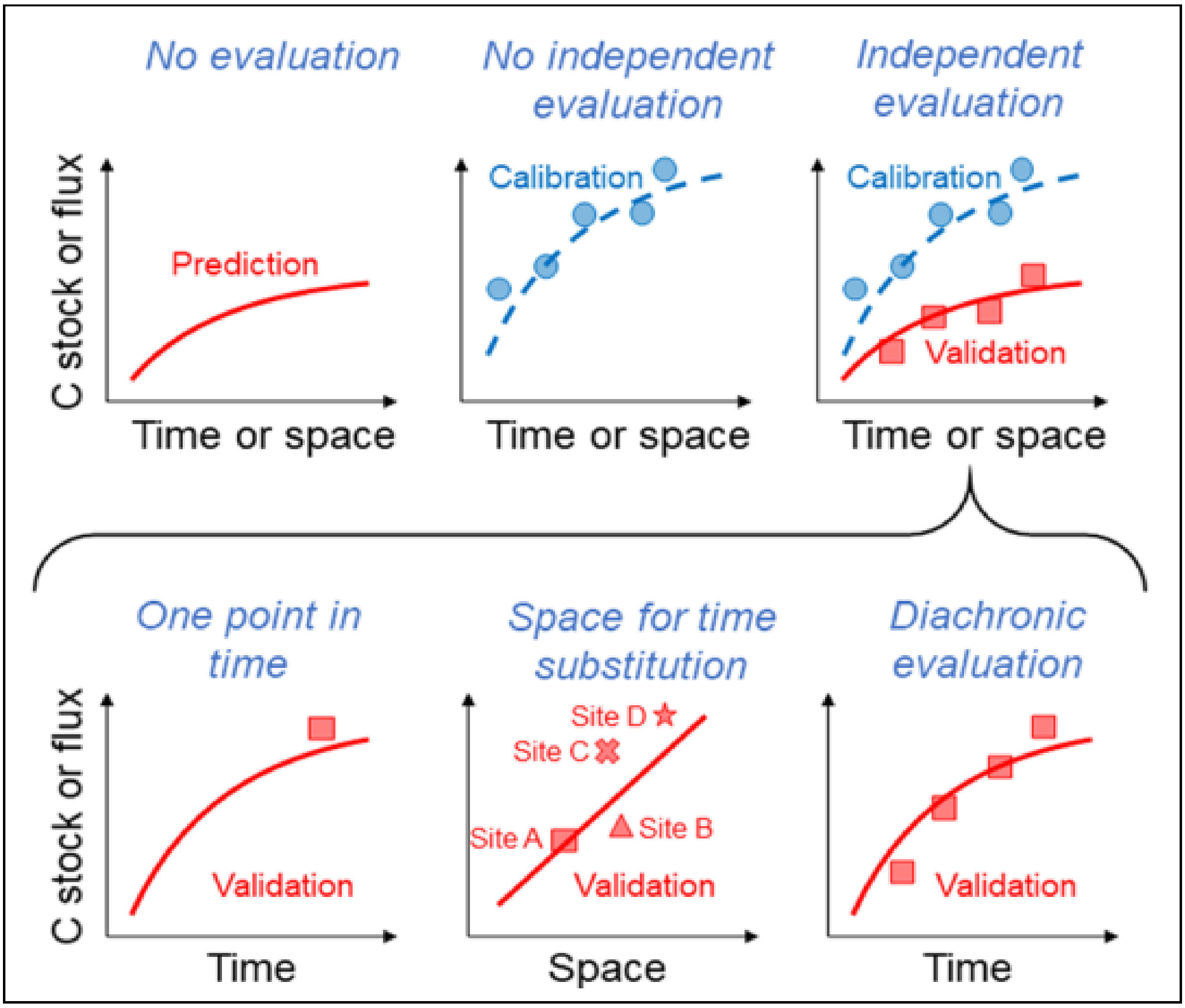


Figure 3. Figure 1c from Le Noe et al. (Le Noë et al., 2023): "A schematic representation of (…) links (…) between SOC models and [the] empirical field"

5.4.2 Requirements for Data Collection

The **validation data** shall represent the different levels of SOC stocks:[3]

a) The data shall be presented in units of tonnes of carbon per hectare [tC/ha].

b) The SOC stock data shall be calculated as the product of the SOC concentration of the fine soil (sieved soil particles at <2mm) in [%] with the fine soil stock (FSS), i.e., the fine soil mass per unit area in [t/ha].

   1. If FSS is measured directly, Equations 3a and 3b in Table 2 in these guidelines shall be applied (Poeplau et al., 2017).

   2. If FSS is not measured directly, it shall be computed as the product of the (gravimetric) bulk density (g cm^{-3}) of the fine soil and the soil depth (cm), deducted by the coarse fragment ratio (sieved soil particles at <2mm), as shown in Equation 4 in Table 2 in these guidelines.

3 The data shall be representative of the project in terms of spatial extent (total area and radius), temporal extent (length of project and length of calculation periods), and the management practices taking place in the project scenario.

3. When comparing two different measurements of a time series, the concept of equivalent soil masses (Wendt & Hauser, 2013) shall always be considered. See Equations 5a and 5b in Table 2 in these guidelines.

c) To ensure comparability across datasets, the following data harmonization steps shall be taken prior to analysis:

1. All SOC and bulk density (BD) values shall be converted to common depth intervals aligned with the modelling objective (e.g., 0–30 cm), using pedotransfer functions or depth-weighted averaging when necessary.

2. Units shall be harmonized across all datasets (e.g., converting g/cm³ to t/ha or %SOC to g/kg where applicable).

3. Differences in analytical methods (e.g., Walkley-Black versus dry combustion) shall be documented. When possible, correction factors shall be applied based on peer-reviewed cross-lab comparison studies.

4. Metadata (e.g., sampling date, method, equipment, and lab technique) shall accompany all datasets to support transparent harmonization.

The sampling units of the **validation data** shall be consistent with the prediction units of the model that is validated. For example, a model that produces farm-level predictions shall be validated using data that was sampled on a farm level.

If there is a mismatch between sampling units and prediction units, the unit with the finer spatial granularity shall be aggregated (e.g., by calculating the mean SOC stock of point locations to represent the mean SOC stock at a field level) to match the unit with the coarser granularity. For example, if the SOC model produces farm-level predictions but SOC reference data is available at individual point locations sampled via probability sampling within farms, then all points within the boundaries of a farm shall be aggregated to provide a single farm-level mean reference value.

The **validation data** from the project site shall capture the variance of the SOC stock within the spatial boundaries of the project with sufficient precision to quantify the changes within a **90% confidence interval** (in correspondence to the uncertainty quantification of the SOC FM, Section 9, pp. 26-28)**.** In cases where the sampled SOC reference data is aggregated (i.e., a field- or farm-level mean from point-level measurements) to be consistent with the prediction unit of the model, the thresholds apply after aggregation. The equations to calculate the required number of samples are shown in Box 5.

The **validation data** shall be representative for the project context (modelling units):

a) The validation data shall be sampled over the soil depth interval stated in the modelling objective (e.g., 0-30 cm or 0-50 cm).

b) The validation data from within the project boundaries shall be collected in accordance with Section 16.2 of the SOC FM, pp. 35-36.

c) Directed stratified sampling (with stratification based on additional data and composite samples aggregated within each stratum) shall be used in accordance with the protocols noted in the SOC FM (p. 44) or Annex 3.2 of the Food and Agricultural Organization (FAO) Global Soil Organic Carbon -

Monitoring, Reporting, and Verification (GSOC MRV) Protocol (FAO, 2020, pp. 89-91, referred to as "directed stratified sampling" in this protocol.) for any validation data.

d) When previously collected data is used for model validation that was collected without following the directed stratified sampling design referenced above, justification shall be provided for how the data is unbiasedly representative of the project context (detailed further in Section 5.4.3 of these guidelines, "How to Acquire Data"). In the best case, the data was still sampled via a probability sampling design to satisfy the criterion of unbiasedness.

For the **calibration data,** there are no hard requirements, although the data shall be similarly representative of the project context as the validation data to make model calibration most effective. However, data that does not satisfy some of the requirements for validation data may still be used as calibration data.[4] This also means that calibration data does not need to come from within the project site itself.

The **validation data** shall be statistically independent of the **calibration data**.

a) To ensure statistical independence, the following criteria shall be fulfilled:

    1. The validation data and the calibration data shall be mutually exclusive, i.e., no single data point that is part of the validation data may be part of the calibration data and vice versa.

    2. The minimum geographical distance between any validation data point and any calibration data point shall be 30 metres. This entails that all data points taken at the same location at different points in time and different soil depths shall either be completely assigned to the validation data or completely assigned to the calibration data.

b) Ideally, calibration data and validation data shall be drawn randomly without replacement from the same superpopulation of data (in the best case, from the same sampling campaign(s)). Care shall be taken to enforce the minimum distance constraint mentioned above.

c) To prove that the above requirements are fulfilled:

    1. The model validation report shall clearly state the minimum geographical distance (in metres) between any validation data point and any calibration data point as well as the coordinate reference system in which the distance was computed.

    2. The geographical locations, time stamps, and soil depths of all validation data points and all calibration data points shall be made available as a geospatial file (geopackage, geojson, or other open standard geospatial file format) along with the model validation report allowing for simple verification of the minimum distance constraint. The actual measurements (e.g., SOC stock values) or model predictions for these data points may be

4 These guidelines and the SOC FM evaluate the suitability of modelling SOC stock change quantification based on the model validation and verification. Before a model is allowed to be used for direct quantification with the SOC FM, it shall be validated as noted in Section 6.1 of these guidelines.

omitted for reasons of confidentiality but shall be made available to the VVB upon request for the purpose of official validation and verification.

When the goal is to collect or generate a time series of SOC stock data, special care shall be taken to guarantee the adequacy of the time series data, as explained in Box 6 in these guidelines.

Table 2. Calculation of SOC stocks

Method	Equation	Equation Number
Poeplau FSS Method (Poeplau et al., 2017)	$FSS = \frac{mass_{finesoil}}{volume_{sample}} \times depth \times 0.01$ where *mass*$_{finesoil}$ is given in [g], *volume*$_{sample}$ is given in [cm^{-3}] and *depth* is given in [cm]; the resulting unit is [t/ha]. $SOCstock = SOCcon_{finesoil} \times FSS$ where *SOC*$_{confinesoil}$ is given in [%]; the resulting unit is [t/ha].	3a 3b
BD Method	$SOC_{stock} = SOCcon_{fine\ soil} \times BD_{fine\ soil} \times depth \times (1 - CFR)$ where *SOCcon*$_{fine\ soil}$ is given in [%], *BD*$_{fine\ soil}$ is given in [g cm^{-3}], *depth* is given in [cm] and *CFR* is given in [%]; the resulting unit is [t/ha].	4
Equivalent soil masses (Wendt & Hauser, 2013)	$M_{SOIL(DL)} = \frac{mass}{area} = \frac{M_{SAMPLE(DL)}}{\pi\left(\frac{D}{2}\right)^2 \times N} \times 10000.$ where • $M_{SOIL(DL)}$ is the soil mass represented by a soil sample depth layer in Mg/ha. • $M_{SAMPLE(DL)}$ is the dry sample mass represented by a soil sample depth layer by the area sampled by the probe or auger in g. • π(D/2)2 is the cross-sectional area of the probe's or auger's inside diameter in mm. • *N* is the number of cores sampled. Furthermore: $M_{OC(DL)} = M_{SOIL(DL)} \times C_{OC(DL)},$	5a

Table 2. Calculation of SOC stocks		
	where $M_{OC(DL)}$ is the mass of organic carbon represented by a soil sample depth layer in kg/ha. $C_{OC(DL)}$ is the concentration of organic carbon represented by a soil sample depth layer in g/kg.	5b

Box 5. Confidence interval-based estimation of sample size versus minimum detectable difference

While many SOC sampling protocols and methodologies recommend the use of a minimum detectable difference (MDD) formula for determining the required sample size, we do not recommend this. The MDD is mainly intended for use in a statistical t-test, which is not required as part of the SOC FM. Instead, a 90% confidence interval is required (SOC FM Section 9, pp. 26-28), which shall be sufficiently narrow to allow for robust quantification of SOC stock changes and to avoid excess uncertainty deductions. To determine the required sample size (for design-based estimation) for this 90% confidence interval, two things are needed:

n

1. The desired degree of precision (or margin of error) in terms of SOC stock change
2. An estimate of either the variance of SOC stock change (paired sampling scenario) or an estimate of the spatial variance of the SOC stock at the two temporal snapshots

Correspondingly, there are two potential formulas for determining the required sample size:

1. For paired sampling, with a change (temporal) variance estimate:

$n = \dfrac{z_{\alpha/2}^2 s^2}{MoE^2}$	Equation 6

2. For independent sampling, with a spatial (snapshot) variance estimate:

$n = \dfrac{z_{\alpha/2}^2 2s^2}{MoE^2}$	Equation 7

Where:

= required sample size

$z_{\alpha/2}$ = z-value of the normal distribution at confidence level 100(1-α)% (two-sided)

α = significance level (for the SOC FM 90%CI α= 0.1)

s^2 = estimate of the variance of 1. SOC stock change 2. SOC stock snapshot

MoE = Desired margin of error (degree of precision)

We strongly recommend that such derived sample sizes are the basis for any quantification approach (even model-based), as this allows for switching to a design-based approach if a model falls short during verification.

Box 6. Requirements for collecting SOC stock time series data

While high-quality time series data can greatly simplify the model validation process for SOC sequestration projects, it is quite difficult to obtain such data. Creating a time series of paired soil data points requires careful planning in both sampling and measurement. Between the multiple measurements, consistency is key, i.e., the same sampling and measurement protocols shall be applied and preferably also the same devices and personnel to carry out the work.

Sites shall be representative and accessible and shall be consistently sampled at fixed locations and regular intervals (at least in five-year intervals). Because of the well-documented micro-scale variability of SOC stocks, composite samples within a spatial radius are required to establish SOC stock time series data. The spatial radius to cover the composite shall be recorded and shall be at least as large as the spatial imprecision of the GPS/global navigation satellite system (GNSS) device used to measure the (central) spatial location of the composite sample. Standardised protocols shall be used for collecting soil samples at consistent depths. Repeated sampling campaigns shall consider the equivalent soil mass principle (see Table 2 in these guidelines). Proper sample handling, including preservation and accurate labelling, is essential to prevent degradation. Detailed records of sampling conditions and environmental factors shall be maintained, and a robust data management system shall be used to organise data. Quality control measures, such as duplicates and calibration of equipment, are necessary to ensure accuracy. Finally, appropriate statistical methods shall be applied to analyse temporal trends in the data. This approach ensures reliable, consistent, and interpretable soil data over time.

5.4.3 How to Acquire Data

Validation data that fulfils the requirements noted above can be obtained from different sources. Depending on which data source is used, the data will be more or less representative of the project area. The following shall be considered for SOC model validation:

- **Data Source A:** Direct SOC measurements[5] from *within* the project area
- **Data Source B:** Direct SOC measurements from *similar* reference sites *adjacent to* the project area covering the same management changes as in the project scenario over the same spatial and temporal extent
- **Data Source C:** Direct SOC measurements from *similar* reference sites *not adjacent to* the project area but covering the same management changes as in the project scenario over the same spatial and temporal extent
- **Data Source D:** Direct SOC measurements from *similar* reference sites *not adjacent to* the project area, covering the same management changes as in the project scenario but over a different spatial or temporal extent

It is strongly recommended to use validation data from Data Sources A or B to ensure maximal representativeness for the project area. Depending on the data source, additional uncertainty deductions might be applied by Gold Standard. The inclusion of Data Sources B, C, and D is a pragmatic approach for allowing project developers to validate and use models for estimation in the absence of validation data from the project site itself (i.e., in the first years of a project before resampling is conducted). As the project progresses and data is sampled from within the project's spatial boundary over time for verification, the model shall be revalidated on this data from within the project's spatial boundary to continue using the model for model-based estimates.

For Data Sources B, C, and D, the similarity of reference sites to the project area shall be justified on the basis of the respective modelling unit defined in accordance with Sections 6 and 7 of the SOC FM (pp. 13-20), based on a common climate zone, soil texture class, vegetation type, and management history. This justification shall be included in the model validation report to allow independent assessment of the representativeness of the validation data for the project area.

5.4.4 Outputs of Data Collection

The main output of the data acquisition step is electronic files with all collected SOC reference data points used for model calibration and validation, i.e.:

a) The measured SOC stock data points for all sampling units

b) The exact dates when the individual samples were collected (and, optionally, the dates when they were analysed in the laboratory)

c) The exact locations of the sampling units (point coordinates or exact spatial boundaries)

d) The exact soil depth intervals at which the samples were collected

Furthermore, the following metadata shall be recorded in the model validation report:

5 "Direct" means field-based, on-the-ground measurements using widely accepted soil sampling protocols (Aynekulu et al., 2011; FAO, 2020). Soil spectrometry may be accepted if the statistical relationship with traditionally accepted field measurements is characterised as strong and verified for the project area.

e) The full description of the sample preparation before the lab analysis (e.g., compositing of samples, homogenisation, sieving, drying)

f) The full description of the lab methods and measurements carried out (e.g., International Organization for Standardization [ISO] 10694 dry combustion, volumetric/gravimetric quantification of BD/coarse fragment ratio)

g) The uncertainty of the lab measurements (if possible, the standard error of the mean of the lab measurements)

Lastly, the model validation report shall contain:

h) Proof that the validation data and the calibration data are statistically independent as described in Section 5.4.2 of these guidelines

i) If Data Sources B, C, or D are used, justification that the sampled reference sites are similar to the project area, as described in Section 5.4.3 of these guidelines

5.5 Calibrate the Model

5.5.1 Rationale of Model Calibration

Calibration/training is a critical step in ensuring the accuracy and reliability of a SOC model for a sequestration project. It involves adjusting the model parameters so that its outputs closely match observed data. Data representative of the project context and algorithms can be used to test and adjust the model parameters, such that the model predictions match the observed validation data. This section is a guide for the process of calibrating a SOC model for a specific project.

Model calibration adjusts model parameters within plausible ranges to minimise the discrepancy between the model outputs and observed field data. This process fine-tunes the model to local conditions, enhancing its predictive accuracy.

Calibration is a crucial process for adapting a SOC model to a specific sequestration project, thereby enhancing its predictive accuracy and reliability. Carefully gathering data, understanding model parameters, and iteratively refining the model through calibration and validation ensures that a SOC model provides valuable insights for effective carbon sequestration management. Calibration is not just a technical exercise; it's a critical step in aligning the model with the real-world dynamics of a project site.

Calibration Challenges

- Data limitations: Limited or low-quality data can constrain calibration accuracy.
- Model structure: Inherent limitations in the model structure may prevent perfect calibration. Understanding these limitations is essential for interpreting model results.
- Computational resources: Some calibration methods, especially automated ones, may require significant computational resources.

5.5.2 Calibration Requirements

Considerations for Effective Calibration

- Data quality: Ensure the quality and representativeness of the data used for calibration. High-quality data leads to more reliable calibration.
- Temporal and spatial scale: Calibrate the model at a temporal and spatial scale that matches the project requirements.
- Model complexity: The complexity of the model can impact the ease and effectiveness of calibration. More complex models may offer more detailed insights but may require more extensive calibration.
- Uncertainty analysis: Acknowledge and analyse uncertainties that are inherent in the model and input data. Understanding these uncertainties is crucial for interpreting model predictions.

5.5.3 How to Calibrate a Model

A. Steps in SOC Model Calibration

a) Gather local data: Collect field data relevant to SOC dynamics in the project area, including soil carbon measurements, soil properties (texture, pH, moisture), climate data, land use history, and management practices. This data forms the basis for calibration.

b) Understand the model parameters: Get familiar with the parameters of the chosen SOC model. Understand which parameters are most influential and how they relate to the physical processes in the project area.

c) Make initial parameter setting: Begin with default or literature-based parameter values. These values provide a starting point for calibration.

d) Choose a calibration method: Common methods include manual calibration (trial and error), automated algorithms (e.g., genetic algorithms, Monte Carlo simulations, machine learning models), or a combination of both.

e) Calibrate against measured reference data: Adjust the model parameters to minimise the difference between model predictions and measured reference data. Section 5.4 of these guidelines, “Collect the Required Data,” provides details on how to obtain calibration data. Use the statistical metrics from Section 5.3.3 of these guidelines, “Model Selection Requirements,” to quantify the model performance on the calibration data.

f) Iterative Refinement: Calibration is an iterative process. Continuously refine the parameters and rerun the model until a satisfactory level of agreement is reached between the model outputs and observed data.

g) Validation: After calibration, validate the model using a held-out validation dataset as described in Section 5.4 of these guidelines, “Collect the Required Data.” Successful validation ensures that the model is reliable and not just overfitted to the calibration data set.

B. Outputs of Model Calibration

The foremost output is the final calibrated model. In addition, the calibration process shall be documented, and the documentation shall be submitted in the context of the project. The model's name and version shall be documented together with a description of the data and the method that was used to calibrate/train the model.

With the calibrated model, SOC stock snapshots or changes can be modelled in the project area and context. However, to officially use the model for quantification of SOC within the context of a SOC FM project, the model shall be appropriately validated, such that the uncertainty of model outputs and predictions can be quantified appropriately. Therefore, model validation, predictions, uncertainty quantification, and model verification are all covered in depth in the following section on uncertainty quantification.

6. UNCERTAINTY QUANTIFICATION

6.1 Validate the Model

This and the following sections cover the official validation, prediction, and verification process when a model shall be used for purely model-based estimation of the SOC stocks within a project's spatial boundaries. For simple design-based or model-assisted estimation, no official model validation or verification is required, as the SOC stock quantification is ultimately based on the sample of ground-measured observations. This holds true even when a model is used to quantify the variance to determine strata and the required sample size or for model-assisted estimation after the sampling process.

6.1.1 Rationale of Model Validation

The goal of model validation is to demonstrate that a SOC model is suitable for achieving the stated modelling objective for a given modelling unit within the project area. The SOC model selected for a modelling unit of the project area shall be able to accurately predict SOC stocks within that modelling unit and provide reliable uncertainty estimates. In the model validation process, the predictive accuracy of the model shall be assessed using independent observations (validation data) that are representative for the specific modelling unit and were not used for model selection or calibration. (See Section 5.4.2 of these guidelines, "Requirements for Data Collection.") The model validation metrics and thresholds are chosen to ensure that (1) the model is sufficiently calibrated to allow for the quantification of SOC stocks in the modelling units and (2) the model provides reliable uncertainty estimates.

6.1.2 Validation Metrics

The three metrics that shall be computed and assessed for model validation are:

1. Bias
2. Root mean squared error (RMSE)
3. Coefficient of determination (R^2 score/R-squared)

Bias describes the average difference between the model predictions and the measured data. *RMSE* represents the approximate average error of the model on the level of the prediction unit. The *R^2 score* describes the fraction of the variation in the measured data that is accounted for by the linear regression of predicted to ground-measured data.

Definitions and short descriptions of every metric are given below. All required equations are listed in Table 3 in these guidelines.

When computing these validation metrics, it is important to maintain consistency between the prediction unit of the model and the sampling unit of the validation data. Ideally, the prediction unit of the model is the same as the sampling unit of the validation data. For example, a model that provides farm-level SOC predictions should be validated using a dataset of farm-level SOC reference data, where each data point is a composite sample across a farm.

If the prediction unit of the model differs from the sampling unit of the validation data, <u>the finer unit shall be aggregated to match the coarser unit</u> before the validation metrics are computed. For example, if the model produces predictions for individual points (in geographical space) and the validation data contains farm-level SOC reference data, all point predictions within the boundaries of a farm shall be aggregated to produce a farm-level prediction. The validation metrics are then computed on a farm level.

<u>A SOC model can be validated only up to the spatial level of granularity (e.g., point, field, farm) of the sampling units of the validation data (or coarser), regardless of the prediction units.</u> For example, a SOC model that produces predictions for individual points but was validated on farm-level SOC reference data is considered as validated only for farm-level predictions. The point predictions cannot be viewed as reliable unless they are validated on point-level SOC reference data.

For simplicity, in the remainder of this section, it is assumed that the prediction units and the sampling units of the validation data are the same.

Bias

The Bias quantifies the structural error that leads to systematic overestimation or underestimation of the measured SOC reference data. The Bias is calculated by averaging the signed model errors/residuals (differences between model predictions and measured reference data). See Equation 17 in Table 3 in these guidelines. A positive Bias means the model tends to overestimate the measured values, whereas a negative Bias means the model tends to underestimate them. The absolute value of the Bias indicates how strongly the model overestimates or underestimates the mean of the measured reference data in units of the data (i.e., SOC stock [t/ha]). The Bias shall be close to zero to show that the model does not systematically overestimate or underestimate the measured values.

RMSE

The RMSE quantifies how close, on average, the predicted SOC values are to the observed SOC values. It is calculated by taking the square root of the mean squared error (MSE). See Equation 18 in Table 3 in these guidelines. Since the individual errors are squared before averaging, larger errors are more strongly penalised than smaller errors. The MSE can be decomposed into the squared Bias and the variance of the model errors; for this reason, the RMSE jointly captures the Bias and the precision of a model. The RMSE has the same unit as the data (i.e., SOC stock [t/ha]) and ranges from zero to infinity. Lower RMSE values indicate a better model fit; the closer to zero, the better.

R^2 score

The R^2 score quantifies the proportion of the variation of the measured reference data that is explained by the model. It can be used to assess how well a SOC model explains the variability of SOC on the ground. The R^2 is score is dimensionless and can take values between minus infinity and one. The value shall be positive for the model to be useful in practice. Higher R^2 values indicate a better fit of the model to the data; a value close to one means that the SOC model effectively explains all of the variability of SOC on the ground. When the mean of the ground measured observations is used as a prediction for all values in the dataset, the R^2 value will be zero. The R^2 score can be interpreted as a rescaled variant of RMSE that relates the model errors to the variance of the measured data. See Equation 19 in Table 3 in these guidelines.

Table 3. Accuracy and uncertainty metrics		
Symbol/Name	**Equation**	**Equation Number**
Single measured data point	X, where $X \in \mathbb{R}$	8
Sequence of measured data points and indexed data point	$(X_n)_{n \in \mathbb{N}} = (X_1, X_2, X_3, , X_n)$, where n = total number of data points or sample size and $X_i \in (X_n)_{n \in \mathbb{N}}$, where: i = unique index of an individual data point	9 9b 9c
Significance level	α, typically chosen as 0.05 to correspond to 95% confidence	10
Single prediction at the prediction unit	$\hat{X}$ where $\hat{X} \in \mathbb{R}$	11
Mean model prediction	$\bar{\hat{X}} = \frac{1}{n} \sum_{i=1}^{n} \hat{X}_i$	12

Table 3. Accuracy and uncertainty metrics		
Validation data mean	$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	13
Variance of the validation data	$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$	14
Standard deviation of the validation data	$s = \sqrt{s^2}$	15
Single model error/residual	$r_i = \hat{X}_i - X_i$	16
Bias	$Bias = \frac{1}{n}\sum_{i=1}^{n} r_i = \hat{\bar{X}} - \bar{X}$	17
RMSE	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{X}_i - X_i)^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} r_i^2}$	18
R^2	$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{X}_i - X_i)^2}{\sum_{i=1}^{n}(\bar{X} - X_i)^2}$	19
Standard error of the validation data mean	$SE = \sqrt{\frac{1}{n}s^2}$	20
Margin of error (MoE) of the two-sided 90% confidence interval around the validation data mean	$MoE = SE \times t_{(1-0.9)/2,n-1}$	21
Prediction interval with coverage probability $1-\alpha$	$I_{1-\alpha}(\hat{X}_i) = [a_i, b_i]$ where a_i is the upper bound of the interval (how these are calculated depends on the model)	22
Prediction interval coverage probability (mean of the interval coverage indicator function)	$PICP = \frac{1}{n}\sum_{i}^{n} \delta(a_i \leq X_i \leq b_i)$ with the indicator function	23a

Table 3. Accuracy and uncertainty metrics		
	$\delta(t) = \begin{cases} 1, \text{if t is TRUE} \\ 0, \text{else} \end{cases}$	23b
Standard deviation of the interval coverage indicator function	$s_{PICP} = \sqrt{\frac{1}{n-1}\sum_i (\delta(a_i \leq X_i \leq b_i) - PIC$	24
Test statistic for the t-test to verify that the prediction interval coverage probability (PICP) is no less than 90%	$t = \frac{PICP - 0.9}{s_{PICP}/\sqrt{n}}$	25

6.1.3 Requirements for Model Validation

For successful model validation, all of the metrics described above shall be computed and assessed on independent validation data, i.e., data that was not used for model selection or calibration. Section 5.4 of this document, "Collect the Required Data," details the exact requirements for validation data and how it should be collected. If the modelling objectives of the project differ across modelling units, model validation shall be carried out separately for every modelling unit. All results of the model validation process shall be collected in a model validation report and made available for independent review.

In a best-case scenario, there is an abundance of data that can be split into multiple validation datasets to (1) allow validation of candidate models during model selection and calibration and (2) for final model validation. If there is data scarcity, cross-validation can be used (Arlot and Celisse, 2010). See "Annex B: Requirements for Cross-Validation" in these guidelines.

Once the final validation data has been determined, the metrics from Section 6.1.2 of these guidelines, "Validation Metrics," shall be computed on that data. The following sections provide details on how the metrics shall be reported and which thresholds shall be satisfied.

Requirements for Bias

The model validation report shall demonstrate that the Bias of the model is close enough to zero. More specifically, the absolute value of the Bias on the validation data shall not be greater than the half-width of the two-sided 90% confidence interval around the validation data mean. This requirement ensures that the mean model prediction over the validation data lies within the confidence interval around the validation data mean and that the overall model error is smaller or equal to the accepted margin of error, as noted in Section 9 of the SOC FM (pp. 26-27).

The model validation report shall contain a table showing the following:

a) Number of validation data points

b) Mean model prediction over the validation data points (see Equation 12 in these guidelines)

c) Validation data mean (see Equation 13 in these guidelines)

d) Standard error of the validation data mean (see Equation 20 in these guidelines)

e) Half-width of the two-sided 90% confidence interval around the validation data mean (see Equation 21 in these guidelines)

f) Bias value (see Equation 17 in these guidelines)

"Example 1: Validation of Bias" in these guidelines demonstrates how the validation results can be presented in practice, including a visualisation of the above quantities. It is recommended but not required to add such a visualisation to the model validation report to facilitate interpretation of the results.

Requirements for RMSE and R^2 score

The model validation report shall demonstrate that the model captures variability of the validation data so that the individual model predictions are better than a mean prediction.

More specifically, the RMSE computed on the validation data shall be less than the standard deviation of the validation data (Chai and Draxler, 2014). Equivalently, the R^2 score shall be greater than zero.

The model validation report shall clearly state the relevant quantities and visualise the relationship between model predictions and measured data. In particular, the report shall contain a table showing:

a) Number of validation data points

b) Standard deviation of the validation data (see Equation 15 in these guidelines)

c) RMSE computed on the validation data (see Equation 18 in these guidelines)

d) R^2 score computed on the validation data (see Equation 19 in these guidelines)

Furthermore, the validation report shall contain two scatter plots, each with corresponding linear regression lines (optionally including their confidence intervals) and 1:1 lines for comparison:

1. A regression plot shows the measured reference values of the validation data on the vertical axis (y-axis) and the model predictions on the horizontal axis (x-axis); the 1:1 line is the straight line that satisfies y = x.
2. A residual plot shows the residuals on the vertical axis (y-axis) and the model predictions on the horizontal axis (x-axis); the 1:1 line is the straight line that satisfies y = 0.

Systematic differences between the above-referenced plots and the corresponding 1:1 lines may indicate model misspecification and shall be explained in the report. If there is evidence for heteroscedastic residuals in the plots (e.g., larger residual variance for larger predictions), the model validation report shall state how heteroscedasticity is addressed (e.g., by log-transforming the data), particularly in the light of providing conservative estimates for ΔCSOC, t-0.[6]

Example 2 demonstrates how the validation results can be presented in practice and is based on the same data as Example 1. Both examples follow.

Example 1: Validation of Bias

Table E1.1. Bias validation results	
Number of validation data points	93
Mean model prediction	10.29 t/ha
Measured validation data mean	11.81 t/ha
Standard error of the validation data mean	1.19 t/ha
Half-width of the 90% confidence interval around the validation data mean	±2.36 t/ha
Bias	-1.52 t/ha

6 The presence of systematic patterns or heteroscedasticity in the regression plot or residual plot indicates that model residuals are not statistically independent of the model predictions. A lack of this statistical independence would invalidate the statistical analyses commonly performed with empirical SOC models to quantify uncertainties of SOC stock or SOC stock change predictions, including model-based geostatistical approaches (Szatmári et al., 2021) and design-based approaches as described in Annex A.3.

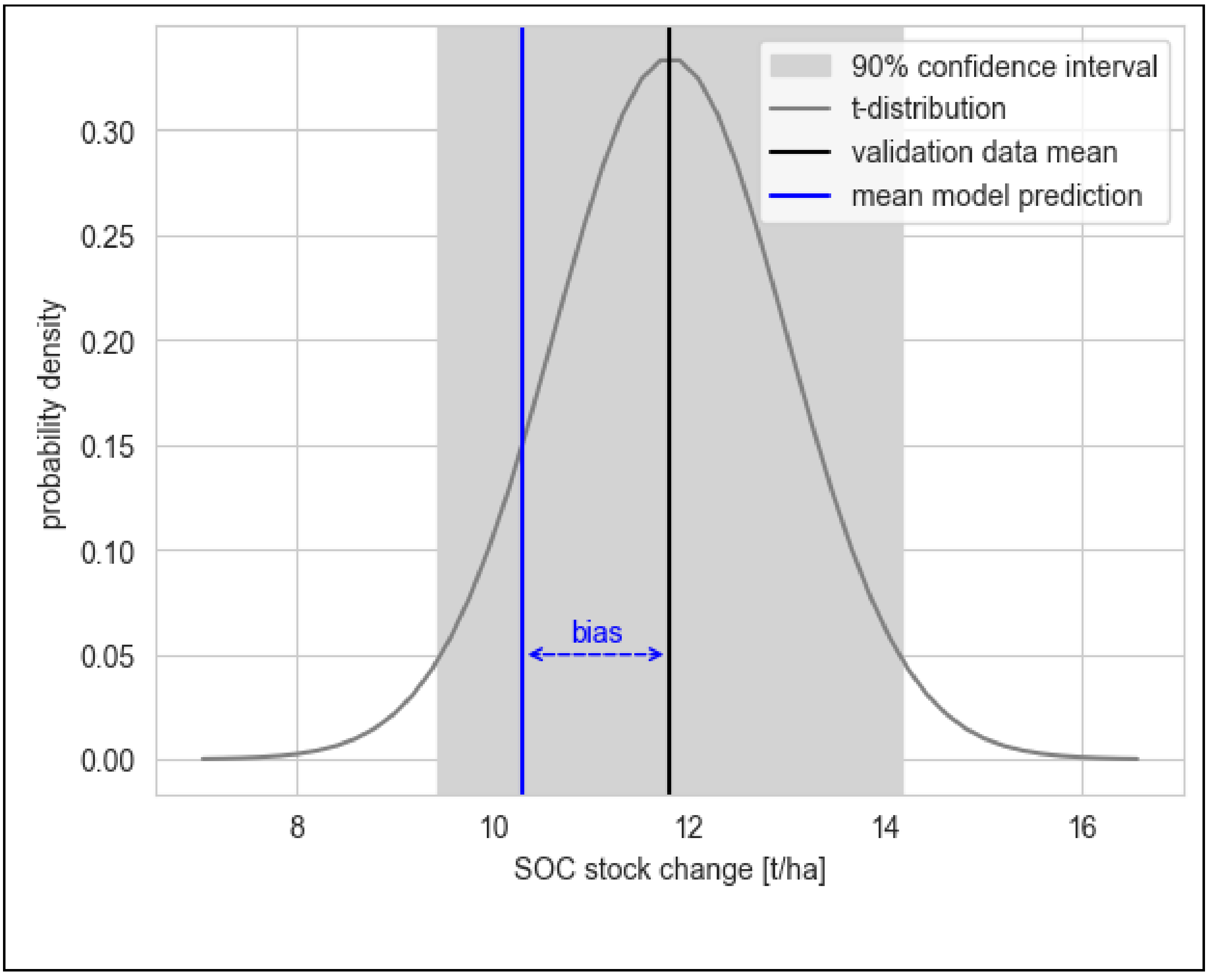


Figure E1.2. Visualisation of the Bias validation results

The absolute value of the Bias (1.52 t/ha) is not greater than the half-width of the 90% confidence interval around the estimated validation data mean (2.36 t/ha). The mean model prediction lies within the confidence interval. The location and scale of the t-distribution are given by the validation data mean and its standard error; the degrees of freedom are given by n - 1, where n is the number of validation data points.

Example 2: Validation of RMSE and R^2 score

Table E2.1. RMSE and R^2 score validation results	
Number of validation data points	93
Standard deviation of the validation data	11.42 t/ha
RMSE	5.31 t/ha
R^2 score	0.78

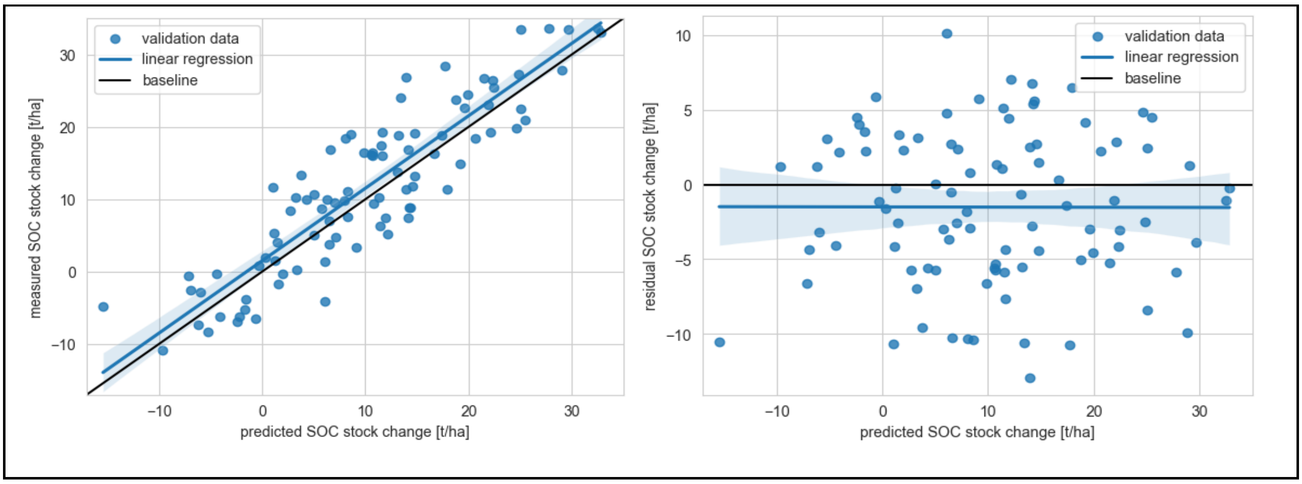


Figure E2.2. Scatter plots for RMSE and R^2 validation

The plot on the left demonstrates that the model well approximates the measured values; the variability around the regression line is small compared to the total variability. In both plots, the regression lines are parallel to the respective 1:1 lines but do not match the 1:1 lines exactly due to the non-zero Bias. Apart from the Bias, there is no apparent structure in the plots, meaning that the model is well specified. In particular, there is no evidence for heteroscedasticity.

6.1.4 How to Validate a Model

Steps in the Final Model Validation

For every modelling unit of the project, a suitable model shall be validated by performing the following steps:

1. Use a hold-out validation dataset (a dataset that is independent of the data used at any stage during model calibration), as outlined in Section 5.4 of these guidelines, "Collect the Required Data."
2. Apply the final calibrated model on the validation data to obtain model predictions for every validation data point. The exact details of this step are model-dependent.
3. Calculate all validation metrics noted in Section 5.3.3 of these guidelines, "Model Selection Requirements," as well as the threshold values and visualisations in Section 5.3.4 these guidelines, "How to Select a Model."
4. Make sure that all requirements from "How to Select a Model" are satisfied.

Note that a single model may be applied in multiple modelling units but that the model shall be validated separately for each modelling unit in which it is applied.

In cases where model validation does not succeed, other quantification approaches shall be used for the project, e.g., the **design-based estimation** approach (Approach 1 in the SOC FM) or the **model-assisted estimation** approach (as noted in Box 1 in these guidelines). Potentially, collecting more in situ data from the project area will enable better calibration and a successful validation of the model for later calculation/crediting periods. There is also the possibility of retrying validation with a different model selection, as described in Section 5.3 of these guidelines, "Select an Appropriate Model."

6.1.5 Outputs of Model Validation

A model validation report shall include for every modelling unit of the project:

a) A table summarising the Bias validation result as described in Section 5.3.4 of these guidelines, "How to Select a Model," with an optional visualisation of the quantities (see Example 1 in these guidelines).

b) A table and two scatter plots summarising the RMSE and R^2 score validation results as described in "How to Select a Model" (see Example 2 in these guidelines).

c) If cross-validation is used, a table summarising the distances between cross-validation folds as described in "How to Select a Model" (see Example 3 in these guidelines).

d) All data used to calculate the validation metrics (i.e., all model predictions with the corresponding measured reference values) shall be made available, along with the model validation report in a format that allows simple reproduction of the model validation results by an independent reviewer. Geographical locations, time stamps, soil depths, and any other information that would uniquely identify the data points in space and time may be omitted for reasons of confidentiality.

6.2 Make Predictions and Estimate the Uncertainty of the Predictions

6.2.1 Rationale of Predictions and Uncertainty Estimation

Upon successful model selection, calibration, and validation, the SOC models shall be applied in the project context to issue carbon credits for the quantification period. For that purpose, model predictions shall be made for every prediction unit in every modelling unit, and all quantities required to issue carbon credits for the project shall be derived from these model predictions. In particular, the uncertainties of the predictions and all derived quantities shall be estimated to apply the required uncertainty deductions described in Section 9 of the SOC FM. An additional safety discount based on the source of the validation data shall be applied to account for the uncertainty about the applicability of a model in the project context.

Uncertainty in SOC model outputs arises from various sources, including measurement errors in input data, variability in environmental conditions, model structural inadequacies, and parameter uncertainties. All of these uncertainties can propagate through the model and affect the model predictions and derived quantities. Therefore, statistically sound quantification of the uncertainties of the SOC model outputs is critical for any model-based SOC sequestration project.

Details on how uncertainties shall be expressed and which uncertainties shall be quantified are given below, along with examples for different uncertainty quantification approaches. Whether an uncertainty quantification approach is suitable for a specific model highly depends on the model and must be theoretically well justified. The chosen quantification approach shall be tested for its reliability within the project context as part of the model validation process. (See Section 5.3 of these guidelines, "Select an Appropriate Model.")

A type of higher-level uncertainty that is particularly hard to quantify is the uncertainty about whether the model is applicable in the project context (and provides reliable uncertainty estimates in the first place). Model applicability crucially depends on the representativeness of the validation data for the project context. Therefore, these guidelines define an additional safety discount depending on the data source of the validation data. (See Section 5.4.3 of these guidelines, "How to Acquire Data.") The safety discount must be added to the uncertainty deductions noted in Section 9 of the SOC FM.

6.2.2 Requirements for Predictions and Uncertainty Estimation

The following SOC model outputs must be computed:

a) SOC stock predictions for every prediction unit of every modelling unit of the project
b) The mean of these SOC stock predictions over the entire project area

Uncertainty estimates shall be provided for all of these SOC model outputs. Uncertainties shall be expressed using two-sided (bounded) intervals that contain the quantity of interest with a nominal coverage probability of 90%. The exact way the intervals are constructed and interpreted depends on the model and the underlying probability interpretation (frequentist or Bayesian). The following types of intervals are allowed for uncertainty quantification:

- **Confidence intervals**, whenever the model regards a quantity as a fixed but unknown population parameter (frequentist perspective)
- **Prediction intervals**, whenever the model regards a quantity as an unobserved random variable (frequentist perspective)
- **Credible intervals (prior intervals, posterior intervals)**, whenever the model regards a quantity as an unobserved random variable (Bayesian perspective)

Which type of interval is suitable for a specific SOC model output depends on the model:

- In a classical regression approach based on probability samples, the individual predictions would be considered random variables (with associated *prediction intervals*), while the mean values and mean difference values would be considered fixed but unknown population parameters (with associated *confidence intervals*).
- When following a geostatistical model-based approach, the individual predictions at each modelling unit and their means would be random variables, and *prediction intervals* for all of these quantities would be provided.
- In a process-based model, uncertainties would be computed in a Bayesian way by assuming prior distributions on the input parameters and constructing *credible intervals* for all quantities using Monte Carlo simulations (effectively treating all model outputs as unobserved random variables).

The process of model validation should ensure that the model is suitable for the declared modelling objectives within the project’s spatial boundaries. For model-based estimation approaches, spatial, temporal, and spatial-temporal autocorrelation shall be properly evaluated and taken into account during prediction and uncertainty estimation. For an example of how to do this, see “Annex A: Quantification of SOC Stock Change Using Digital Soil Mapping” in these guidelines.

6.2.3 How to Predict and Estimate Uncertainties

The exact steps and equations to obtain the SOC model outputs and associated uncertainties described above are highly model-dependent. Therefore, only general guidance is given here:

1. Run the SOC model(s) on every prediction unit of every modelling unit to obtain SOC stocks for the start and end of the respective calculation period.[7]
2. Calculate the total SOC stocks for the project scenario across all modelling units with Equation 6 of the SOC FM (p. 20) for the start and end of the respective calculation period.
3. Estimate the uncertainties for the mean SOC stock change using well-established statistical methods from textbooks or peer-reviewed literature that are suitable for the specific SOC model. The methods may involve one or more of the following:
    - **Monte Carlo simulations** to obtain credible intervals for process-based model outputs: Run the model multiple times with randomised inputs and internal parameter sets that follow strong prior distributions. Use the empirical quantiles of the simulated SOC model outputs to construct interval bounds.
    - **Bootstrapping** to obtain confidence intervals for population parameters in frequentist models: Create multiple bootstrap samples from the calibration data by random selection with replacement and estimate the parameter independently from every bootstrap sample. Use the empirical quantiles of the bootstrap replicates of the parameter estimate to construct interval bounds.
    - **Quantile Regression** (Koenker, 2005) to obtain prediction intervals in a regression model: Algorithms such as the Quantile Regression Forest (Meinshausen, 2006) or Conformalised Quantile Regression (Romano et al., 2019) directly estimate prediction intervals from covariates.
4. Calculate the model output uncertainty (UNC) defined in the SOC FM, Chapter 9, Step 3, Equation 11, by dividing the half-width of the interval around the SOC stock change prediction by the total SOC stock change estimate.

7 Model-based estimation SOC stock predictions must be sent to the Gold Standard Secretariat and the VVB before the validation/verification data is sampled from the project area. This means that there can be a long time period between running the model for the start and end of the calculation period. On request by the VVB, the exact reproducible script or sequence of steps for running the model for any prediction used for quantification purposes in the project must be made available by the project developer.

5. Obtain the total uncertainty (TU) by adding a safety discount to UNC based on the data source of the validation data (Table 4 below) and proceed with Step 4 in the SOC FM, Chapter 9, using the TU.

Table 4. Safety discounts based on data source	
Data Source	Safety discount
A	0%
B	10%
C	30%
D	50%

Table 4 provides a heuristic orientation for the safety discount, and the values of the safety discounts by data source for validation may be adjusted in coordination with the Gold Standard Secretariat and the respective project's VVB. As the project continues, model validation shall be conducted on Data Source A as soon as it becomes available within the project.

6.2.4 Outputs of Predictions and Uncertainty Estimation

a) All SOC model outputs listed in Section 6.2.3 of these guidelines, "How to Predict and Estimate Uncertainties," along with intervals that capture their uncertainties

b) UNC as defined in the SOC FM

c) TU that combines UNC and the safety discount

6.3 Verify the Model Predictions

6.3.1 Rationale for Model Verification

Verifying the predictive performance of a SOC model within the context of a specific soil carbon sequestration project is the final critical step in model application. This process involves comparing the model predictions with actual outcomes to assess accuracy and reliability in real-world conditions. This is sometimes referred to as "model true-up" (Lavallee et al., 2024). This section provides a comprehensive approach to verifying the prediction performance of a SOC model in a GS4GG SOC FM project.

Understanding Model Verification

Verification differs from calibration and validation as it specifically focuses on how well the model performs in predicting the quantities to calculate SOC stock change outcomes in the project.

Verifying the prediction performance of a locally calibrated and validated SOC model is essential to ensure its effectiveness in a specific soil carbon sequestration project. This process not only tests the model accuracy in real-world scenarios but also builds confidence among stakeholders in its predictions. Successful verification involves a thorough comparison of model outputs with actual SOC data, continuous monitoring, and adaptability to refine the model as more data becomes available. Rigorously verifying the SOC model enhances its value as a reliable tool for guiding soil carbon sequestration efforts and contributing to broader climate change mitigation strategies.

6.3.2 Requirements for Model Verification

Considerations for Effective Verification

a) Data quality and representativeness: The accuracy of model verification is heavily dependent on the quality and representativeness of the independent data used. Only data from Data Source A in Section 5.4.3 of these guidelines, "How to Acquire Data," is allowed for model verification.

b) Realistic expectations: Set realistic expectations for model performance. All models have inherent limitations and uncertainties.

c) Adaptability: Be prepared to adapt the model if verification reveals significant discrepancies between predictions and observations.

d) Stakeholder engagement: Involve stakeholders in the verification process to ensure transparency and build trust in the model's predictions.

Challenges in Model Verification

a) Data limitations: Limited availability of relevant, high-quality independent data can be a significant challenge.

b) Dynamic environmental conditions: Changes in environmental conditions, such as unusual weather events or management practices, can impact model performance.

c) Model complexity: Highly complex models might perform well in calibration and validation but might struggle in real-world applications.

6.3.3 How to Verify a Model

Steps in Model Verification

a) Re-establish the verification criteria:

1. The model must be unbiased (Bias < half-width of 90% confidence interval).
2. The model error must be small (RMSE < Standard deviation of the verification data).
3. The model must capture variability in the data ($R^2 > 0$).

b) Collect and prepare independent data: Gather independent SOC data from the project site (Data Source A) that was not used in the model calibration or validation. This data shall represent the conditions under which the model is expected to operate. The data shall be collected in line with well-established soil sampling protocols (Aynekulu et al., 2011; FAO, 2020).

c) Run the model for prediction: Use the model to make predictions for the period and conditions corresponding to the independent data collected.

d) Compare predicted and observed data: Compare the model's predictions with the observed SOC data. Quantify the model's predictive performance using the metrics from Section 6.1.2 of these guidelines, “Validation Metrics”: Bias, RMSE, R^2 score.

e) Analyse discrepancies: If discrepancies exist between predicted and observed values, analyse these differences to understand their causes. This may involve examining model assumptions, data quality, or unexpected environmental changes. If the model violates any of the verification criteria, the project proponent must submit a plan to address potential over- or under-issuance of Gold Standard Verified Emission Reductions (GS-VERs) due to the model’s predictions.

f) Continuously monitor and update: Model verification is an ongoing process. The model's performance shall be verified at least every five years, and the model shall be updated as more data becomes available or as project conditions change.

6.3.4 Outputs of Model Verification

The required output of verification must be the verification data with all its metadata as described in Section 5.4.4 of these guidelines, “Outputs of Data Collection.” Additionally, the calculated verification metrics and documentation of whether the criteria were met shall be submitted.

In case the model verification criteria were not met, a plan shall be submitted to describe how over- or under-crediting (issuance of GS-VERs) due to the model predictions is addressed by the project developer. The latest version of the performance shortfall guidelines[8] shall provide the specific requirements for such potential over- or under-crediting. If model predictions have failed verification, the project developer must notify the Gold Standard Secretariat and VVB immediately. Within a 90-day period, the project developer can conduct a root-cause analysis of the model’s shortcomings, improve the model via recalibration, and attempt to revalidate. Upon successful revalidation, the model can be applied for model-based estimations in the project context and same verification cycle/stage again. The revalidation shall follow the same procedure as noted in Section 6.1 of these guidelines and shall include the data from the verification event where the model predictions have failed. If the model cannot be revalidated, the project developer shall switch the quantification approach to direct measurements until a new version of the model can successfully be validated and applied again in the project context.

7. MONITORING METHODOLOGY

7.1. DATA AND PARAMETERS MONITORED

Parameter ID	**1**

8 Gold Standard Performance shortfall guidelines: https://globalgoals.goldstandard.org/501g-pr-performance-shortfall-guidelines/

Data/Parameter:	$\Delta C_{SOC,\ t\text{-}0}$
Data unit:	Mass of SOC in tonnes (Mg)
Description:	SOC mass change between time t=t and t=0
Source of data:	Directly modelled or derived as difference of SOC_t - SOC_0 (See Equation 2)
Measurement procedures (if any):	As appropriate to the measurement / estimation procedure.
Monitoring frequency:	Every calculation period
QA/QC procedures:	-
Any comment:	-

Parameter ID	**2**
Data/Parameter:	SOC_t
Data unit:	Total mass of SOC in tonnes (Mg)
Description:	SOC mass snapshot at time t=t
Source of data:	Design-based (sampling), model-assisted, or model-based estimate
Measurement procedures (if any):	As appropriate to the measurement / estimation procedure.
Monitoring frequency:	Every calculation period
QA/QC procedures:	-
Any comment:	-

Parameter ID	**3**
Data/Parameter:	UD
Data unit:	Unitless ratio ([0,1])
Description:	Uncertainty deduction: proportion of $\Delta C_{SOC,\ t\text{-}0}$ that cannot be issued as credits and that shall be deducted because of the quantification uncertainty
Source of data:	Uncertainty estimation in line with a) design-based (sampling), b) model-assisted, or c) model-based estimates as noted in Section 6.2 of these guidelines.
Measurement procedures (if any):	-

Monitoring frequency:	Every calculation period
QA/QC procedures:	-
Any comment:	-

Parameter ID	**4**
Data/Parameter:	$M_{OC(DL)}$/SOC stock
Data unit:	Mass of SOC in tonnes per hectare (ha) (Mg/ha)
Description:	SOC stock: mass of SOC C per unit area over a fixed depth column
Source of data:	Modelled directly or calculated from Equation 3b, 4, or 5b in these guidelines
Measurement procedures (if any):	Measurements shall be in line with the FAO GSOC MRV protocol.
Monitoring frequency:	Every calculation period
QA/QC procedures:	-
Any comment:	-

Parameter ID	**5**
Data/Parameter:	$M_{SOIL(DL)}$
Data unit:	Equivalent mass of fine soil in tonnes per ha (Mg/ha)
Description:	Mass of the fine soil (dried and sieved)
Source of data:	Measured from whole soil core used for organic carbon analysis, dried and sieved
Measurement procedures (if any):	As appropriate to the measurement / estimation procedure.
Monitoring frequency:	Every calculation period, if a variant of design-based estimation is used
QA/QC procedures:	-
Any comment:	-

Parameter ID	**6**
Data/Parameter:	$C_{OC(DL)}$/SOCcon

Data unit:	SOC concentration in % (gC/gSoil * 100%)
Description:	SOC concentration as measured by dry combustion and inorganic carbon removed
Source of data:	Measured fine soil mass (g, whole sample, dried and sieved)
Measurement procedures (if any):	As appropriate to the measurement / estimation procedure.
Monitoring frequency:	Every calculation period, if a variant of design-based estimation is used
QA/QC procedures:	-
Any comment:	-

Parameter ID	**7**
Data/Parameter:	FSS
Data unit:	FSS in tonnes per hectare (Mg/ha)
Description:	Fine soil stock, calculated as the ratio of the mass of a dried and sieved soil core per the known volume of the untreated soil core
Source of data:	See Equation 3a in these guidelines.
Measurement procedures (if any):	As appropriate to the measurement / estimation procedure.
Monitoring frequency:	Every calculation period, if the FSS method is used to determine the SOC stock
QA/QC procedures:	-
Any comment:	-

Parameter ID	**8**
Data/Parameter:	$BD_{fine\ soil}$
Data unit:	BD of the fine soil in grams per cm^3 (g/cm^3)
Description:	Bulk Density, the mass of soil per unit volume, measured on a sub-sampled aliquot
Source of data:	Measured with the intact core method on field sampled data
Measurement procedures (if any):	As appropriate to the measurement / estimation procedure.

Monitoring frequency:	Every calculation period, if the BD method is used to determine the SOC stock (see Equation 4 in these guidelines)
QA/QC procedures:	-
Any comment:	-

Parameter ID	**9**
Data/Parameter:	n
Data unit:	Count (unitless)
Description:	(Estimated) number of (composite) samples in a design-based estimation (sample size)
Source of data:	Estimated with Equations 6, 7 in these guidelines or determined by constraints of feasibility
Measurement procedures (if any):	As appropriate to the measurement / estimation procedure.
Monitoring frequency:	Every calculation period, if a variant of design-based estimation is used
QA/QC procedures:	-
Any comment:	-

Parameter ID	**10**
Data/Parameter:	α
Data unit:	Significance level (%)
Description:	The probability of incorrectly rejecting the null hypothesis (calling an effect significant although it is not)
Source of data:	Determined by project developer
Measurement procedures (if any):	As appropriate to the estimation procedure.
Monitoring frequency:	Once determined, but reported every calculation period, if α changes for any reason
QA/QC procedures:	-
Any comment:	-

Parameter ID	**11**
Data/Parameter:	X_i

Data unit:	Instance of a random variable (e.g., SOC stock or SOC stock change, units of the variable of interest apply, e.g., Mg/ha)
Description:	Random variable, e.g., SOC stock or SOC stock change
Source of data:	Ground measured data (+ derived through Equations 3b, 4, 5a in these guidelines)
Measurement procedures (if any):	As appropriate to the measurement / estimation procedure.
Monitoring frequency:	Every calculation period, if some form of design-based estimation is used
QA/QC procedures:	-
Any comment:	-

Parameter ID	**12**
Data/Parameter:	$s^2(X)$; s^2 for short
Data unit:	(Estimated) variance of a random variable (squared unit of measurement, e.g., $(Mg/ha)^2$ for SOC stock variance)
Description:	(Estimated) variance of a random variable—measure of variability of the random variable around the arithmetic mean (sum of squared differences of measured data from the mean)
Source of data:	Estimated (e.g., through model predictions or based on design-based estimation) via Equation 14 in these guidelines
Measurement procedures (if any):	-
Monitoring frequency:	Every calculation period, if some form of design-based estimation is used
QA/QC procedures:	-
Any comment:	-

Parameter ID	**13**
Data/Parameter:	$\overline{X}$
Data unit:	Arithmetic mean (e.g., SOC stock or SOC stock change; units of the variable of interest apply, e.g., Mg/ha)
Description:	Arithmetic mean of measured instances of a random variable
Source of data:	See Equation 13 in these guidelines.

Measurement procedures (if any):	-
Monitoring frequency:	Every calculation period, if some form of design-based estimation is used
QA/QC procedures:	-
Any comment:	-

Parameter ID	**14**
Data/Parameter:	$SE(\overline{X})$
Data unit:	Standard error of the mean (e.g., SOC stock or SOC stock change; units of the variable of interest apply, e.g., Mg/ha)
Description:	Uncertainty measure of ground-measured (validation) data: standard error as standard deviation of the measured data divided by the square root of the number of samples n
Source of data:	See Equation 20 in these guidelines.
Measurement procedures (if any):	-
Monitoring frequency:	Every calculation period, if some form of design-based estimation is used
QA/QC procedures:	-
Any comment:	-

Parameter ID	**15**
Data/Parameter:	$\widehat{X}_i$
Data unit:	Instance of a model prediction of a random variable (e.g., SOC stock or SOC stock change; units of the variable of interest apply, e.g., Mg/ha)
Description:	Instance of a model prediction of SOC stock or SOC stock change
Source of data:	Model prediction
Measurement procedures (if any):	-
Monitoring frequency:	Every calculation period, when modelling is used for deriving the variance or mean estimate of SOC stock or SOC stock change
QA/QC procedures:	-

Any comment:	-

Parameter ID	**16**
Data/Parameter:	r_i
Data unit:	Instance of a residual (e.g., SOC stock or SOC stock change; units of the variable of interest apply, e.g., Mg/ha)
Description:	Instance of a residual of SOC stock or SOC stock change: difference between a measured instance of a random variable and a prediction of that instance of the random variable
Source of data:	See Equation 16 in these guidelines.
Measurement procedures (if any):	As appropriate to the measurement / estimation procedure.
Monitoring frequency:	Every calculation period, when modelling is used
QA/QC procedures:	-
Any comment:	-

Parameter ID	**17**
Data/Parameter:	Bias
Data unit:	Units of the random variable in question, e.g., SOC stock or SOC stock change: Mg/ha
Description:	Model structural error: arithmetic mean of the residuals (without squaring residuals or taking the absolute value)
Source of data:	See Equation 17 in these guidelines.
Measurement procedures (if any):	-
Monitoring frequency:	Every calculation period, when modelling is used
QA/QC procedures:	-
Any comment:	-

Parameter ID	**18**
Data/Parameter:	RMSE
Data unit:	Units of the random variable in question, e.g., SOC stock or SOC stock change: Mg/ha

Description:	Square root of the arithmetic mean of the squared residuals, showing the average error of the model and penalising greater residuals over lower residuals
Source of data:	See Equation 18 in these guidelines.
Measurement procedures (if any):	-
Monitoring frequency:	Every calculation period, when modelling is used
QA/QC procedures:	-
Any comment:	-

Parameter ID	**19**
Data/Parameter:	R^2 score
Data unit:	Unitless ratio
Description:	Shows relationship of the mean squared error to the variance in the measured data
Source of data:	See Equation 19 in these guidelines.
Measurement procedures (if any):	-
Monitoring frequency:	Every calculation period, when modelling is used
QA/QC procedures:	-
Any comment:	-

8. ANNEXES

Annex A: Quantification of SOC Stock Change Using Digital Soil Mapping

Recent advances in digital soil mapping (DSM) have made it possible to quantify SOC stocks on a farm scale, regional scale, and global scale with decent accuracies. These DSMs are often produced with remote sensing-based statistical models and can be recomputed annually as soon as the required remote sensing data becomes available. For this reason, DSM is an appealing approach to quantifying SOC stock changes over time.

This section exemplifies (1) how to validate a statistical model for DSM in the project context and (2) how to obtain statistically sound SOC stock change estimates from SOC stock maps produced by that model, including uncertainties. The statistical analysis follows a model-based approach and can be applied only when probability samples from the project area are available.

A.1 Simulated Example

To illustrate the approach, the spatial distribution of SOC stocks in a fictitious project area ("ground-truth") and DSMs produced by a SOC model were simulated for two points in time, 2015 and 2020, before a hypothetical project start. In the simulation, it is assumed that SOC is in an equilibrium state and has roughly the same mean (~50 t/ha) and the same standard deviation (~12 t/ha) in both years. Figure A.1 visualises the simulated ground-truth, model predictions, and model residuals, i.e., the differences between ground-truth and predictions. In practice, only the model predictions (DSMs) are available for the complete project area. Information about the ground-truth and residuals must be obtained by collecting soil samples.

A.2 Model Validation With SOC Stock Data

Validation of the model is demonstrated using SOC stock data only (not SOC stock change data). It is assumed that 100 soil samples have been taken with simple random sampling[9] in 2015 and another 100 samples in 2020 for the sole purpose of model validation. These samples have not been used to calibrate the SOC stock model that is used to produce the DSMs.

Figure A.2 depicts the locations of the soil samples taken in both years. It also shows the regression plots and residual plots mentioned in Section 5.3.4 of this document,

9 Using simple random sampling or stratified simple random sampling with proportional allocation. Neyman allocation is not compatible with the statistical analysis used in this example. Random locations shall be selected independently in every time step. A sampling design in which the same locations are revisited in subsequent sampling rounds would also invalidate the analysis.

"How to Select a Model," in this case, separately for the SOC stock predictions in the two years. Figure A.3 shows the regression plot and residual plot jointly for all 200 SOC stock validation data points. Table A.4 shows the validation metrics and other quantities listed in "How to Select a Model," computed separately for the SOC stock predictions in both years and as well as jointly for the two years.

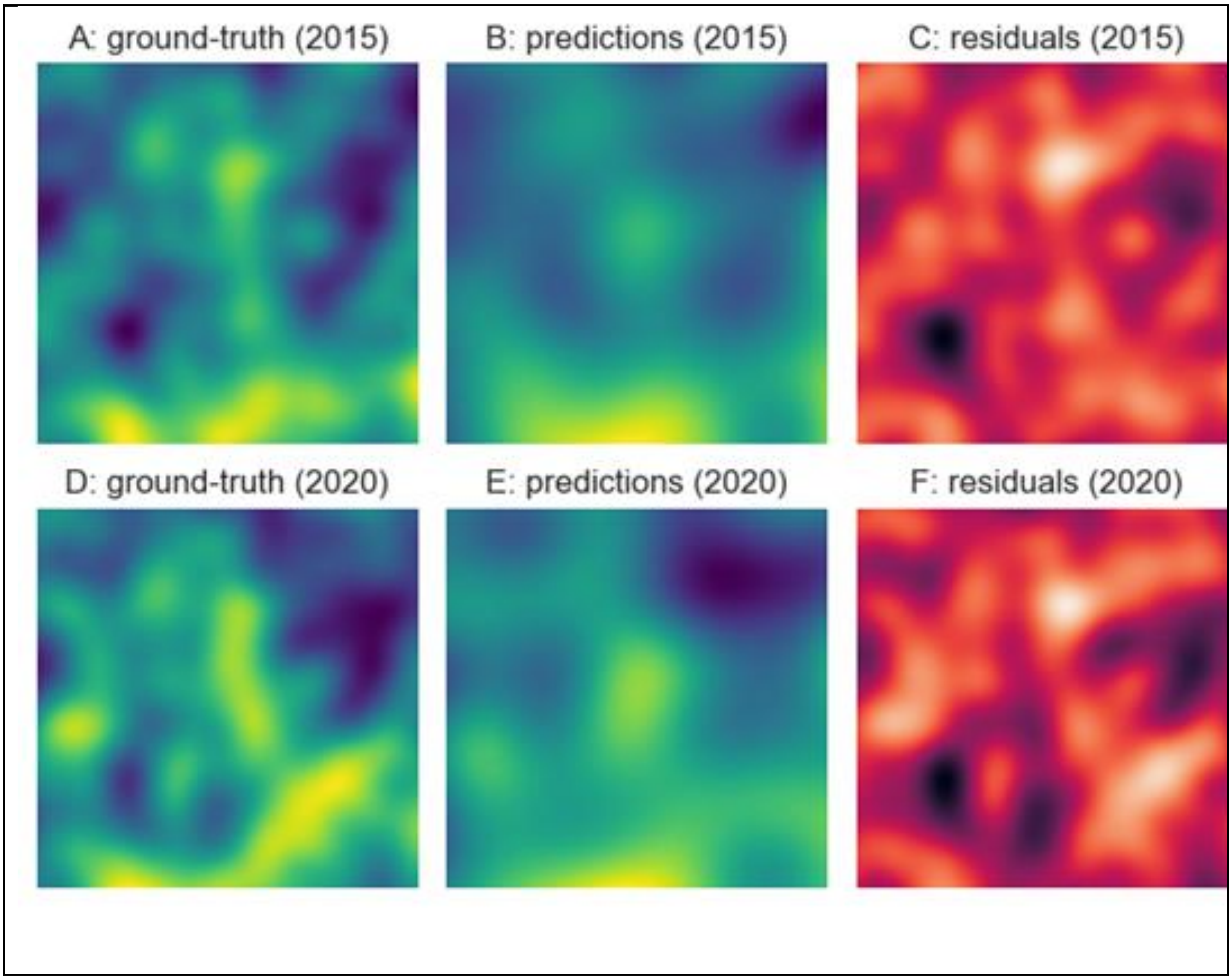


Figure A.1. Simulated ground-truth that shows the spatial distribution of SOC stocks in a fictitious project area in 2015 and 2020, before project start (**A, D**). Simulated model predictions (DSM) for the project area at the same points in time (**B, E**). Model residuals (**C, F**).

The regression plots and scatter plots closely follow the expected patterns (diagonal lines and horizontal lines, respectively; there is no heteroscedasticity). They provide visual evidence that the model properly explains the variability of SOC stocks, individually at each time point and jointly across all time points. The validation metrics confirm this observation quantitatively in all cases:

- The (absolute) **Bias** values are well below the margins of error of the validation data means (half-widths of the 90% confidence intervals).
- The **RMSE** values are well below the standard deviations of the validation data.
- the **R^2 scores** are well above zero.

With these observations, the SOC stock model is successfully validated for the project context.

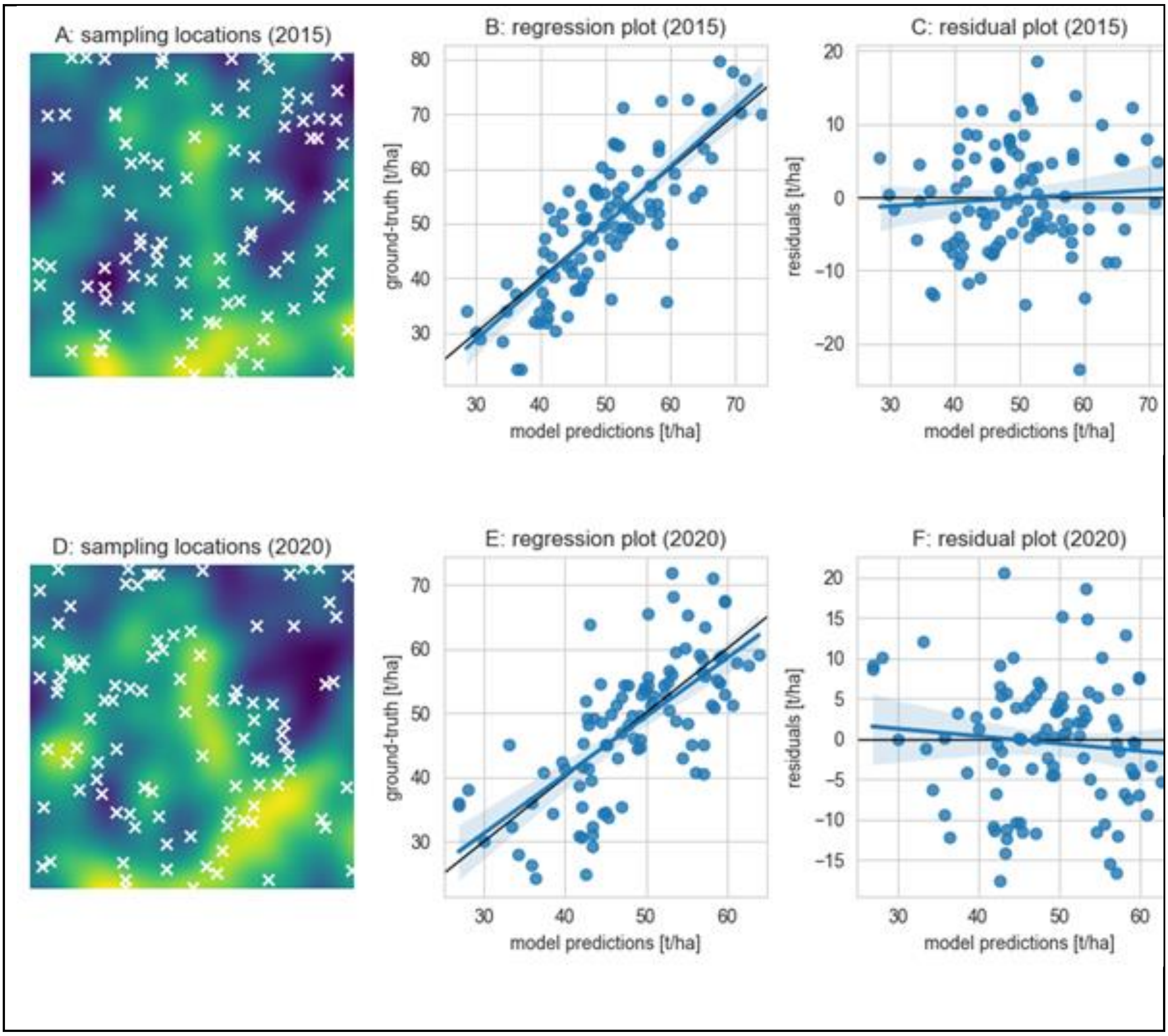


Figure A.2. Sampling locations of the validation data overlaid on the SOC stock ground-truth for 2015 and 2020 (**A, D**), along with regression plots (**B, E**) and residual plots (**C, F**) created separately for the two years. The regression plots and residual plots contain linear trend lines fitted to the respective data (blue lines), along with the corresponding 95% confidence intervals (shaded blue areas) and 1:1 lines for comparison (thin black lines).

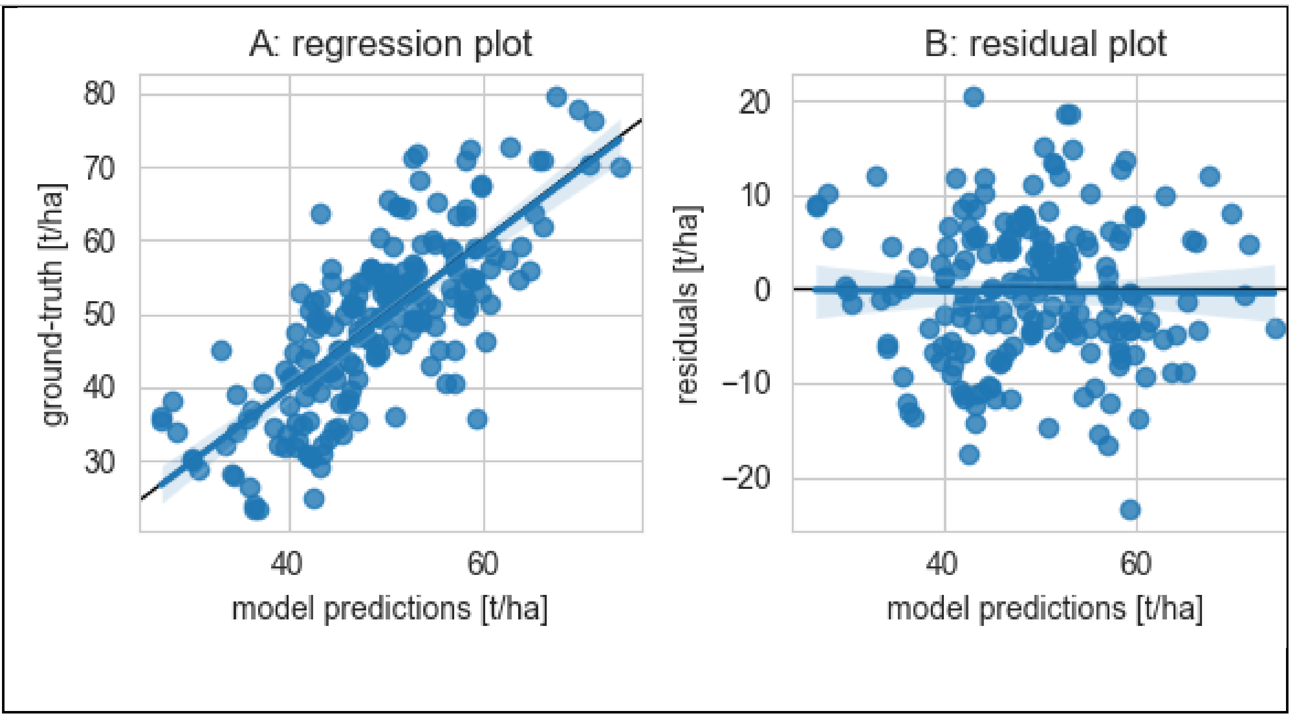


Figure A.3. Regression plot (**A**) and residual plot (**B**) across all 200 SOC stock data points taken in 2015 and 2020, along with linear trend lines (blue lines), 95% confidence intervals (shaded blue areas), and 1:1 lines for comparison (thin black lines).

Table A.4. Model validation metrics on SOC stock data

Year	**2015**	**2020**	**2015 + 2020**
Number of validation data points	100	100	200
Mean model prediction	49.82 t/ha	48.42 t/ha	49.12 t/ha
Validation data mean	49.69 t/ha	48.01 t/ha	48.85 t/ha
Standard error of the validation data mean	1.26 t/ha	1.08 t/ha	0.83 t/ha
Half-width of the 90% confidence interval around the validation data mean	±2.09 t/ha	±1.80 t/ha	±1.37 t/ha
Validation data standard deviation	12.57 t/ha	10.85 t/ha	11.74 t/ha
Bias	-0.13 t/ha	-0.40 t/ha	-0.27 t/ha
RMSE	7.40 t/ha	7.76 t/ha	7.59 t/ha
R^2 score	0.65	0.48	0.58

A.3 Change Prediction and Uncertainty Estimation Exemplified

The goal of this section is to showcase exemplarily how to obtain model-based estimates for the mean SOC stock change and its uncertainty in compliance with Section 6.2 of these guidelines, using the following:

- **DSMs** produced by the SOC stock model during the project period
- **Soil samples** taken with simple random sampling before the project period

In this scenario, the main challenge is to temporally extrapolate the residuals computed from the soil samples into the project period and take them into account for statistically sound change prediction and uncertainty estimation. A geostatistical approach is not required due to the use of probability samples. Instead, two assumptions are made:

1. The residuals are statistically independent of the model predictions.
2. The residuals follow a temporal pattern that can be forecast into the project period.

The first assumption is supported by the residual plots: There is no apparent relationship between the residuals and the model predictions and no indication of heteroscedasticity. (See "Annex B: Requirements for Cross-Validation" and Section 5.3.4, "How to Select a Model," in these guidelines.) The second assumption is much harder to verify without taking additional samples during the project period and will always be speculative until true-up. By assuming a model for the temporal behaviour of the residuals, the approach becomes model-based. There is a plethora of time series models (e.g., AR, ARIMA, etc.) that can be employed for this purpose (Hamilton 1994). The approach yields accurate estimates and uncertainties only if the model adequately represents the true temporal dynamics of the residuals.

For simplicity, this example assumes that the residuals follow a linear trend over time.[10] For this purpose, a linear regression model can be fitted to the validation data residuals (as a function of time).[11] The linear regression model is then used to forecast the residual mean into the project period. During the project period, the mean SOC stock in any given year is estimated as the sum of the mean SOC model prediction (mean of the DSM for that year) and the residual mean forecast from the linear regression model for that year. The mean SOC stock change from project start to any given year within the project period is estimated by subtracting the two SOC stock means obtained in the previous step. Uncertainties of the mean SOC stock predictions and the mean SOC stock change predictions are determined from the uncertainty of the residual mean forecast.

The approach outlined above is visualised in this Figure A.5 for the simulated example from the previous section. All required equations to obtain SOC stock change estimates with uncertainties (as in Figure A.5.D) are listed in Table A.6.

10 A linear trend in the residuals captures models that steadily improve over time (e.g., because more data becomes available to recalibrate a model) as well as models that steadily deteriorate (e.g., because a model was calibrated on historic data and is applied in

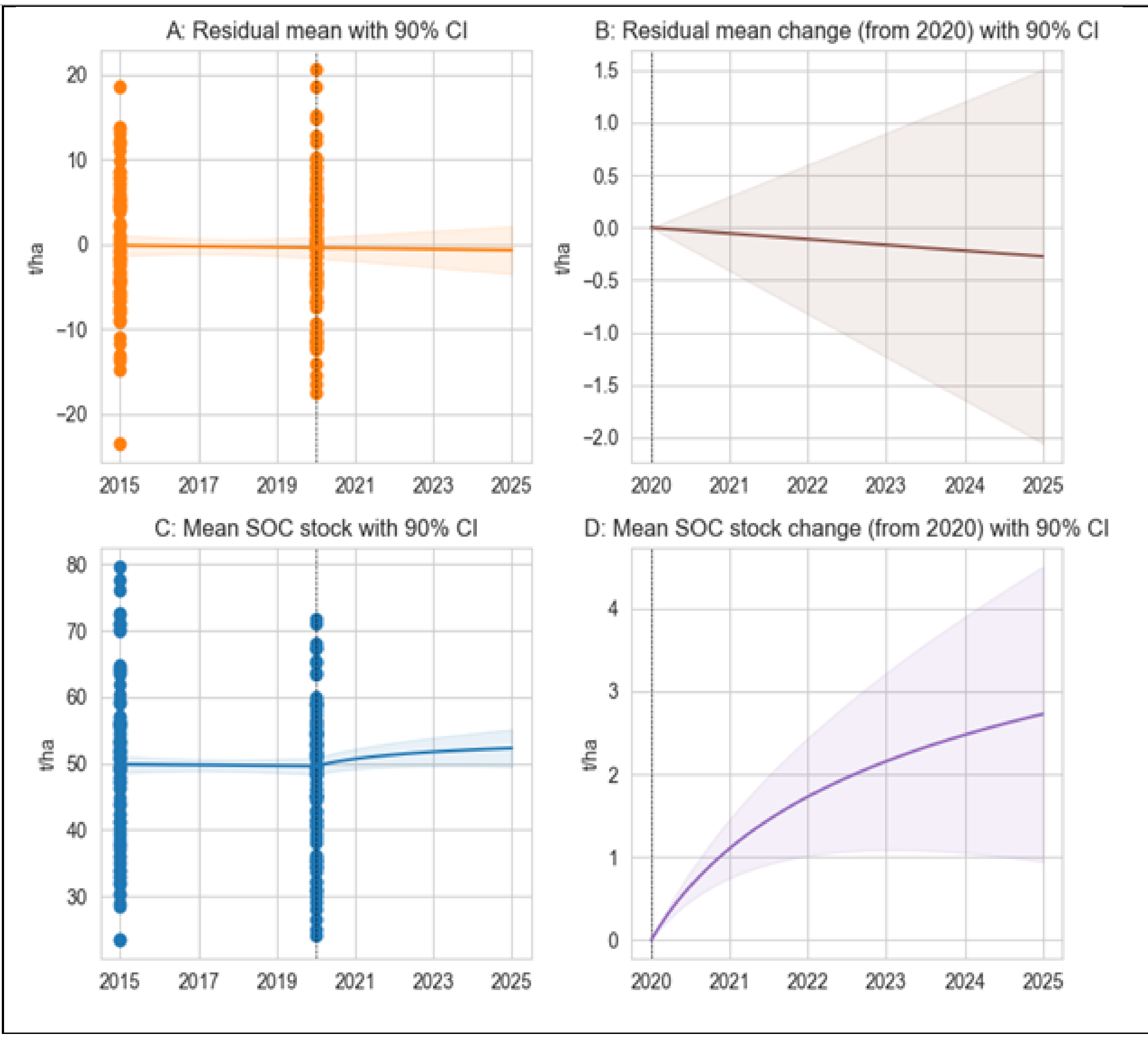


Figure A.5. Visualisation of a model-based SOC stock change quantification and uncertainty estimation approach based on SOC stock predictions, assuming a linear temporal trend in the residuals

The dotted vertical lines in the plots indicate the project start, solid lines indicate mean predictions, and shaded areas indicate 90% confidence intervals (CIs) for the respective means. **A:** SOC stock validation data residuals shown as a function of time, along with a linear regression fit and a forecast of the residual mean into the project period. **B:** The linear regression model allows estimating the change of the residual mean from project start to any year within the project period. **C:** The mean

conditions that are increasingly different from the historic calibration data). It also captures models with a constant performance over time (a zero trend), without *assuming* that the model performance is constant.

11 A linear regression model can be fitted only if validation data from at least *two points in time* is available. If validation data from only *one point in time is* available, e.g., soil samples taken at project start, there is no information about the temporal behaviour of the residuals. As a workaround, if there is reason to believe that SOC was in an equilibrium state for, e.g., five years before the project start, it is permissible to assume that the soil samples taken at project start are also representative for the SOC distribution five years before. In this case, the validation data residuals should be split randomly (50:50), and half of the residuals should be treated as if the samples were taken five years before when fitting the linear regression model.

SOC stocks in any given year (before or after the project start) are estimated by adding the residual mean prediction of the linear regression model to the mean SOC stock prediction obtained from the SOC stock model. **D:** Resulting estimates for the mean SOC stock change from project start to any year within the project period.

Table A.6: Equations for SOC stock change quantification with a linear residual model

Inputs		**Equation Number**
Validation data residuals of the SOC stock model	e_n	26
Validation data sampling times	t_n	27
Validation data residual mean	$\bar{\mathrm{e}} := \frac{1}{N}\sum_n \mathrm{e}_n$	28
Validation data mean sampling time	$\bar{t} := \frac{1}{N}\sum_n t_n$	29
SOC stock model mean prediction in the project area at time *t* (obtained from the DSM)	$\mu_{\mathrm{pred}}(t)$	30
Linear regression parameters		
Slope	$\widehat{\beta_1} := \frac{\sum_n (t_n - \bar{t})(\mathrm{e}_n - \bar{\mathrm{e}})}{\sum_n (t_n - \bar{t})^2}$	31
Intercept	$\widehat{\beta_0} := \bar{\mathrm{e}} - \widehat{\beta_1} \cdot \bar{t}$	32
Regression error variance	$\widehat{\sigma^2}_{\mathrm{resid}} := \frac{1}{N-2}\sum_n \left(\mathrm{e}_n - \left(\widehat{\beta_0} + \widehat{\beta_1} \cdot t_n\right)\right)^2$	33
SOC stock change estimator		
SOC stock mean change estimate from time t to time t’	$\widehat{\mu}_\Delta(t,t') := \mu_{\mathrm{pred}}(t') - \mu_{\mathrm{pred}}(t) + (t'-t)\widehat{\beta_1}$	34
Uncertainty		

Margin of error of the SOC stock mean change estimate	$T_{\alpha/2,N-2}\sqrt{\frac{(t'-t)^2}{\sum_n (t_n-\bar{t})^2}\widehat{\sigma^2}_{\mathrm{resid}}}$	35

Annex B: Requirements for Cross-Validation

Cross-validation (Arlot and Celisse, 2010) may be used during model calibration and for validation of the final model. Cross-validation is a statistical technique to deal with the problem of data scarcity. During cross-validation, a dataset is randomly split into multiple non-overlapping data subsets ("folds"). The folds are used in an iterative process of model calibration and validation; in iteration, a single fold is used for model validation, while the other folds are used for model calibration. The overall model performance is then estimated by averaging the performance metrics obtained in each iteration. The following general requirements apply for cross-validation:

- A minimum of five cross-validation folds shall be used (10 recommended) to reduce the cross-validation uncertainty.
- The minimum number of data points per fold must be 20. In the special case of leave-one-out cross-validation, each fold may contain a single data point only.
- Overall, a minimum of 100 data points shall be used for model calibration when opting for a cross-validation approach.

The procedure that splits the data into different folds shall ensure that there is no data leakage from the training folds to the validation fold. During cross-validation, data leakage can happen in two ways:

1. The exact same data point occurs in multiple cross-validation folds.
2. A cross-validation fold contains a data point that is geographically very close to a data point that lies within a different fold.

The first type of data leakage ca be avoided by following the standard cross-validation protocol. The second type of data leakage shall be addressed explicitly by enforcing a minimum geographical distance between the data points across different folds. The model validation report must demonstrate that the smallest geographical distance between any two data points from two different folds is greater than 30 metres.[12]

For this purpose, the report shall contain a table with the smallest distances (in metres) between any two data points across every pair of folds, as in Example 3 below. In the special case of leave-one-out cross-validation, where each fold contains a single data point only, the report shall clearly state the smallest distance between any two validation points; there is no need to include the table with all pairwise distances. In

12 This threshold was chosen to avoid data leakage due to imprecision of the GPS devices used to record the geographical locations. For example, two points may be taken at different depth intervals at the same location and stored along with two separate GPS devices. readings. Due to GPS device imprecision, the stored locations may be several metres apart. The threshold was explicitly not chosen to avoid *spatial autocorrelation* and achieve *spatial independence* under a specific *geostatistical model*.

any case, the report shall indicate in which coordinate reference system the distances have been computed.

Example 3: Validation of distances for five-fold cross-validation

Table E3.1. Smallest pairwise point distances between cross-validation folds (EPSG:6933)

Fold	**1**	**2**	**3**	**4**	**5**
1	0 m	-	-	-	
2	242 m	0 m	-	-	-
3	301 m	187 m	0 m	-	-
4	193 m	115 m	221 m	0 m	-
5	267 m	152 m	199 m	239 m	0 m

Annex C: Table With Critical Values From the T-Distribution

n_p	t_{np}	n_p	t_{np}	n_p	t_{np}	n_p	t_{np}	n_p	t_{np}	n_p	t_{np}	n_p	t_{np}
		31	1.6973	61	1.6706	91	1.6620	121	1.6577	151	1.6551	181	1.6534
		32	1.6955	62	1.6702	92	1.6618	122	1.6575	152	1.6550	182	1.6533
3	2.9200	33	1.6939	63	1.6698	93	1.6616	123	1.6574	153	1.6549	183	1.6533
4	2.3534	34	1.6924	64	1.6694	94	1.6614	124	1.6573	154	1.6549	184	1.6532
5	2.1319	35	1.6909	65	1.6690	95	1.6612	125	1.6572	155	1.6548	185	1.6532
6	2.0150	36	1.6896	66	1.6686	96	1.6610	126	1.6571	156	1.6547	186	1.6531
7	1.9432	37	1.6883	67	1.6683	97	1.6609	127	1.6570	157	1.6547	187	1.6531
8	1.8946	38	1.6871	68	1.6679	98	1.6607	128	1.6570	158	1.6546	188	1.6531
9	1.8595	39	1.6859	69	1.6676	99	1.6606	129	1.6568	159	1.6546	189	1.6530
10	1.8331	40	1.6849	70	1.6673	100	1.6604	130	1.6568	160	1.6545	190	1.6529
11	1.8124	41	1.6839	71	1.6669	101	1.6602	131	1.6567	161	1.6544	191	1.6529
12	1.7959	42	1.6829	72	1.6666	102	1.6601	132	1.6566	162	1.6544	192	1.6529
13	1.7823	43	1.6820	73	1.6663	103	1.6599	133	1.6565	163	1.6543	193	1.6528
14	1.7709	44	1.6811	74	1.6660	104	1.6598	134	1.6564	164	1.6543	194	1.6528
15	1.7613	45	1.6802	75	1.6657	105	1.6596	135	1.6563	165	1.6542	195	1.6528
16	1.7530	46	1.6794	76	1.6654	106	1.6595	136	1.6562	166	1.6542	196	1.6527
17	1.7459	47	1.6787	77	1.6652	107	1.6593	137	1.6561	167	1.6541	197	1.6527
18	1.7396	48	1.6779	78	1.6649	108	1.6592	138	1.6561	168	1.6540	198	1.6526
19	1.7341	49	1.6772	79	1.6646	109	1.6591	139	1.6560	169	1.6540	199	1.6526
20	1.7291	50	1.6766	80	1.6644	110	1.6589	140	1.6559	170	1.6539	≥200	1.6525
21	1.7247	51	1.6759	81	1.6641	111	1.6588	141	1.6558	171	1.6539		
22	1.7207	52	1.6753	82	1.6639	112	1.6587	142	1.6557	172	1.6538		
23	1.7172	53	1.6747	83	1.6636	113	1.6586	143	1.6557	173	1.6537		
24	1.7139	54	1.6741	84	1.6634	114	1.6585	144	1.6556	174	1.6537		
25	1.7109	55	1.6736	85	1.6632	115	1.6583	145	1.6555	175	1.6537		
26	1.7081	56	1.6730	86	1.6630	116	1.6582	146	1.6554	176	1.6536		
27	1.7056	57	1.6725	87	1.6628	117	1.6581	147	1.6554	177	1.6536		
28	1.7033	58	1.6720	88	1.6626	118	1.6580	148	1.6553	178	1.6535		
29	1.7011	59	1.6715	89	1.6623	119	1.6579	149	1.6552	179	1.6535		
30	1.6991	60	1.6711	90	1.6622	120	1.6578	150	1.6551	180	1.6534		

Figure A.7. Table 6 from the SOC FM (p. 27): "t-values (t_{np}) applicable in Equation 8. Select appropriate t_{np} value depending on the number of samples (np) measured for parameter p."

Annex D: Bibliography

Arlot, S., Celisse, A., 2010. A survey of cross-validation procedures for model selection. Stat. Surv. 4. https://doi.org/10.1214/09-SS054

Aynekulu, E., Vagen, T.G., Shephard, K., Winowiecki, L., 2011. A Protocol for Modelling, Measurement and Monitoring Soil Carbon Stocks in Agricultural Landscapes.

Bradford, M.A., Eash, L., Polussa, A., Jevon, F.V., Kuebbing, S.E., Hammac, W.A., Rosenzweig, S., Oldfield, E.E., 2023. Testing the feasibility of quantifying change in agricultural soil carbon stocks through empirical sampling. Geoderma 440, 116719. https://doi.org/10.1016/j.geoderma.2023.116719

Brus, D.J., 2000. Using regression models in design-based estimation of spatial means of soil properties. Eur. J. Soil Sci. 51, 159–172. https://doi.org/10.1046/j.1365-2389.2000.00277.x

Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7, 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Dvorakova, K., Heiden, U., Pepers, K., Staats, G., van Os, G., van Wesemael, B., 2023. Improving soil organic carbon predictions from a Sentinel–2 soil composite by assessing surface conditions and uncertainties. Geoderma 429, 116128. https://doi.org/10.1016/j.geoderma.2022.116128

FAO, 2020. A protocol for measurement, monitoring, reporting and verification of soil organic carbon in agricultural landscapes. FAO ;

Garsia, A., Moinet, A., Vazquez, C., Creamer, R.E., Moinet, G.Y.K., 2023. The challenge of selecting an appropriate soil organic carbon simulation model: A comprehensive global review and validation assessment. Glob. Change Biol. 29, 5760–5774. https://doi.org/10.1111/gcb.16896

GS4GG, TREES Consulting, 2020. Gold Standard Soil Organic Carbon Framework Methodology.

Hamilton, J.D., 1994. Time series analysis. Princeton University Press, Princeton, N.J.

Hengl, T., Miller, M.A.E., Križan, J., Shepherd, K.D., Sila, A., Kilibarda, M., Antonijević, O., Glušica, L., Dobermann, A., Haefele, S.M., McGrath, S.P., Acquah, G.E., Collinson, J., Parente, L., Sheykhmousa, M., Saito, K., Johnson, J.-M., Chamberlin, J., Silatsa, F.B.T., Yemefack, M., Wendt, J., MacMillan, R.A., Wheeler, I., Crouch, J., 2021. African soil properties and nutrients mapped at 30 m spatial resolution using two-scale ensemble machine learning. Sci. Rep. 11, 6130. https://doi.org/10.1038/s41598-021-85639-y

Henry, B.K., Dalal, R., Harrison, M., Keating, B., 2022. Creating frameworks to foster soil carbon sequestration. https://doi.org/10.19103/AS.2022.0106.25

Heuvelink, G.B.M., Angelini, M.E., Poggio, L., Bai, Z., Batjes, N.H., 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. Eur. J. Soil Sci. 72, 1607–1623. https://doi.org/10.1111/ejss.12998

ISMC, 2023. Model Portal - Soil Modeling [WWW Document]. URL https://www.soil-modeling.org/resources-links/model-portal (accessed 12.20.23).

Koenker, R., 2005. Quantile regression. Cambridge University Press, Cambridge.

Lavallee, J.M., McConkey, B.G., Kyker-Snowman, E., Dietze, M.C., Harrison, M.T., Liu, K., Smith, W., Grant, B., Karunaratne, S., Marklein, A., 2024. Identifying challenges and advancing guidance for using process-based models in soil emission reduction and removal projects. Environmental Defense Fund, New York.

Le Noë, J., Manzoni, S., Abramoff, R., Bölscher, T., Bruni, E., Cardinael, R., Ciais, P., Chenu, C., Clivot, H., Derrien, D., Ferchaud, F., Garnier, P., Goll, D., Lashermes, G., Martin, M., Rasse, D., Rees, F., Sainte-Marie, J., Salmon, E., Schiedung, M., Schimel, J., Wieder, W., Abiven, S., Barré, P., Cécillon, L., Guenet, B., 2023. Soil organic carbon models need independent time-series validation for reliable prediction. Commun. Earth Environ. 4, 158. https://doi.org/10.1038/s43247-023-00830-5

Meinshausen, N., 2006. Quantile Regression Forests. J. Mach. Learn. Res. 7.

Poeplau, C., Vos, C., Don, A., 2017. Soil organic carbon stocks are systematically overestimated by misuse of the parameters bulk density and rock fragment content. SOIL 3, 61–66. https://doi.org/10.5194/soil-3-61-2017

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

Reddy, S., Panichelli, L., Waterworth, R.M., Federici, S., Green, C., 2019. 2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories Volume 4 AFOLU Chapter 3 Consistent Representation of Land.

Romano, Y., Patterson, E., Candès, E.J., 2019. Conformalized Quantile Regression.

Rykiel, E.J., 1996. Testing ecological models: the meaning of validation. Ecol. Model. 90, 229–244. https://doi.org/10.1016/0304-3800(95)00152-2

Särndal, C.-E., Swensson, B., Wretman, J., 2003. Model Assisted Survey Sampling. Springer Science & Business Media.

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

Wendt, J. W., & Hauser, S. (2013). An equivalent soil mass procedure for monitoring soil organic carbon in multiple soil layers. European Journal of Soil Science, 64(1), 58-65.

Zepp, S., Heiden, U., Bachmann, M., Wiesmeier, M., Steininger, M., van Wesemael, B., 2021. Estimation of Soil Organic Carbon Contents in Croplands of Bavaria from SCMaP Soil Reflectance Composites. Remote Sens. 13, 3141. https://doi.org/10.3390/rs13163141

DOCUMENT HISTORY

Version	Date	Description
1.0	08/09/2025	First version released
