Verified Carbon Standard

VCS Module

VMD0053

MODEL CALIBRATION, VALIDATION AND UNCERTAINTY GUIDANCE FOR BIOGEOCHEMICAL MODELING FOR AGRICULTURAL LAND MANAGEMENT PROJECTS

Version 2.1

25 March 2025

Sectoral Scope 14

Version 1.0 of this module was developed by TerraCarbon LLC and Indigo Ag. Version 1.0 was approved on 19 October 2020.

Version 2.0 of this module was developed by Verra and TerraCarbon LLC in consultation with several external experts.

Version 2.1 of this module was developed by Verra.

CONTENTS

1 SUMMARY DESCRIPTION 4

2 SOURCES 7

3 DEFINITIONS 7

4 APPLICABILITY CONDITIONS 8

5 PROCEDURES 9

5.1 Model Calibration 10

5.2 Model Validation 14

5.3 Substitution for Missing Crop Types 32

6 DATA AND PARAMETERS 33

6.1 Data and Parameters Available at Validation 33

6.2 Data and Parameters Monitored 36

7 REFERENCES 37

APPENDIX 1: ASSESSMENT BY INDEPENDENT MODELING EXPERT (IME) 39

A1.1 Steps for Model Assessment by an IME 39

A 1.2 Minimum Qualifications of IMEs 40

1 SUMMARY DESCRIPTION

This module provides procedures for calibration, validation and verification of empirical or process-based models used to estimate carbon stock changes and GHG emission fluxes. It is used in conjunction with VCS Program methodologies that include the option to use biogeochemical models for GHG flux quantification of agricultural land management (ALM) activities. This module must be used for all GHGs and carbon pools for which biogeochemical process-based models are employed to quantify baseline and the project emissions. Figure 1 shows the workflow for the use of models within the VCS project cycle.

A Model Validation Report (MVR) must be generated by the project proponent following the guidance in this module. The MVR is designed to support independent expert review, by an independent modeling expert (IME), of a model proposed for use in an ALM project. The MVR is also designed to support independent verification that a model is valid and used appropriately to quantify GHG benefits in a specific project according to the requirements specified in the methodology applied and this document. This ensures that the model is appropriately tested for model performance with defined bounds for model prediction error. As shown in Figure 1, the MVR must be submitted for project verification.

The requirements and guidance in this document fall into two main categories:

- **Category 1:** Standardized best practices for use of peer-reviewed observed experimental data to test a model and determine model prediction error; and
- **Category 2**: Standardized demonstration of acceptable fit and a lack of bias when a model is used to estimate soil organic carbon (SOC) stock change and, where applicable, flux change of N_2O and CH_4.

Requirements under Category 1 address the importance of using high-quality observed experimental data of soil-based GHG emissions reductions or carbon dioxide removals (i.e.., reductions in N_2O or CH_4 flux, soil carbon sequestration) from controlled research trials or approved data sources as the basis of evaluating model performance. Changes in ALM practices have diverse impacts on soil-based emissions. Soil-based emissions are also highly variable. There has been rapid growth in new studies and experimental methods to capture this variance, increase precision and reduce uncertainty. Requirements described in this module are intended to ensure that appropriate and consistent methods are followed to locate, aggregate and use observed data for model improvement and testing.

Requirements under Category 2 provide guidance for model calibration, validation and the determination of model prediction error in the context of measurement uncertainties. These are highly technical processes that vary widely across scientific research areas. The MVR aims to ensure that model validation is specific to the model proposed for use in the project, is appropriate for the land management system and biophysical conditions occurring in the project and requirements related to the assessment of model bias and fit have been met. Model validation must be documented in an MVR which must show that the same model version and parameter sets are used, and that all project domain and practice category/crop

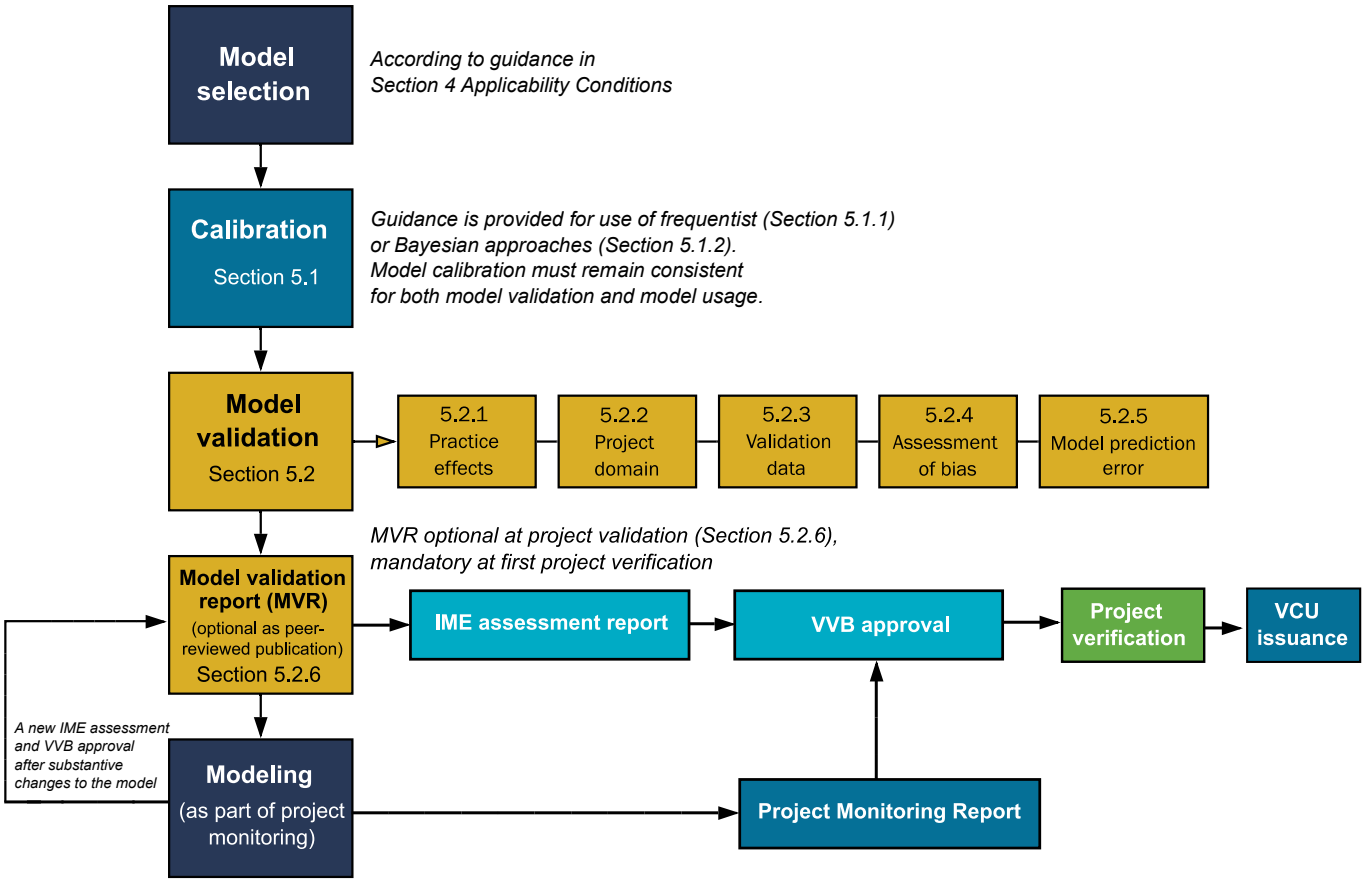
functional group (PC/CFG) combinations have met minimum requirements for model validation. MVRs must be independently assessed by an IME, or accepted for publication in one of the peer-reviewed publications listed in Table 3 and reviewed by an IME as per the procedures outlined in Section 5.2.6. MVRs will be public documents.

Where a project area remains constant or is only expanded to include new fields that already fit within the validated project domain, the existing MVR should be used for each subsequent monitoring report. Where the project is expanded to new practice categories, crop functional groups or emissions sources, or the model is changed in a way that substantively affects model runs and the estimated emission reductions and removals (ERRs), the MVR must be:

1) Revised, re-submitted and reviewed by an IME, or
2) Submitted and accepted for publication as a new journal article in one of the peer-reviewed journals listed in Table 3 and reviewed by an IME.

In both cases, the IME must assess the MVR or peer-reviewed publication to confirm its appropriateness for the project domain (see Figure 1).

Figure 1: Steps for using models for GHG quantification in ALM projects



2 SOURCES

This module is based on the following documents:

- *VMD0053 Model Calibration, Validation, and Uncertainty Guidance for the Methodology for Improved Agricultural Land Management, v2.0*
- *Requirements and Guidance for Model Calibration, Validation, Uncertainty, and Verification for Soil Enrichment Projects, Version 1.0a – Climate Action Reserve*

3 DEFINITIONS

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

Calibration
Any process involving the adjustment of parameters and constants within a model so that the model more accurately simulates measured values.

Climate zone
Geographic zone as defined in the *2019 Refinements to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories,* Volume 4, Chapter 3, Figure 3A.5.1[1] on page 3.47.

Crop functional group (CFG)
Broad category of crop species with similar characteristics (e.g., grasses, legumes, non-legume broadleaf species, rice varieties)

Goodness of fit
A characterization of the discrepancy between measured and modeled values

Independent modeling expert (IME)
An individual or organization that has demonstrated competency in quantifying GHG fluxes, in particular SOC stock changes, in agricultural land management using biogeochemical models and is independent from the project proponent

Model-driving input data
Data needed to execute a model run, such as meteorological time series data, or rates of fertilizer application, crop identities or seed values for random number generation

[1] Consider the first Corrigenda to the *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories* replacing the original climate zone map, available at: https://www.ipcc-nggip.iges.or.jp/public/2019rf/corrigenda1.html

Model prediction error
The uncertainty in a model prediction as determined from comparison to direct measurements. Measurements used to determine model prediction error must be the same as those used to validate the model.

Model validation
The process of evaluating model performance relative to measured values, with a validated model having demonstrated satisfactory performance in terms of goodness of fit and characterization of model prediction error. Model validation must use datasets independent of those used in model calibration, unless a statistical approach like k-folding is applied (e.g., in a data-limited situation).

Model Validation Report (MVR)
A document which must demonstrate that model calibration and validation are specific to the model being proposed for use in the project and appropriate for the land management system and biophysical conditions occurring in the project, and that requirements related to the assessment of model bias and fit have been met. Data sources for calibration and validation datasets must be specified in the MVR, which is prepared by the project proponent.

Model version
A uniquely traceable record of all files needed to reproduce a given model output from its calibrated parameter set and model-driving input data. These (collectively the "model files") should include source code, internal parameters that are not adjusted during calibration, default values for parameters or input data and any other information that may change model behavior. A model version must change any time there is a change in any of the model files. For a given parameter set and set of model-driving inputs, any copy of the model reporting the same model version must always produce the same output.

Parameter set
The set of mathematical values and constants contained in a model that characterizes the biophysical and biogeochemical system being represented.

Pooled measurement uncertainty (PMU)
An estimate of the typical uncertainty associated with experimental measurements of the emissions change resulting from a given practice change. It is computed from the observed variation between replicate measurements.

4 APPLICABILITY CONDITIONS

This module applies where empirical or process-based models used to estimate SOC stock changes/GHG emissions meet all of the following conditions. Models must be:

1) Publicly available, though not necessarily free of charge, from a reputable and recognized source (e.g., the model developer's website, IPCC or government agency). Sufficient conceptual documentation of inputs, outputs and information on how the model functionally represents carbon stock and GHG dynamics must be accessible to the public. Providing the source code or an API for independent replication of calculations is not necessary;
2) Appropriately reviewed, tested and shown in peer-reviewed scientific studies to successfully simulate changes in carbon stock and trace gas emissions resulting from changes in ALM introduced by the project activity;
3) Able to support repetition of the project model simulations. This includes clear versioning of the model used in the project and stable software support of that version, as well as fully reported sources and values for all parameters used with the model version and the project activity. The same model version must be used in the baseline and project scenarios. For stochastic models, the seeding sequence to the random number generator must be provided so that model runs may be reproduced. Where multiple sets of parameter values are used in the project, full reporting must include clear identification of the sources of varying parameter sets as well as how they were applied to estimate stock change/emissions in the project. Acceptable sources include peer-reviewed literature and statements from appropriate expert groups (i.e., that demonstrate evidence of expertise with the model via authorship of peer-reviewed model publications or authorship of reports for entities supporting climate-smart agriculture, such as FAO or a comparable organization). Project proponents must describe the datasets and statistical processes used to set parameter values (i.e., the parameterization or calibration procedure); and
4) Validated per datasets and procedures detailed in Section 5.2. Model prediction error is calculated using datasets described in Section 5.2.5, using the same parameter sets applied to estimate stock changes/emissions in the project. Note that this means every parameter set must be validated separately.

5 PROCEDURES

The model calibration and validation procedures described in this section must be adhered to when quantifying baseline and project emissions to determine soil carbon stock changes or GHG emission fluxes.

The requirements for developing the baseline and project scenarios of the applied methodology must be adhered to.[2]

[2] Where a discrepancy between the module's and the methodology's requirements occurs, the project must justify which approach is more accurate and conservative.

5.1 Model Calibration

Model calibration is a variable and model-specific set of processes. Some examples include:

- *Statistical procedures to optimize rates of mass flow and the simulation of internal model pools (e.g., optimizing the allocation of daily net primary production to root growth to more accurately simulate observed root growth for a given crop);*
- *Adjusting model parameters with directly measured values (e.g., setting the simulated fraction of plant residue left on the soil surface after harvest using an average of observed values); and*
- *"Tuning" a set of model parameters that it may not be possible to measure directly, using overall model performance and an understanding of model sensitivities (e.g., adjusting a constant downregulating of the rate of soil biological processes under moisture-limited conditions using measures of soil respiration).*

Deterministic models, where the same inputs always result in the same outputs, may have different calibration processes than stochastic models, which include random variability. Mechanistic models, which are based on mathematical representations of mechanisms within the modeled system, are more generalizable with fewer data than empirical models. Empirical models are based on statistical synthesis of observations and should not be extended outside of where observations are available.

Data used for model calibration must be independent from data used for model validation (i.e., using a separate process and separate datasets). Further, for either process, the quality of measured datasets (i.e., rigor of the experimental design, accuracy of observations, applicability to the system that a model is being calibrated or validated to simulate) will determine the quality of the model output. Datasets for calibration and validation may be either kept completely isolated from each other or drawn from a single pool using a statistical process that guarantees independence, such as k-folding.

Calibration and validation data must be demonstrably independent. This requirement is met where datasets used for calibration and validation do not overlap in experimental research locations and are not taken from the same experimental study. Where calibration and validation datasets for SOC change or trace gas flux do overlap in either experimental study or research location, independence between the datasets used for calibration and validation should be demonstrated at the PC/CFG combination level (Section 5.1.2). For example, where root measurements and N_2O flux measurements from a subset of treatments in a tilled soybean/corn rotation experiment are used for model calibration, the N_2O flux measurements from the remaining treatments in the same study must not be used as validation data for either the corn or the soy CFG and tillage practice effect combinations. However, if at the same research facility N_2O flux was measured in a demonstrably separate corn/soy rotation experiment (separate in space or time, with separate experimental design or intention), those data are permissible for inclusion in model validation. In some cases, depending on the model, it may be defensible to use cultivar-specific measurements of crop growth to calibrate modeled crop growth, while using SOC change or trace gas flux change measurements from the same

study to validate model performance. Such cases must be clearly explained and presented for review in the MVR. Note that SOC stock changes in calibration and validation datasets need not be calculated on an equivalent soil mass (ESM) basis.

This module does not prescribe a single model calibration procedure. However, the selected calibration procedure must be reported to ensure model parameters and parameter sets were generated appropriately and meet the following requirements:

1) The parameter sets used when validating the model are the same as those used when the model is applied to simulate baselines and project practices; and
2) The data used for model calibration and validation are separate.

In this context, "parameter sets" refer to all values internal to a model that determine how input data drive model performance and behavior, and that are changed using processes independent of model-driving input datasets. This means model parameters that are not dependent on input datasets when the model is run (e.g., through a Bayesian statistical procedure) must be declared and shown to be set appropriately following the above calibration requirements.

Model parameters should be as generalizable as possible across the project domain, with minimal use of different parameter sets. However, different parameter sets may be used where they are defined at the scale of either IPCC climate zones or nationally defined agricultural land regions, for example Land Resource Regions in the US (Section 5.2.2). Where a project is using nationally defined agricultural land regions, the definition must be approved by a VVB and parameter sets should be declared at a minimum for each individual agricultural land region included in the project. This information must be included in the MVR and assessed and approved by the IME.

The same parameter set should be used to simulate all CFGs and PCs within that defined land area. An exception may be made for crop growth parameters, for example to reflect different maturity groups within a large land region. The use of varying crop growth parameters must be clearly defined in the MVR and presented as parameter sets specific within each land area boundary where the crop is simulated. This ensures appropriate use in model validation and project simulations.

Because biogeochemical models often contain many parameters, different strategies may be employed for calibration. General guidance for frequentist and Bayesian approaches are provided in Sections 5.1.1 and 5.1.2 respectively.

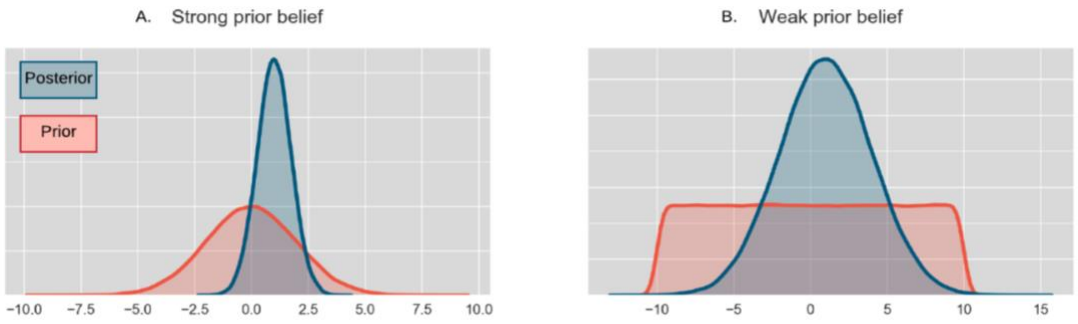
5.1.1 Guidance on Model Calibration using Frequentist Approaches

Wallach et al. (2019) provide helpful guidance on common approaches to frequentist model calibration, including how to decide how many and which parameters to estimate, whether to calibrate in stages and how to avoid over-parameterization (i.e., where the model fits the data well but has poor predictive ability). Examples of model calibration are abundant in the peer-reviewed literature and span a wide range of complexity and automation in their approaches (e.g., Bruun et al., 2003; Liang et al., 2009; Yeluripati et al., 2009).

5.1.2 Guidance on Bayesian Methods for Calibration, Validation and Error

Model calibration may also be completed using Bayesian statistical methods, which apply a probabilistic approach to integrating existing knowledge and observed data (Wikle & Berliner, 2007). Bayesian statistical approaches are an emerging area of development in soil biogeochemical modeling. They typically require implementing Markov Chain Monte Carlo methods for sampling probability distributions. This is often computationally demanding with soil biogeochemical models, which may have dozens to hundreds of parameters or more. Parameter values in these types of models are also sometimes difficult to constrain (i.e., using data or existing knowledge to set limits on the range of values that a parameter may have and defining its probability distribution across that range). Where there is little prior knowledge about a parameter value, "uninformative priors" or "weakly informative priors" are used to represent what is known or believed about the parameter. The resulting posterior distribution, or the distribution that represents the integration of prior knowledge and observed data, may be wide unless the observed data are strongly informative (i.e., have highly accurate and precise values). Figure 2 illustrates a strong prior belief (A) versus a weak prior belief (B).

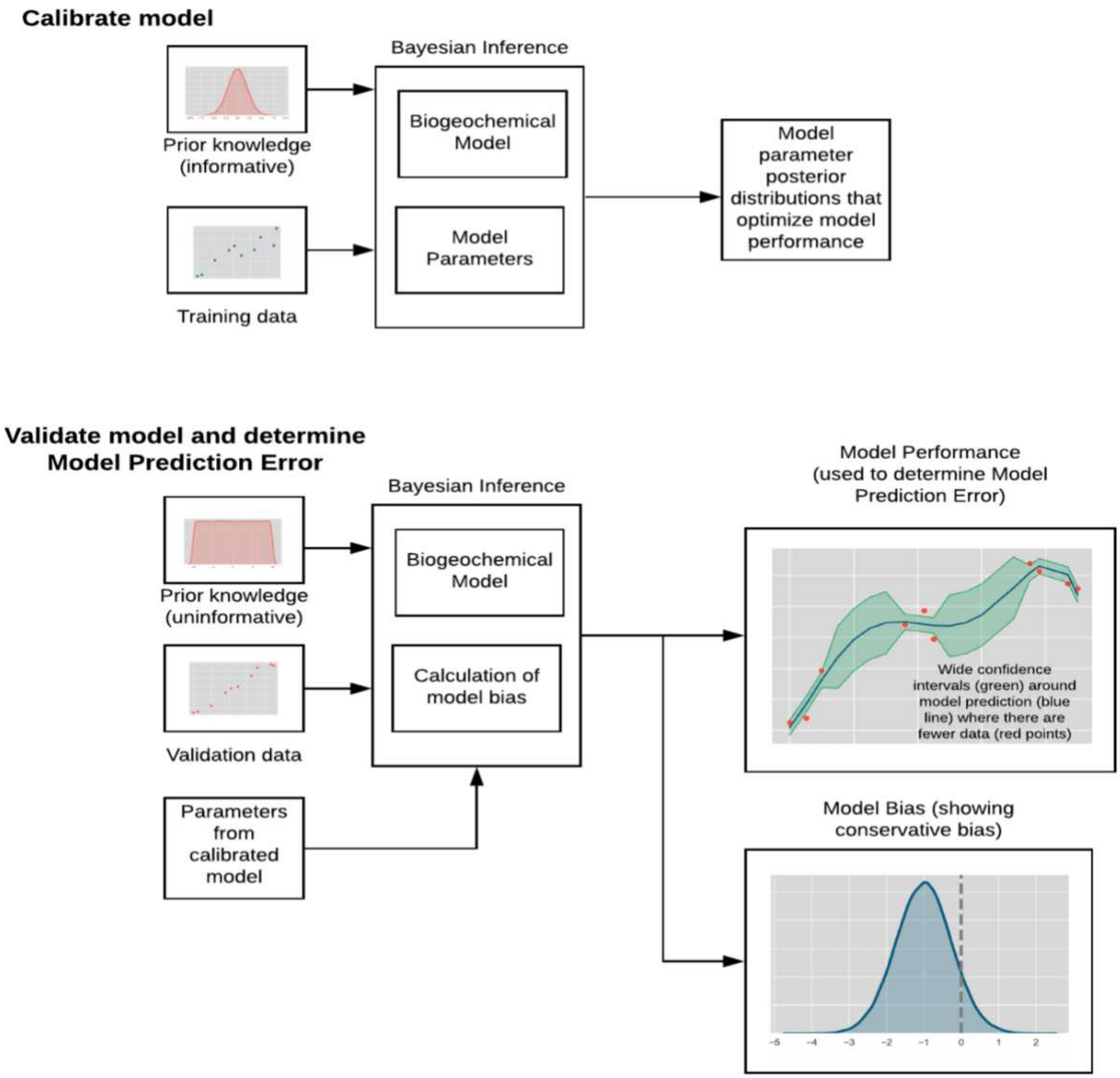
Figure 2: Comparison of prior and posterior distributions when there is A) strong prior belief (e.g., strong and consistent evidence and prior analyses) and B) weak prior belief (e.g., weak or variable evidence or no prior analyses)



Across dozens or hundreds of parameters, Bayesian methods may be complex to implement and require large quantities of data. Despite these challenges, Bayesian methods provide a coherent mathematical framework to integrate diverse sources of information into model parameterization, as evidenced in their central role in the developing field of ecological forecasting (Dietze, 2017), as well as in the Predictive Ecosystem Analyzer Project data-model integration system.[3] A Bayesian approach is encouraged for model validation and model prediction error, as the confidence intervals around model predictions will be directly based on the availability and variance of observed data. Figure 3 presents a conceptual workflow for a Bayesian approach to these analyses.

[3] Available at: www.pecanproject.org

Figure 3: Conceptual framework for Bayesian approach to model calibration and validation



In the example in Figure 3, model calibration is a separate analytical process from validating model performance and determining model prediction error. In a fully integrated analysis, informative posteriors from model calibration may be used as priors in model validation.

Box 1: Summary of requirements described in Section 5.1

The following are required for the MVR:

- Model version, as defined in Section 2
- Description of the model calibration process, including the adjustment of model parameters with directly measured values (e.g., leaf area index or harvest index, or increases in plant productivity due to genetic improvements)
- Documentation of all internal model parameter sets, including proof that parameter sets are defined at a resolution no finer than one climate zone or one nationally defined agricultural land region, depending on which is declared by the project (Section 5.2.2). Where there is justification to claim an allowance for crop growth parameter sets to vary within climate zones/nationally defined agricultural land regions (e.g., varying maturity groups), documentation must be provided for each zone/region where the crop will be simulated. The documentation must specify all crop growth parameter sets used in the zone/region and the rules used to select which parameter set is used for a given simulation. Documentation of calibration for crop growth modeling and validation of the crop growth model is not a requirement, unless the IME deems such documentation necessary given the nature of the project activities.
- Justification for splitting of experimental data between calibration and validation (where applicable), clearly described at the CFG/PC/emission source (ES) combination level

The following are required upon request from the IME:

- Datasets used for model calibration, including but not limited to full citation, experimental locations, specific crops and practices studied, climate zones/nationally defined agricultural land regions, soil textures and clay contents and number of observations

5.2 Model Validation

Model validation is the process of evaluating model performance relative to measured values, with a validated model having demonstrated satisfactory performance in terms of goodness of fit and characterization of model prediction error. This section outlines required procedures for model validation. Note that it is not acceptable to validate a model and then adjust model parameters when using the model to simulate project baselines and practices. All parameter sets must be validated following the guidance in this section. Where the minimum requirements do not result in all parameter sets being validated and additional steps are not taken to validate all parameter sets, unvalidated parameter sets are not approved for use in the project.

5.2.1 Declare Practice Categories Requiring Evaluation

For every practice considered additional within the project, the model must be shown to have an acceptable goodness of fit and unbiased representation of the underlying biogeochemical process governing the effect of that practice. To demonstrate this, each practice must be binned into the practice categories (PCs) shown in Table 1 to demonstrate the domain of practice effects and the categories requiring evaluation. Validating model performance and

uncertainty within a PC is accomplished using any practice effect in the category domain, evaluated using appropriate experimental data meeting the requirements described below. Projects are encouraged to evaluate a range of practice effects in each PC domain.

The project proponent must declare all practice effects requiring evaluation for the project.

Table 1: Practice categories and their associated practice effects requiring biogeochemical performance evaluation

Practice Category Requiring Evaluation	Domain of Practice Effects
Inorganic nitrogen fertilizer application	• Magnitude • Form: inorganic N fertilizers • Timing • Method: surface, subsurface or irrigation-based application
Organic amendments application	• Magnitude • Form: includes but is not limited to biochar, mulch, compost, animal manure • Timing • Method: surface, subsurface or irrigation-based application • Variation in C:N ratio
Water management/irrigation/flooding (e.g., paddy rice)	• Magnitude • Timing • Source of water • Method of irrigation
Soil disturbance and/or residue management	• Soil disturbance including tillage and compaction • Residue management: soil exposure after harvest, physical incorporation of green manure • Seeding practice (e.g., direct seeding of rice)
Land cultivation practices, planting and harvesting (e.g., crop rotations, cover crops)	• Variety of crops grown (including improved varieties) • Increasing crop rooting depth • May include cover crops • May include soil preparations such as changing soil pH through liming
Grazing practices	• Presence/absence of grazing • Stocking density • Forage type or quality • Species of grazers • Mixed or single species herds

	• Loading weight • Grazing time • Rest/recovery periods

5.2.2 Define the Project Domain

For each PC declared in the project description, the model must be evaluated in terms of its fit and bias in estimating emission reductions. Evaluation of each PC begins with defining the project domain in terms of its biophysical attributes. Specifically, the project proponent must declare the unique CFGs, climate zones/nationally defined agricultural land regions and soil attributes associated with each declared PC.

Step 1: Declare Project Crop Functional Groups

CFGs for each PC must be declared. Individual crop types may be grouped into functional groups across crops sharing unique combinations of the following attributes:

1) N fixation (Y/N);
2) Annual/perennial (A/P) (defined in accordance with the NRCS Conservation Compliance categorization of crops[4]);
3) Photosynthetic pathway (C3/C4/CAM);
4) Growth form (tree/shrub/herbaceous – trees and shrubs have woody plant growth, whereas herbaceous species do not grow woody plant material);
5) Crop variety (e.g., rice variety); and/or
6) Flooded/not flooded.

Step 2: Declare Climate Zones or Nationally Defined Agricultural Land Regions

A project may use either climate zones or nationally defined agricultural regions to define its project domain. Where using climate zones, the full list of climate zones encompassed in the project domain must be declared for each PC, following the climate zone definitions given in the *2019 Refinements to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*. Where using nationally defined agricultural land regions (e.g., Land Resource Regions in the United States), the regions must be approved as appropriate for the project by a VVB. The full list of defined land boundaries encompassed in the project domain must then be declared.

Step 3: Declare Project Soils

Soils are to be declared for each practice category in terms of:

1) Soil textural class and

[4] Available at: https://www.nrcs.usda.gov/2014-farm-bill-conservation-compliance-crop-list

2) Associated range in possible clay content of each class, according to the standard FAO/USDA soil texture triangle.[5]

Soil texture classes include sand, loamy sand, sandy loam, loam, silt loam, silt, sandy clay loam, clay loam, silty clay loam, sandy clay, silty clay and clay.

Box 2: Summary of requirements described in Sections 5.2.1 and 5.2.2

The following are required for the MVR:

- List of combinations of PCs and CFGs occurring in the project
- List of combinations of PCs, CFGs and ESs validated
- List of climate zones/nationally defined agricultural land regions included in the project domain
- List of soil texture classes and associated clay content in the project domain

The following are required upon request from the IME:

- List of specific crops and practices occurring in the project, and a description of how these were binned into the PCs and CFGs validated

5.2.3 Gather Data to Validate Model Performance and Uncertainty

Requirement 1: Generalized Dataset Attributes

Datasets to validate model performance and uncertainty for each declared PC/CFG/ES combination from Section 5.2.1 and 5.2.2 must include measurements for each modeled quantity, where the modeled quantity is the change in the flux of emissions to the atmosphere for SOC, N_2O and/or CH_4 that results from the adoption of any practice associated with that effect. Datasets may include individual PCs as well as combinations of PCs (e.g., "stacked" practices), provided the PC in question is experimentally varied and measured within the study. Some hypothetical examples of acceptable experimental treatments to evaluate PCs are given in Table 2.

[5] The FAO *World reference base for soil resources 2014* (available at: https://www.fao.org/3/i3794en/I3794en.pdf) uses the same classification scheme for soil texture classes as the USDA. The USDA Soil Texture Calculator (available at: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs142p2_054167) may be used to determine the soil texture class based on percent sand and clay content.

Table 2. Examples of acceptable experimental treatments for use in evaluating practice categories

Experimental Treatment	Practice Category
Comparison of two different application rates of urea	Inorganic nitrogen fertilizer application
Comparison of continuously flooded rice management to alternate wetting and drying (AWD)	Water management/irrigation/flooding
Comparison of conventional tillage using moldboard plow to strip tillage	Soil disturbance and/or residue management
Comparison of conventional seeding to direct seeded rice (DSR) in rice systems	Soil disturbance and/or residue management
Comparison of single-crop rotation to double-crop rotation; comparison of no cover crop to with cover crop	Cropping practices, planting and harvesting (e.g., crop rotations, cover crops)

Datasets to validate model performance and uncertainty must adhere to the following guidelines:

- Measured datasets must be drawn from peer-reviewed and published experimental datasets with measurements of the emissions source(s) of interest (SOC stock change and/or N_2O and CH_4 change, as applicable), ideally using control plots to test the PC. Datasets may also be drawn from a benchmark database maintained by a third party or from measurements made within the project boundary, where approved by the IME. The use of datasets from a benchmark database should include full citation of the database as well as a description of how datasets were extracted, including exclusion criteria for any records not used in the validation.
- All validation dataset sources must be reported. The same measurement dataset sources may be used for validating multiple PCs, where appropriate.
- Selection of validation datasets physically closest to the project geographic location should be prioritized. At minimum, validation datasets should come from the same climate zones, nationally defined agricultural land regions, countries and continents. Where datasets do not match these requirements, the project proponent must provide rationale demonstrating why the datasets are still appropriate for model validation. A justifiable rationale may include scenarios in which data that meet these requirements are poor quality or insufficient, requiring the use of data that are not proximal to the project area and model domain. In such cases, the requirements of the applied methodology to adjust for potential mismatches at future crediting events must be adhered to.

- Studies must report sufficient information on location, management, starting soil conditions and other model inputs to be modeled (i.e., providing enough information such that model inputs have low uncertainty relative to modeled results and allowing the model to be appropriately initialized). The amount of information needed to initialize and run a model is specific to the model and emissions source. Therefore, the reported information required to initialize and model a study should be described for the model version and parameter sets being validated, and any processes used to address unreported information fully described in the MVR.

- Studies reporting the effects of changing multiple practices concurrently ("stacked" practice changes) may be used provided that the composite of all studies used to validate a PC/CFG/ES combination contains at least one study that isolates the effect of the practice change being validated.

- In the case of SOC stocks, repeat measurements of SOC stock change must be able to capture multi-year changes, as practice effects on SOC may combine short and long-term changes in soil biogeochemical processes. Measurements from paired fields leveraging space-for-time analysis methods that approximate multi-year changes may be used for SOC validation. Newer methods for SOC stock monitoring are becoming available that are able to observe changes with greater precision at shorter time intervals. New and novel methods for SOC monitoring will be acceptable where there is peer-reviewed support of the method or independent expert support, both of which must be approved by the IME. New methods for SOC monitoring must be able to demonstrate accurate measurement of multi-year impacts on SOC stock changes. Measured datasets of SOC stock change may be made at any depth and it is likely that the depth increments used across different studies will vary. The model may be used either to predict SOC stock change at the same depths as the observations in individual studies, or data from across all studies may be adjusted, using a weighted averaging approach, to common depth increments that match the project or model. SOC stock changes in validation datasets need not have been calculated on an ESM basis.

- In the case of N_2O and CH_4 flux, any combination of measurements from chambers and/or eddy covariance flux towers from experimental sites are acceptable. Methods of temporal aggregation should be documented in the MVR (e.g., Mishurov & Kiely, 2011; Turner et al., 2016), as should the portions of the calendar year covered by aggregated N_2O and/or CH_4 measurements. Justification must be provided where portions of the year are missing.

- Project proponents are expected to use a process for selecting data for validating model performance and uncertainty that results in the assembly of validation datasets that are representative of the range of peer-reviewed observed results. Project proponents must describe the methods, selection process and data manipulations used to create the dataset applied in the model validation process. This includes describing search terms and databases used to identify available datasets, criteria used to select dataset sources, origin of extracted data (e.g., figures, tables, databases with DOI), original units of data and data uncertainty, and data manipulations used to convert original

units into the units described above. It is strongly recommended that project proponents follow PRISMA guidelines[6] for transparent reporting of meta-analysis and systematic review methods. The project proponent should report the number of validation data measurements of each data type (SOC, N_2O and CH_4) for each PC/CFG project domain combination. The project proponent should also include a histogram showing the range of validation data values (e.g., measured SOC change). Where validation data are unevenly distributed across the project domain (e.g., almost all validation data are reported in sandy soils, with only a few in soils with higher clay content), the method used to link validation data to model structural error (described in more detail in Section 5.2.5) should demonstrate that this discrepancy is addressed.

Requirement 2: Specific Dataset Requirements to Validate Model

The specific minimum requirements for validating model performance and uncertainty for a PC/CFG/ES combination are set based on the geographic extent of a project (i.e., the declared climate zones or nationally defined agricultural land regions), as well as the soil attributes encountered within the project (i.e., the declared soil textural classes and clay contents).

For all PC/CFG/ES combinations, each climate zone or nationally defined agricultural land region – depending on which is used – must be represented in the validation dataset. Additionally, the three soil textural classes expected to be most predominant in the project's geographic extent must be included in the dataset. The dataset must cover a range of soil clay content that spans 15 percentage points. Where clay content in the project area varies by less than 15 percentage points, the project proponent must demonstrate that the dataset covers a representative range of soil texture conditions likely to be encountered across the project site. Once validated, a PC/CFG/ES combination is approved for crediting within all declared climate zones/nationally defined agricultural land regions and for all declared soil textures.

The purpose of these minimum requirements is to ensure testing for generalized model performance (i.e., that a model has not been hyper-calibrated or overfitted for a specific combination of factors that leads to poor model performance in other contexts). It is in a project proponent's interest to exceed these minimums and validate the model across more climate zones/nationally defined agricultural land regions, soil texture classes and clay contents. This is because model prediction error must use the same dataset as model validation and will penalize the use of few data points (see Section 5.2.5). Where the available data fail to meet one of the minimum requirements due to data scarcity or fail one minimum while exceeding the other requirements in a way that supports a demonstrable test of generalized model performance, a case may be made for a valid exception to Requirement 2. Following are two examples:

- Only two of three declared climate zones are included in the validation dataset because no data could be found for the third, but five or more soil types are included (as opposed to three) and the closest geographic extent between experimental sites is at least 500 km.

[6] Available at: http://prisma-statement.org/

- Only two of three declared climate zones are included because no data could be found for the third, but five or more different soil types are included with a span in clay content greater than or equal to 30 percent.

Any such cases must be addressed explicitly in the MVR and must be approved by the IME and reviewed by the VVB.

Note that all model parameter sets used in crediting must be validated for each PC/CFG/ES combination (see Section 5.1). Where model parameter sets vary by climate zone/nationally defined agricultural land region, additional measurement datasets beyond the minimum described above may be required to ensure that all parameter sets are validated.

Special Rules for Practice Categories

For studies used to validate model performance and uncertainty in the cropping PC, any CFG occurring within the experimental period of measurements may be counted toward validation. For example, where two rotations are compared, one with a repeating corn-soy rotation and the other introducing a cover crop between corn and soy, the study may be used to validate all three of the CFGs associated with corn, soy and the cover crop for the cropping PC, provided that experimental measurements spanned at least one full rotation.

Where grazing practices have been validated on pasture and a CFG has been validated for either the cropping or soil disturbance PCs, the model may be considered validated for grazing on residue for that CFG. For grazing practices, pasture may be defined as any perennial grass or legume. C3 and C4 grasses do not need to be validated separately for pasture grazing.

For rice cropping systems, inorganic sulfur fertilizer application[7] may be considered an extra PC eligible for crediting due to its effects in reducing methane emissions. Validation of the inorganic sulfur fertilizer application PC is analogous to the inorganic nitrogen fertilizer application PC and encompasses the same domain of practice effects to be used in validation (i.e., magnitude, form, timing or method for sulfur fertilizer applied, with form encompassing inorganic S fertilizers and method encompassing surface, subsurface or irrigation-based application).

For studies focused on grass blends that include a mixture of C3 and C4, or N-fixing and non N-fixing, all CFGs represented in the blend may be considered represented in that study.

When validating a model for the organic amendments application PC, data from all CFGs classified as “annual” may be pooled and the validation result may be considered applicable for crediting of organic amendment practices in any annual CFG. Each perennial CFG must still be validated separately.

When validating a model for the inorganic N fertilizer application PC, validation data may be scarce for CFGs that fix N (e.g., soybean) because these crops are often grown without

[7] Note that inorganic sulfur fertilizer application is not an eligible project activity under VM0051, v1.0.

fertilization. Therefore, the model may be considered validated for an annual, herbaceous, C3, N-fixing crop where both of the following apply:

1) Inorganic N fertilizer application has been successfully validated for another annual CFG, and

2) The annual, herbaceous, C3, N-fixing CFG has been successfully validated for the cropping, planting and harvesting PC.

Cropping systems using irrigation as a normal part of management separate from practices intended to reduce emissions (i.e., where irrigation is present in both project and baseline) are not required to have the water management/irrigation PC validated.

Box 3: Summary of requirements described in Section 5.2.3

The following are required for the MVR:

- Full description of data requirements to initialize and run the model version and parameter sets accurately, as well as the process for addressing missing information
- A full accounting of the studies comprising the validation dataset for each PC/CFG/ES combination. Study attributes should include:
  - Citation,
  - Climate zone/nationally defined agricultural land region,
  - PC and CFGs being studied,
  - Soil texture(s) and clay contents being studied,
  - Experimental time period,
  - Depths of SOC measurements,
  - Measurement technique (e.g., dry combustion for SOC, chambers for N_2O),
  - Methods of temporal aggregation used for observations of N_2O and CH_4,
  - Portions of the calendar year covered by all N_2O and/or CH_4 measurements, with justification provided when portions are missing,
  - Number of observations used in validation,
  - Measurement uncertainty associated with replicates, where reported, and
  - Experimental location (only when split between calibration and validation).

The following are required upon request from the IME:

- Additional details for validation studies including, but not limited to:
  - Experimental location and corresponding climate zone/nationally defined agricultural land region;
  - Specific crops and practices being studied;
  - Original units of measurements; and
  - Mathematical transformations performed on measurement data.

- Study-specific use of data to initialize and run the model, as well as a record for completing missing information using the process described in the MVR

5.2.4 Assessment of Bias for Each Practice Category

For each PC/CFG/ES declared in Section 5.2.1, the model must be shown to be unbiased in estimating the change in SOC stocks, N_2O or CH_4 fluxes for the project domain defined in Section 5.2.2. This must be demonstrated using measured data that meet the requirements of Section 5.2.3. Bias, as a simplified version of average relative error (FAO, 2019), is calculated between measured data and model predictions. Bias indicates the average tendency of the modeled estimates to be larger or smaller than their observed counterparts (Moriasi et al., 2007). Positive values indicate model overestimation bias, meaning that the model overestimates the practice effect. A negative value indicates the model underestimates the practice effect.

The calculation of bias is defined as:

$$bias = \sum_{i=1}^{n}(P_i - O_i)/n \quad (1)$$

Where:

P_i	=	Predicted (i.e., modeled) change in SOC, N_2O or CH_4 for the *i*th observation of the practice change
O_i	=	Observed (i.e., measured) change in SOC, N_2O or CH_4 for the *i*th observation of the practice change
i	=	Index of observation within study
n	=	Number of observations in study

Model bias should be calculated for each study and a mean bias computed as the unweighted mean of all biases from individual studies. The mean bias should be less than or equal to an estimate of pooled measurement uncertainty (PMU). PMU is defined as the pooled standard error of all the measured values for a practice change, where standard error is derived from replicates of the measurements (Figure 4). Because not all studies will report measurement standard error, PMU may be computed from all studies found in an MVR that use the same measurement technique for the emissions reduction being evaluated (note that this implies studies evaluating different CFGs may be pooled together). Studies that use different measurement techniques at different time points should be conservatively excluded from both groups of studies for either technique.

Where it is not possible to obtain PMU, a default replacement value may be used for PMU that is based on typical measurement error for a given measurement technique. For SOC stock change, replacement PMU values must be based on the combined measurement error of the SOC content and bulk density measurement techniques, relevant to the project area and subject to IME approval and VVB review. Peer-reviewed literature, national/regional soil

inventory datasets or data from standards or quality assurance programs, such as the North American Proficiency Testing Program,[8] are suggested.

Finally, depending on the improved practice being evaluated, the observed rate of change for SOC stocks and associated uncertainty may differ across depth increments (generally uncertainty increases at lower depths). Studies that comprise a validation dataset are likely to employ a range of depth increments. For models designed to simulate stock changes across an entire depth profile as opposed to at individual depths, model bias should be assessed on the same depth profile equivalent as a study, as opposed to individual depth increments. Given these factors, project proponents may elect to define PMU as a function of cumulative sampling depth based on the observed measurement uncertainty in validation datasets. Total aggregate PMU is then calculated as the simple mean of study-specific PMU-equivalents as determined by the value of that function at that study's cumulative soil sampling depth.

$$\sigma_{meas} = \sqrt{\frac{\sum_{j=1}^{k} \sigma_j^2 (n_j - 1)}{\sum_{j=1}^{k} (n_j - 1)}} \quad (2)$$

Where:

k = Number of observations examined across all studies

σ_j = Standard error of the jth observed change in SOC, N_2O or CH_4

n_j = Number of replicate measurements used in the jth observation

A model is judged as valid where mean model bias is less than PMU, and model prediction error is determined as described in Section 5.2.5. Per-study bias should be reported, ranked from highest to lowest. The intention of reporting per-study bias as well as evaluating mean model bias compared to PMU is to avoid penalizing any one study in terms of measured data or model performance (i.e., where there are few or variable measured data or the model is biased in its prediction).

However, it should be recognized that there may be circumstances in which a model may be performing reasonably well even where mean bias is greater than PMU (e.g., due to limited availability of measured datasets or poor reporting of measured uncertainties). A project proponent is allowed to petition for validating the model for use, where it is clearly justified that the model shows reasonable overall performance given available measured data. Such a petition will need to be approved by an IME and reviewed by the VVB.

In this model evaluation framework, large model biases result in large residuals. Following guidance for model prediction error in Section 5.2.5, this means large model bias in either direction (positive bias or negative bias) will result in large predictive uncertainty, and thus increase credit deductions. Therefore, high model prediction error will be yielded in two circumstances: through low precision of an accurate model or high precision of an inaccurate model.

[8] Available at: https://www.naptprogram.org/

Figures 4 and 5 illustrate a process of meeting the requirements described in this section.

Figure 4: Visual summary of one possible approach to calculate measurement uncertainty of an observed practice change effect.

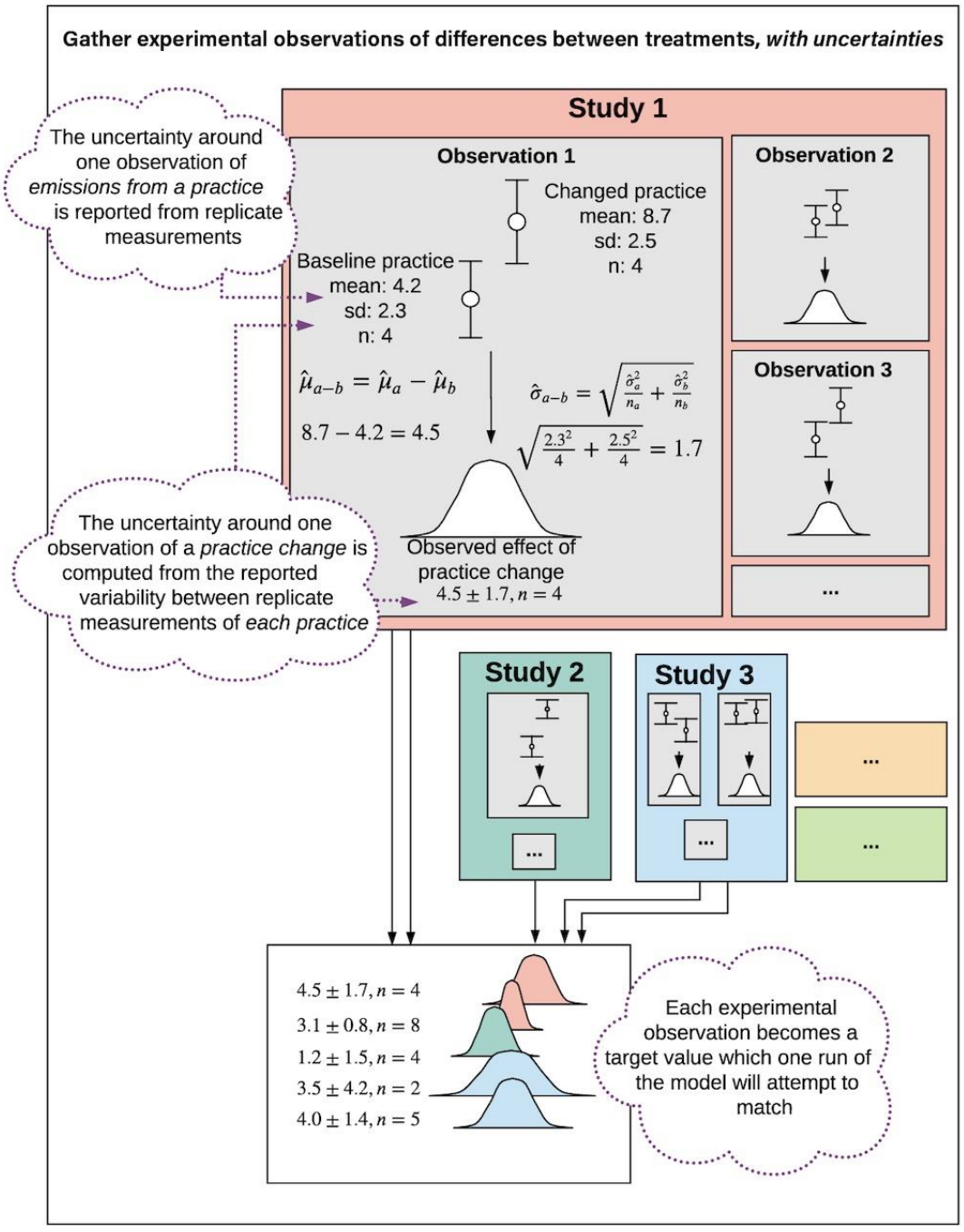
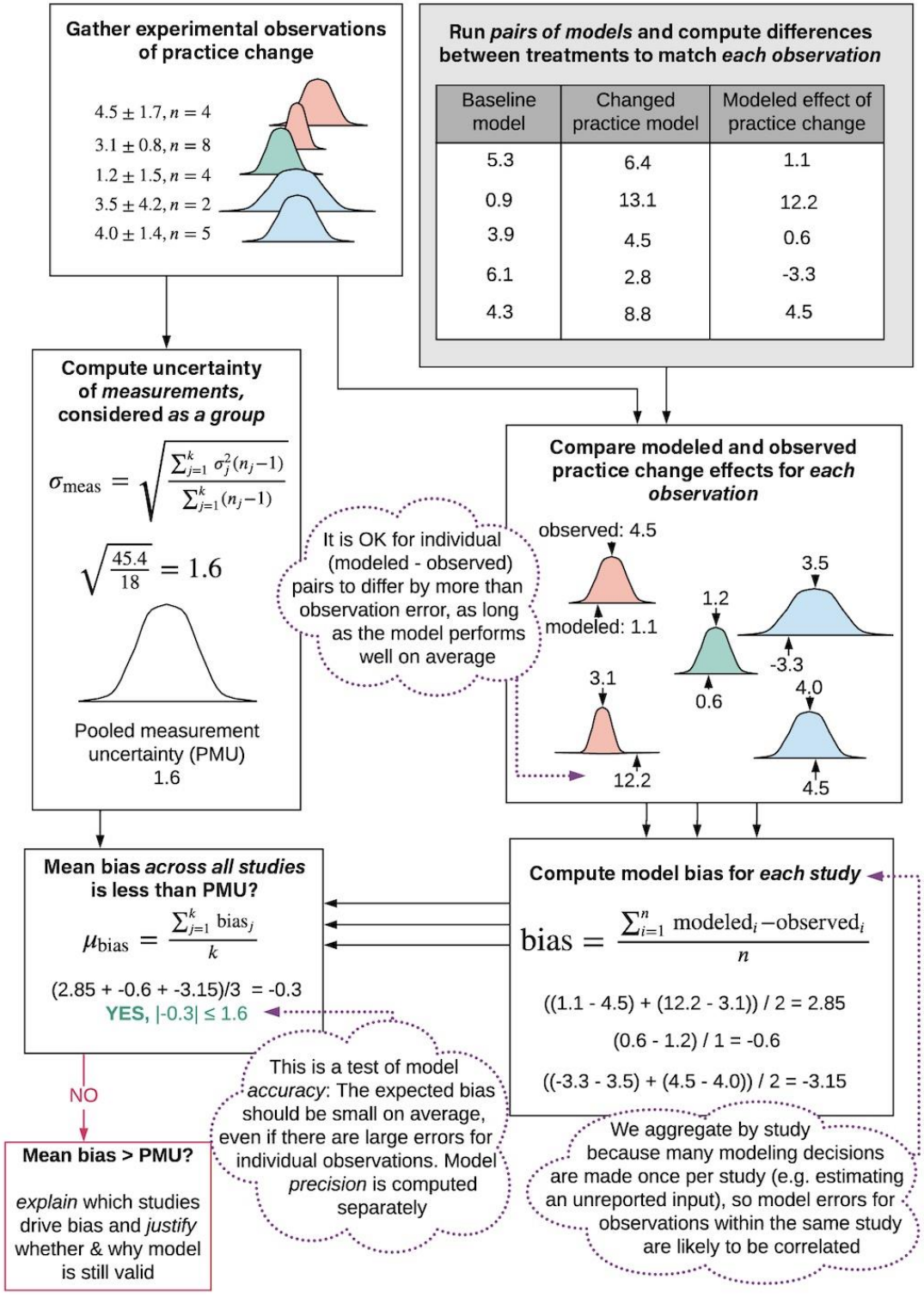


Figure 5: Visual summary of calculations for demonstrating that model bias is on a similar scale to measurement error



Box 4: Summary of requirements described in Section 5.2.4

The following are required for the MVR:

- One complete example derivation of:
  - Calculation of model bias for a study, per Figure 4
  - Calculation of PMU for a single measurement technique, per Figure 5
- All values of PMU used for each PC/CFG/ES combination validated
- All values of study bias for each study in a PC/CFG/ES validation dataset, ranked highest to lowest
- Average bias across all studies in a PC/CFG/ES validation dataset

The following are required upon request from the IME:

- Complete derivations and/or calculations made of PMU, study bias and average model bias for each PC/CFG/ES combination

5.2.5 Using Data to Evaluate Model Prediction Error

To evaluate the model for performance, the same datasets should be used to estimate the uncertainty of model predictions (i.e., the model prediction error) and evaluate model fit. The calculation of model uncertainty bounds associated with a particular prediction (i.e., the prediction interval) should account for cases where there are few validation data (e.g., by using a weakly informative prior when using a Bayesian framework, Figure 2B) as well as for data variability (i.e., with a wider posterior when data are more variable, where using a Bayesian framework). These features enable the model to adequately estimate the confidence in its predictions, as described next.

Two procedures for estimating model prediction error are permitted[9]:

- Analytical Calculation of Error: Where sufficient data from experimental field trials with controls are available, model prediction error is equivalent to the variance of the errors comparing simulated and observed emissions reductions/removals.

- Monte Carlo Simulation: Models with parameter uncertainty are used to simulate observations in the validation dataset with posterior predictive distributions (PPDs). PPDs are then iteratively resampled to generate an estimate of model prediction error.

Users must select one approach or the other for each carbon stock change or GHG flux being considered for validation. Where the Monte Carlo Simulation method is used, all model parameter distributions and other inputs (e.g., model hyperparameters) must be reported. Illustrative examples in this document are based on the Analytical Calculation of Error approach.

[9] Specific requirements and equations for estimating model prediction error in the applied methodology must be adhered to, where applicable.

In the MVR, as a check that model uncertainty bounds have been appropriately set, measured versus modeled results should be compared for each PC/CFG/ES combination for changes in SOC, N_2O and CH_4 (where relevant). A minimum confidence coverage of 90 percent should be demonstrated for 90 percent prediction intervals on independent test data (i.e., the 90 percent prediction intervals should contain the measured value for at least 90 percent of the validation data). Estimation of the 90 percent prediction interval will differ depending on the type of model being tested. For Bayesian models, posterior prediction intervals are a function of parameters calibrated as probability distribution functions. For non-Bayesian models, posterior prediction intervals may be determined by calculating the standard deviation of model residuals (i.e., the difference between modeled and observed values for a series of observations) and multiplying that by the 90 percent z-score of a standard normal distribution. The resulting value is added to and subtracted from the model prediction to give the upper and lower bounds, respectively, of the posterior prediction interval.

The prediction interval should be compared against independent observations[10] that were not used in calibrating model parameter distribution functions nor calculating the standard deviation of residuals. Leave-one-out or k-fold cross validation techniques are recommended to achieve this goal. Calculation of confidence coverage is then based on the total number of tests performed across all iterations.

It should be recognized that there may be circumstances where model uncertainty bounds are appropriately set even where 90 percent confidence coverage is not achieved, for example due to limited availability of measured datasets. A project proponent is allowed to petition for validating the model for use with such error bounds, where it is clearly justified that the model prediction error is appropriately set given available measured data (e.g., where error bounds that cover six out of seven observations or seven out of eight observations are missing, the confidence coverage drops below 90 percent). Such a petition must be approved by the IME and reviewed by the VVB.

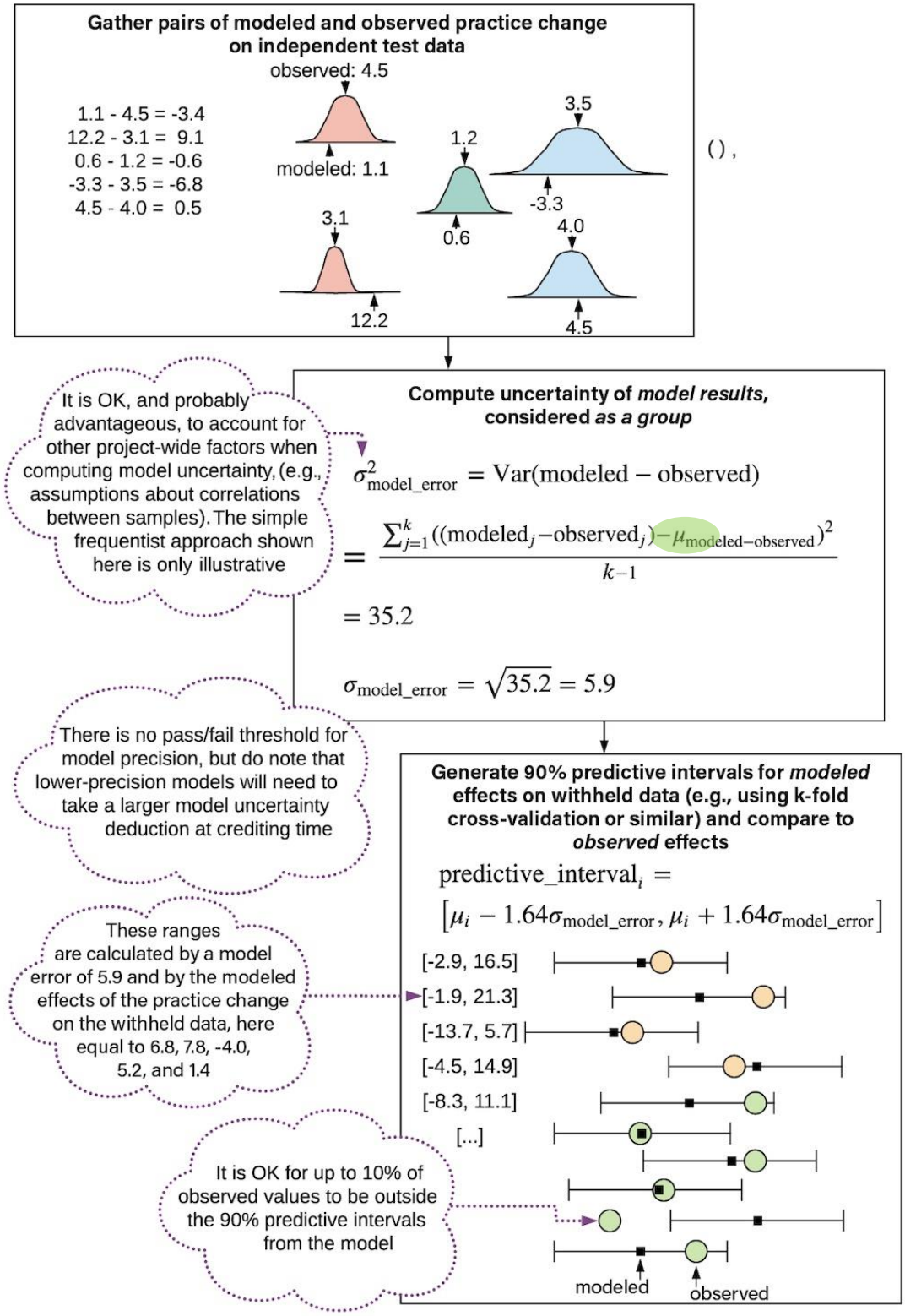
After demonstrating that the model uncertainty bounds have been appropriately set, the model prediction error is estimated using the entire dataset and reported in the MVR. The reported value should be used as the estimate of model prediction error in all uncertainty deduction calculations for the project per the procedures outlined in the applied methodology.

In the MVR, the following must also be included for each PC/CFG/ES combination and for changes in SOC, N_2O and CH_4:

- Scatterplot of the model predictions versus measurements;
- Histogram of residuals (the differences between predictions and measurements); and
- Mean squared error.

[10] An example of observations that are not independent is when SOC is measured from 0 to 4 years and then again at 8 years. There is one observation from 0 to 4 years and another from 4 to 8 years, with the latter simulation also having to run through the 0–4 year period. The measurements over these two periods are not independent and must not be split between the validation and calibration subsets. Further, the period from 0–8 years is not eligible to be included, even if kept in the same subset, since it is just a combination of the two shorter periods so would distort the apparent model prediction error.

Figure 6: Illustrative example of one possible approach to computing model prediction error and testing whether validation data are within predictive intervals. This figure represents one iteration in a k-fold or leave-one-out cross validation approach.



Box 5: Summary of requirements described in Section 5.2.5

The following are required for the MVR:

1) For each PC/CFG/ES combination:
   - Graphs of measured versus modeled results demonstrating that the 90 percent prediction intervals contain the measured value at least 90 percent of the time, per Figure 6
   - Scatterplot of model predictions versus measurements
   - Histograms of residuals (the differences between predictions and measurements)
   - Mean squared error

5.2.6 Review and Approval of Model Validation Reports

An MVR following the above requirements and guidance must be submitted with each monitoring report. Model validation requirements, including approval of MVRs, must be satisfied and confirmed at the latest prior to the completion of project verification activities. However, it is recommended that MVRs are submitted for IME assessment at project validation. Appendix 1 outlines detailed procedures on IME assessment and approval of model use and MVRs under this module .

The MVR must be project-specific, including demonstration of model validation for the project domain and PC/CFG/ES combinations. Further, MVRs must be:

1) Independently assessed by an IME who is hired by the VVB and who must fulfil the minimum qualifications defined by Verra (see Appendix 1 for details on IME minimum qualifications); or

2) Accepted for publication in one of the peer-reviewed journals listed in Table 3 and reviewed by an IME who is hired by the VVB and who must fulfil the minimum qualifications defined by Verra. Where the peer-reviewed publication option is pursued, the following also applies:

   a) Where the paper has passed peer review and has been accepted for publication with revisions that do not change any aspects of model validation following the guidance in this document, it is acceptable as an MVR even before the paper has been published. In this circumstance, the project proponent should submit the peer-reviewed publication and responses to all revisions that clearly demonstrate revisions do not impact model validation.

   b) Model validation may be completed using a different method than explicitly evaluating bias and goodness of fit as described above. The paper must demonstrate that separate datasets were used for model calibration and model validation (see Section 5.1). Model validation must demonstrate that the model was found acceptable for use by the peer reviewers for a given biophysical domain and set of practices.

   c) The biophysical domain and practices used in the publication must be shown to completely meet the same domain requirements laid out in Sections 5.2.2 and 5.2.3, as well as cover the PCs and CFGs identified in Section 5.2.1.

d) The same datasets used in the peer-reviewed model validation must be used to calculate model prediction error in the project and evaluate model uncertainty.

e) The same model version and model parameter values/parameter set values must be used in the paper as are used in the project.

f) As a means of enhancing transparency with peer reviewers, the authors must clearly state the purpose of the paper as being to validate the model for use in generating verifiable carbon credits and therefore the ISO 14064 principles for GHG accounting should be kept in mind.

g) The project proponent must submit a sub-report outlining how the above requirements have been met and clarifying any aspects of the peer-reviewed paper as it pertains to the overall requirements of the MVR.

For each subsequent monitoring report, as long as a project area remains constant or is only expanded to include new fields that already fit within the validated project domain, the existing MVR may be used. Models may be recalibrated or revised based on new data, or a new model may be applied, provided the same model version using the same parameters/parameter sets is applied in the baseline and project scenarios. Where the project is expanded to new PCs, CFGs or emissions sources, or the model is changed in a way that substantively affects model runs and the estimated ERRs, the MVR must be:

1) Revised, reviewed by an IME and re-submitted; or

2) Submitted and accepted for publication as a new journal article in one of the peer-reviewed journals listed in Table 3, and reviewed by an IME.

In both cases, the IME must assess the MVR or peer-reviewed publication to confirm its appropriateness for the project domain. All MVRs will be made publicly available in the Verra registry.[11]

Table 3: Journals approved for publication of model validation reports[12]

Approved Journals for Publishing Model Validation Reports	
Agricultural and Forest Meteorology	Global Change Biology
Agricultural Systems	Journal of Environmental Quality
Agriculture, Ecosystems and Environment	Journal of Geophysical Research - Biogeosciences
Agronomy Journal	Nutrient Cycling in Agroecosystems
Atmospheric Environment	Paddy and Water Environment
Biogeochemistry	Plant & Soil

[11] Project proponents may elect to petition Verra and the VVB performing validation/verification to keep elements of either document confidential where they contain commercially sensitive information, but such decisions are left to Verra's discretion.

[12] The addition of other journals may be proposed through a tool revision.

Biogeosciences	PLoS ONE
Ecological Applications	Rice Journal
Ecological Modeling	Science of the Total Environment
Ecosystems	SOIL
Environmental Modelling and Software	Soil & Tillage Research
Environmental Pollution	Soil Biology & Biochemistry
Field Crops Research	Soil Research
Frontiers in Ecology and the Environment	Soil Science Society of America Journal
Geoderma	Soil Use & Management
Global Biogeochemical Cycles	Vadose Zone Journal

5.3 Substitution for Missing Crop Types

Where it becomes clear during the calibration and validation process that no data are available to validate a specific crop grown in the project, an alternative crop from the same (validated) CFG may be used as a substitute in both the baseline and project simulations. Where there are no other crops from the same CFG that have been validated, and thus the CFG has not been validated, substitutions from other CFGs may be made that meet the specific requirements for the baseline and project simulations outlined below. This method depends on the availability of alternative, conservative CFGs that meet all the above criteria and have been validated; without alternatives, substitution is not possible.

- **Baseline:** Replace the missing crop with a crop from a more conservative, validated CFG, such as an unfertilized perennial grass for an annual herbaceous crop. Conservative in the case of a baseline simulation means emitting fewer GHG emissions than the missing crop and this should be clearly supported with peer-reviewed literature.
- **Project:** Replace the missing crop with a crop from a more conservative, validated CFG. Conservative in the case of a project simulation means emitting more GHG emissions than the missing crop and this should be clearly supported with peer-reviewed literature.

6 DATA AND PARAMETERS

6.1 Data and Parameters Available at Validation

Data/Parameter	P_i
Data unit	t CO_2e
Description	Change in SOC, N_2O or CH_4 predicted by modeling the *i*th validation measurement
Equations	(1)
Source of data	The predicted value of change in SOC, N_2O or CH_4 is modeled.
Value applied	Not applicable
Justification of choice of data or description of measurement methods and procedures applied	An empirical or process-based model used to estimate stock change/emissions that meets applicability conditions of this module should be used.
Purpose of data	Calculation of baseline and project emissions
Comments	None

Data/Parameter	O_i
Data unit	t CO_2e
Description	Change in SOC, N_2O or CH_4 observed in the *i*th validation measurement
Equations	(1)
Source of data	See Section 5.2.3
Value applied	The observed value of change in SOC, N_2O or CH_4 is determined from validation datasets.
Justification of choice of data or description of measurement methods and procedures applied	Validation data meeting requirements in Section 5.2.3 may be used.
Purpose of data	Calculation of baseline and project emissions

Comments	None

Data/Parameter	*n*
Data unit	Number
Description	Number of values in the study used for validation
Equations	(1)
Source of data	See Section 5.2.3
Value applied	The number of values in the validation dataset is determined from the validation data.
Justification of choice of data or description of measurement methods and procedures applied	Validation data meeting requirements in Section 5.2.3 may be used.
Purpose of data	Calculation of baseline and project emissions
Comments	None

Data/Parameter	*i*
Data unit	Number
Description	Index of current observation within a study used for validation
Equations	(1)
Source of data	See Section 5.2.3
Value applied	The value is incremented for each observation within the validation study being considered.
Justification of choice of data or description of measurement methods and procedures applied	Validation data meeting requirements in Section 5.2.3 may be used.
Purpose of data	Calculation of baseline and project emissions
Comments	None

Data/Parameter	k
Data unit	Number
Description	Number of observations used for validation
Equations	(2)
Source of data	See Section 5.2.3
Value applied	The sum of the number of observations in all studies used for validation
Justification of choice of data or description of measurement methods and procedures applied	Validation data meeting requirements in Section 5.2.3 may be used.
Purpose of data	Calculation of baseline and project emissions
Comments	None

Data/Parameter	σ_j
Data unit	t CO_2e
Description	Standard error of the observed change in SOC, N_2O or CH_4 from a given practice in the jth observation
Equations	(2)
Source of data	See Section 5.2.3
Value applied	The standard error is determined from the validation data.
Justification of choice of data or description of measurement methods and procedures applied	Validation data meeting requirements in Section 5.2.3 may be used.
Purpose of data	Calculation of baseline and project emissions
Comments	None

Data/Parameter	n_j
Data unit	Number

Description	Sample size of the *j*th observation
Equations	(2)
Source of data	See Section 5.2.3
Value applied	The sample size is determined from the validation data.
Justification of choice of data or description of measurement methods and procedures applied	Validation data meeting requirements in Section 5.2.3 may be used.
Purpose of data	Calculation of baseline and project emissions
Comments	None

Data/Parameter	*j*
Data unit	Number
Description	Index of current observation within the entire validation dataset
Equations	(2)
Source of data	See Section 5.2.3
Value applied	The value is incremented for each observation across all studies in the validation dataset.
Justification of choice of data or description of measurement methods and procedures applied	Validation data meeting requirements in Section 5.2.3 may be used.
Purpose of data	Calculation of baseline and project emissions
Comments	None

6.2 Data and Parameters Monitored

This section is not applicable; all data and parameters are available at model validation.

7 REFERENCES

Bruun, S., Christensen, B. T., Hansen, E. M., Magid, J., & Jensen, L. S. (2003). Calibration and validation of the soil organic matter dynamics of the *Daisy* model with data from the Askov long-term experiments. *Soil Biology and Biochemistry*, *35*(1), 67–76. https://doi.org/10.1016/S0038-0717(02)00237-7

Climate Action Reserve. Soil Enrichment Protocol. Available at: http://www.climateactionreserve.org/how/protocols/soil-enrichment/

Dietze, M. C. (2017). *Ecological forecasting*. Princeton University Press.

Food and Agriculture Organization of the United Nations (FAO) (2019). *Measuring and modelling soil carbon stocks and stock changes in livestock production systems: Guidelines for assessment*. Available at: http://www.fao.org/3/CA2934EN/ca2934en.pdf

Intergovernmental Panel on Climate Change (2019). *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*. Available at: https://www.ipcc.ch/report/2019-refinement-to-the-2006-ipcc-guidelines-for-national-greenhouse-gas-inventories/

Liang, Y., Gollany, H. T., Rickman, R. W., Albrecht, S. L., Follett, R. F., Wilhelm, W. W., Novak, J. M., & Douglas Jr, C. L. (2009). Simulating soil organic matter with CQESTR (v. 2.0): Model description and validation against long-term experiments across North America. *Ecological Modelling*, *220*(4), 568–581. https://doi.org/10.1016/j.ecolmodel.2008.11.012

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *American Society of Agricultural and Biological Engineers*, *50*(3), 885−900. https://doi.org/10.13031/2013.23153

Mishurov, M., & Kiely, G. (2011). Gap-filling techniques for the annual sums of nitrous oxide fluxes. *Agricultural and Forest Meteorology*, *151*(12), 1763–1767. https://doi.org/10.1016/j.agrformet.2011.07.014

Natural Resources Conservation Service (2014). *2014 Farm Bill - Conservation Compliance Crop List*. United States Department of Agriculture. Available at: https://www.nrcs.usda.gov/2014-farm-bill-conservation-compliance-crop-list

Turner, P. A., Baker, J. M., Griffis, T. J., & Venterea, R. T. (2016). Impact of kura clover living mulch on nitrous oxide emissions in a corn-soybean system. *Journal of Environmental Quality*, *45*(5), 1782–1787. https://doi.org/10.2134/jeq2016.01.0036

Wallach, D., Makowski, D., Wigington, J., & Brun, F. (2019). *Working with Dynamic Crop Models: Methods, Tools and Examples for Agriculture and Environment* (3rd ed.). Elsevier Science.

Wikle, C. K., & Berliner, L. M. (2007). A Bayesian tutorial for data assimilation. *Physica D: Nonlinear Phenomena*, *230*(1–2), 1–16. https://doi.org/10.1016/j.physd.2006.09.017

Yang, J. M., Yang, J. Y., Liu, S., & Hoogenboom, G. (2014). An evaluation of the statistical methods for testing the performance of crop models with observed data. *Agricultural Systems*, *127*, 81–89. https://doi.org/10.1016/j.agsy.2014.01.008

Yeluripati, J. B., van Oijen, M., Wattenbach, M., Neftel, A., Ammann, A., Parton, W. J., & Smith, P. (2009). Bayesian calibration as a tool for initialising the carbon pools of dynamic soil models. *Soil Biology and Biochemistry*, *41*(12), 2579–2583. https://doi.org/10.1016/j.soilbio.2009.08.021

APPENDIX 1: ASSESSMENT BY INDEPENDENT MODELING EXPERT (IME)

This appendix supports approval of a model for use by a project using a quantification approach based on biogeochemical modeling.

A1.1 Steps for Model Assessment by an IME

The following steps must be taken for biogeochemical model use and approval within a VCS ALM project:

1) **Project proponent generates an MVR.**
   The project proponent must generate an MVR demonstrating that a model is valid and used appropriately to issue credits in a specific project according to the requirements specified in the applied methodology and this module.

2) **VVB selects and contracts an IME to assess the MVR.**
   The VVB may select an IME from the list on Verra's website or contract a new IME to review the MVR. New IMEs must fulfil the minimum qualifications defined by Verra (see "Minimum qualifications of IMEs" below).

3) **IME assesses MVR and generates an IME assessment report.**

   The IME must generate an assessment report based on the MVR assessment that:

   a) Confirms that the selected model meets the applicability conditions stated in Section 4 of this module and verifies it is appropriate for the agricultural management system and biophysical conditions occurring in the project;

   b) Assesses the quality of model-driving input data (experimental data of soil emission reductions) and the pooled measurement uncertainty;

   c) Confirms that the calibration procedure and generation of model parameters and parameter sets meet the requirements stated in Section 5 of this module (see Box 1 for details), including the definition of the project domain (see Box 2 for details); and

   d) Confirms that the requirements related to model bias, model prediction error and goodness of fit have been met to estimate SOC stock change and, where applicable to the project, flux change of N_2O and CH_4 (see Boxes 3, 4 and 5 for details).

   Where the peer-reviewed publication option is pursued instead of the MVR, the IME must assess the publication based on the requirements listed in Section 5.2.6(2).

   Project proponents must provide all relevant project documentation to the IME. Project proponents must promptly respond to questions and findings from the IME and submit additional evidence and assist in arranging meetings with stakeholders, as requested. The burden of proof in the assessment process ultimately rests with the project proponent.

4) **IME assessment report submitted to VVB for approval.**
The IME assessment report must be submitted to and approved by the VVB alongside other project documentation as stated in Section 5.2.6 of this module. The IME must keep the VVB informed regarding questions and resolved findings with the project proponent (e.g., through email copies). Where the VVB has questions or concerns about the IME assessment report, the IME and VVB must iterate until such questions or concerns are satisfactorily resolved and the VVB approves the report.

All MVRs and IME assessment reports will be made public as part of the project documentation in the Verra registry.

A 1.2 Minimum Qualifications of IMEs

Verra defines minimum qualifications that IMEs must fulfil to perform evaluation of the use of process-based biogeochemical models under the VCS Program. IMEs must meet the following criteria:

1) Demonstrated competency in quantifying GHG fluxes associated with ALM through the use of biogeochemical models, specifying specialization in certain practices or land uses and regional/country expertise, where relevant. The prospective IME must have at least five years of relevant work experience.
2) Stated ability to assess specific model types based on demonstrated use of the model to be evaluated or conceptually similar models. Prospective IMEs may demonstrate expertise through citation of their peer-reviewed scientific publication(s), or through reference to relevant project reports, presenting or using specific model(s).
3) Demonstrated freedom from conflict of interest. This must be established by disclosing all relevant organizational affiliations and anything else that may give rise to a conflict of interest.
4) Recommended by two references, preferably researchers and academic staff.

The IME Qualification Form must be used to provide evidence demonstrating that the expert meets the above criteria. The IME Qualification Form is available on the VMD0053 webpage.

DOCUMENT HISTORY

Version	Date	Comment
VMD0053, v1.0	19 Oct 2020	Initial version
VMD0053, v2.0	30 May 2023	• Additional guidance on procedures related to different steps of model validation • Clarifications on the role of and process for assessment by the independent modeling expert (IME) • New guidance on minimum qualifications and process to be listed as an IME • General improvements, errata and clarifications
VMD0053, v2.1	25 Mar 2025	• Minor edits expanding applicability to other ALM methodologies in the VCS Program