VCS Module

VMD0053

MODEL CALIBRATION, VALIDATION, AND UNCERTAINTY GUIDANCE FOR THE METHODOLOGY FOR IMPROVED AGRICULTURAL LAND MANAGEMENT

Version 1.0

19 October 2020

Sectoral Scope 14

CONTENTS

1 SOURCES

This module is based on the following document:

- *Requirements and Guidance for Model Calibration, Validation, Uncertainty, and Verification for Soil Enrichment Projects, Version 1.0*

2 SUMMARY DESCRIPTION OF THE MODULE

This module provides procedures for calibration, validation, and verification of empirical or process-based models used to estimate stock change/emissions with application of VM0042 Methodology for Improved Agricultural Management. It is intended to provide a standardized approach to test model performance as a component of credit quantification in a VCS ALM project. This document must be used for any gas or pool for which models are employed for quantification outside of the protocol equations. A Model Validation Report generated by following this document is designed to support independent expert review of a model proposed for use in a VCS ALM project, grouping model testing by combinations of geographic regions, crop types and practice changes where a model may be used to issue credits. The Model Validation Report is also designed to support independent verification that a credit-compliant model is valid and used appropriately to issue credits in a specific project. As written, this document supports validation of a model for use by a VCS ALM project following Approach 1 (Measure and Model), when the baseline is also modeled. When a performance benchmark must be used as the baseline for a project following Approach 1, these guidelines will need to be followed in the context of the modeling approach defined and approved by Verra for the performance benchmark. This will ensure that the model is appropriately tested for model performance and appropriately defined bounds for model prediction error.

The requirements and guidance presented in this document fall into two main categories: 1) standardizing best practices for VCS-appropriate use of peer-reviewed observed experimental data to test a model and determine model prediction error and 2) standardizing demonstration of acceptable fit and a lack of bias when a model is being used to estimate soil organic carbon (SOC) stock change and, if applicable to the project, flux change of N2O and CH4.

Requirements falling into category 1 are meant to address the importance of using high-quality observed experimental data of soil emission reductions as the basis of evaluating model performance. Changes in agricultural practices have a great diversity of impacts on soil emissions. Soil emissions are also highly variable. There has been rapid growth of new studies and experimental methods to capture variance, increase precision, and reduce uncertainty. Further, the format of experimental data can be highly variable across published studies. Requirements described in this document are meant to ensure that appropriate and consistent methods are followed to locate, aggregate, and use observed data for model improvement and testing.

Requirements falling into category 2 are meant to provide specific guidance for using the above datasets for model calibration, validation, and the determination of a model's prediction error, in the context of measurement uncertainties. These are highly technical processes that vary widely across areas of scientific research. The Model Validation Report aims to ensure that model validation is specific to the model being proposed for use in the project, is appropriate for the cropping 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 a Model Validation Report. Requirements for submitting Validation Reports are outlined in Section 5.2.6. Validation Reports must show that all requirements for a specific project have been met, including proof that the same model version and parameter sets are used, and that all project domain and crop functional group/practices category combinations have met minimum requirements for model validation. Validation Reports must either be independently assessed by a third party expert or accepted for publication in one of the peer-reviewed publications listed in Section 5.2.6, and reviewed by an independent third-party expert. Model Validation Reports will be public documents.

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 Model Validation Report can be used. If the project is expanded to new practice categories, new crop functional groups, or the model is changed, the Model Validation Report needs to be:

- Revised, reviewed by an independent third-party expert, and re-submitted, or
- Submitted and accepted for publication as a new journal article in one of the peer-reviewed journals listed in Section 5.2.6 and reviewed by an independent third-party expert.

In both cases, the Model Validation Report or peer-reviewed publication must also be reviewed by an approved VVB to confirm its appropriateness for the project domain.

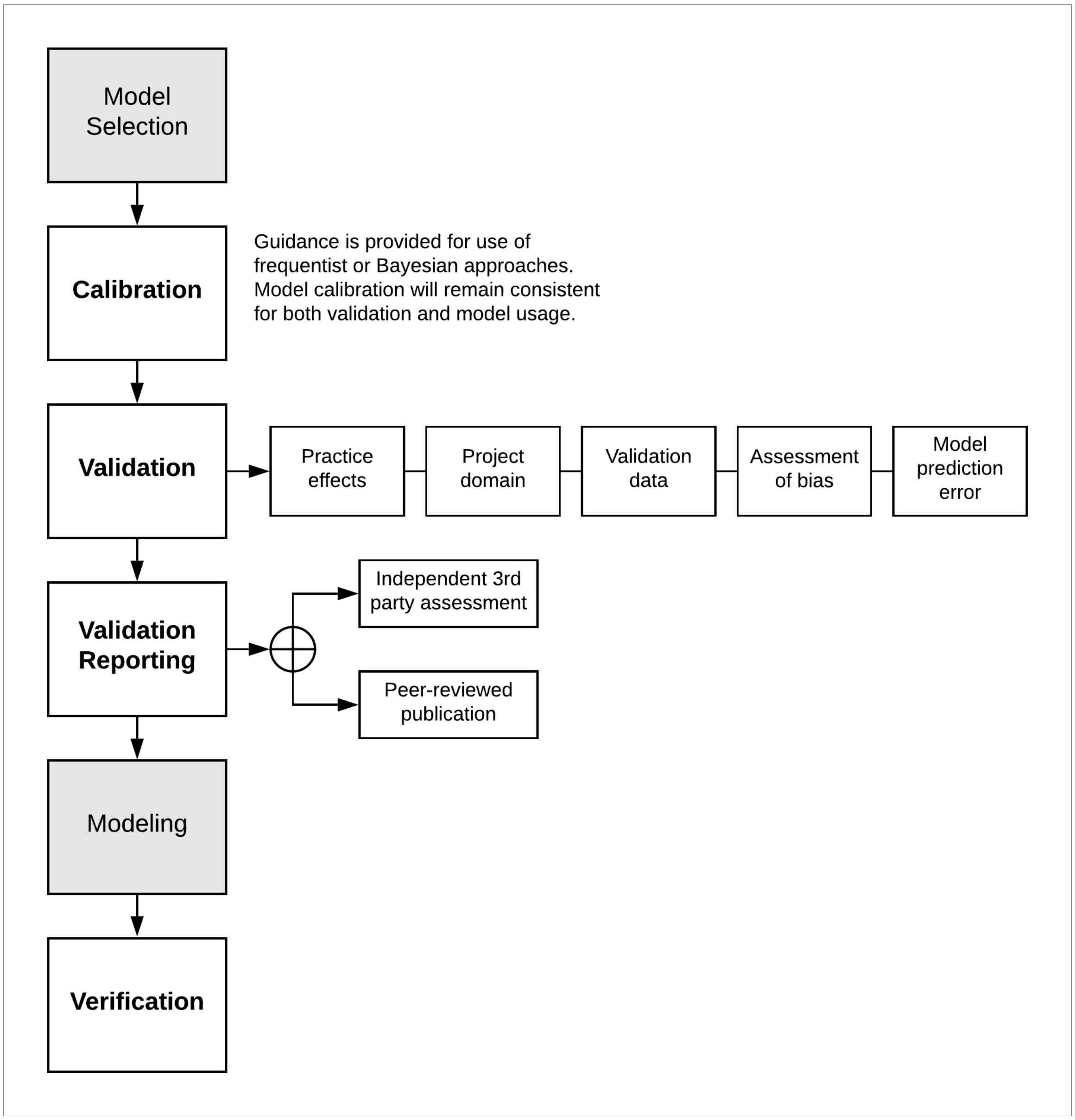

Figure 2.1 Steps related to the use of models for quantification in VCS ALM projects

3 DEFINITIONS

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*.

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

Model-driving input data
Data provided to the model 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

Model prediction error
The uncertainty in a model's prediction as determined from comparison to measurements, required in this document to be the same measurements used to validate the model.

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") may include source code, internal parameters that are not adjusted during calibration, default values for parameters or input data, or generally 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 sets
The set of mathematical values and constants contained in a model that characterizes the biophysical and biogeochemical system being represented.

Pooled measurement uncertainty
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.

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 typically uses datasets independent of datasets used in model calibration, unless (e.g. in a data-limited situation) a statistical approach like k-folding is applied.

4 APPLICABILITY CONDITIONS

This module applies where empirical or process-based models used to estimate stock change/emissions meet specific conditions. Models must be:

1. Publicly-available;

2. Shown in peer-reviewed scientific studies to successfully simulate changes in soil organic carbon and trace gas emissions resulting from changes in agricultural management included in the project description;
3. Able to support repetition of the project model simulations. This includes clear versioning of the model used in the project, stable software support of that version, as well as fully reported sources and values for all parameters used with the project version of the model. For stochastic models, this means providing the seeding sequence to the random number generator so that model runs can be reproduced. Where multiple sets of parameter values are used in the project, full reporting includes clearly identifying 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 can demonstrate evidence of expertise with the model via authorship on peer-reviewed model publications or authorship of reports for entities supporting climate smart agriculture, such as FAO or a comparable organization), and must describe the data sets 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, with model prediction error calculated using datasets as 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

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 a method of harvest using an average of observed values); and
- 'Tuning' a set of model parameters that may not be able to be measured directly using overall model performance and an understanding of model sensitivities (e.g. adjusting a constant downregulating 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, which are based on statistical synthesis of observations and cannot be extended outside of where observations are available.

For any model used with this module, 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, aka "garbage in, garbage out". Datasets for calibration and validation may either be kept completely isolated from each other or may be drawn from a single pool using a statistical process such as k-folding that guarantees independence.

For the purposes of this protocol, calibration and validation data should be demonstrably independent. This requirement can be met if datasets used for calibration and validation do not overlap in experimental research locations and are not taken from the same experimental study. If 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 crop functional group/practice category combination level (Section 5.1.2). For example, if root measurements and N2O flux measurements from a subset of treatments in a tilled soybean/corn rotation experiment were used for model calibration, the N2O flux measurements from the remaining treatments in the same study could not be used as validation data for either the corn or the soy crop functional group and tillage practice effect combinations. However, if at the same research facility N2O flux was measured in a demonstrably separate corn/soy rotation experiment (separate in space or time, with separate experimental design or intention), those data would be permissible for inclusion in model validation. Depending on the model, in some cases 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 should be clearly explained and presented for review in the Model Validation Report.

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

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

'Parameter sets', in this context, mean 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 (for example through a Bayesian statistical procedure) must be declared and shown to be set appropriately following the above calibration requirements. It is encouraged for model parameters to be as generalizable as possible across the project domain, with minimal use of different parameter sets. However, it is acceptable for different parameter sets to be used as long as they are defined at either the scale of IPCC climate zones or at the scale of nationally defined agricultural land regions, for example Land Resource Regions in the US (Section 5.2.2). If a project is using nationally defined agricultural land regions, the definition must be VVB approved, and parameter sets should be declared at a minimum for each individual agricultural land region included in the project. The same parameter set should be used to simulate all crop functional groups and practice categories 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 Model Validation Report and presented as parameter sets specific within each land area boundary where the crop is simulated, to ensure their appropriate use in model validation and project simulations.

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 Section 5.2. If the minimums described in Section 5.2 do not result in all parameter sets being validated, and additional steps are not taken to validate all parameter sets, unvalidated parameter sets cannot be approved for use in the project.

Because biogeochemical models often contain a large number of parameters, different strategies can be employed to perform calibration. General guidance for frequentist and Bayesian approaches are provided in Sections 5.1.1 and 5.1.2.

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 parameters to estimate, 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-review literature and span a wide range of complexity and automation in their approaches (e.g. Bruun et al. 2003, Yeluripati et al. 2009, Liang et al. 2009).

5.1.2 Guidance on Bayesian Methods for Calibration, Validation, and Error

Model calibration can 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 can be computationally demanding with soil biogeochemical models, which can have dozens to hundreds or more parameters. Parameter values in these types of models can also be difficult to constrain, i.e., use data or existing knowledge to set limits on the range of values that a parameter may have, and define its probability distribution across that range. When 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, can be wide unless the observed data are strongly informative, i.e., have highly accurate and precise values. The following figure illustrates a strong prior belief (A) versus a weak prior belief (B).

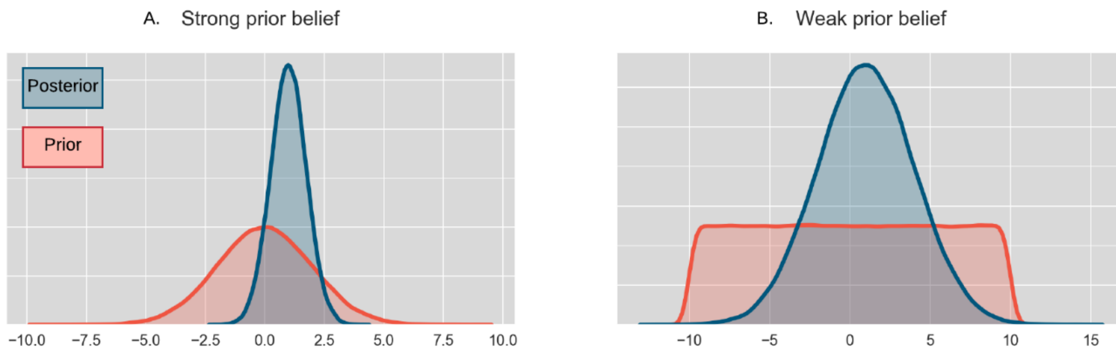

Figure 5.1 Comparison of prior and posterior distributions when there is strong prior belief (e.g., strong and consistent evidence and prior analyses, A), versus weak prior belief (e.g., weak or variable evidence or no prior analyses, B).

Across dozens or hundreds of parameters, Bayesian methods can 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 its central role in the developing field of Ecological Forecasting (Dietze, 2017), as well as in the Predictive Ecosystem Analyzer Project data-model integration system.[1] 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 5.2 presents a conceptual workflow for a Bayesian approach to these analyses.

[1] Accessible at: pecanproject.org

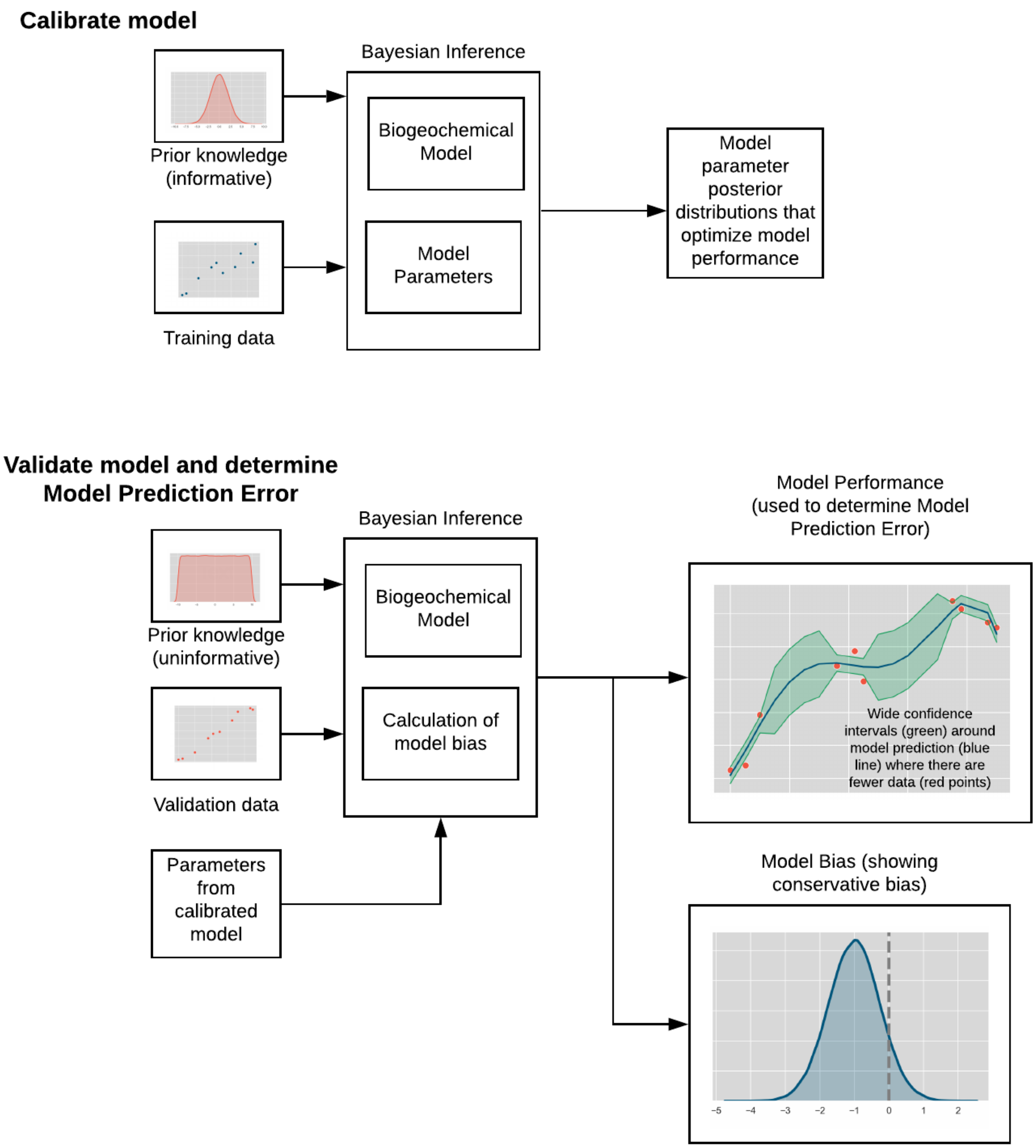

Figure 5.2 Conceptual framework for Bayesian approach to model calibration and validation. In this example model calibration is a separate analytical process from validating model performance and determining model prediction error. In a fully integrated analysis, informative posteriors from model calibration might be used as priors in model validation.

Box 1. Summary of Requirements described in Section 5.1

Required for the Model Validation Report

- Model version, as defined in Section 3.
- Description of the model calibration process, including the adjustment of model parameters with directly measured values (e.g. LAI 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). If there is justification to claim an allowance for crop growth parameter sets to vary within climate zones/nationally defined agricultural land region (e.g. varying maturity groups), crop growth parameter sets and their use must be documented per each LRR where the crop will be simulated.
- Justification for splitting of experimental data between calibration and validation (where applicable), clearly described at the crop functional group/practice category/emissions source combination level.

Required Upon Request of the Verification Team

- 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

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 do so, each practice must be binned into the Practice Categories (PCs) shown in Table 5.3 to demonstrate the domain of practice effects and the categories requiring evaluation. Validating model performance and uncertainty within a practice category can be accomplished using any practice effect in the category domain, evaluated using appropriate experimental data meeting requirements described below. Projects are encouraged to evaluate a range of practice effects in each practice category domain.

Table 5.3. 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, timing, or method for nitrogen fertilizer applied, with form encompassing inorganic N fertilizers, and method encompassing surface, subsurface, or irrigation-based application
Organic amendments application	Magnitude, form, timing, method or variation in C:N ratio for organic amendments applied. Forms include and are not limited to biochar, mulch, compost, and animal manure, and methods encompass surface, subsurface, or irrigation-based application
Water management/irrigation	Magnitude, timing, source or method of irrigation water applied
Soil disturbance and/or residue management	Soil disturbance including tillage and compaction, and residue management encompassing soil exposure after harvest and physical incorporation of green manure
Cropping practices, planting and harvesting (e.g., crop rotations, cover crops)	Variety of crops grown, increasing crop rooting depth, may include cover crops and soil preparations such as changing soil pH through liming
Grazing practices	Any of the following: presence/absence of grazing, stocking density, forage type or quality, species of grazers, mixed or single species herds, loading weight, grazing time, and rest/recovery periods

A project developer must declare all practice effects requiring evaluation for the project.

5.2.2 Define the Project Domain

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

1. Declare Project Crop Functional Groups

Crop functional groups (CFGs) for each practice category must be declared. Individual crop types can be grouped into functional groups across crops sharing unique combinations of the following attributes:

- N fixation (Y/N);
- Annual/perennial (A/P) (defined in accordance with the NRCS Conservation Compliance categorization of crops[2]);
- Photosynthetic pathway (C3/C4/CAM);
- Growth form (tree/shrub/herbaceous) (trees and shrubs have woody plant growth, versus herbaceous species that do not grow woody plant material); and/or
- Flooded/not flooded.

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. If using climate zones, for each practice category, the full list of climate zones encompassed in the project domain must be declared, following the climate zone definitions given in the *2019 Refinements to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories.* If using nationally-defined agricultural land regions, for example Land Resource Regions in the United States, it must be approved by a VVB as appropriate for the project. The full list of defined land boundaries encompassed in the project domain must then be declared.

3. Declare Project Soils

Soils are to be declared for each practice category in terms of (1) soil textural class and (2) the associated clay content[3] of that class. NRCS 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 Section 5.2.1 and 5.2.2

Required for the Model Validation Report

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

Required Upon Request of the Verification Team

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

[2] Resource can be found here: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/programs/farmbill/?cid=stelprdb1262733

[3] See Table A-1 for clay contents of NRCS soil textural classes. https://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/survey/?cid=nrcs143_014055

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 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 practice categories as well as combinations of practice categories (e.g. "stacked" practices), provided the practice category in question is experimentally varied and measured within the study. Some hypothetical examples of acceptable experimental treatments to evaluate practice categories are given in the following table:

Table 5.4. Examples of Acceptable Experimental Treatments to be used in Evaluating Practice Categories

Experimental treatment	Practice Category
Comparison of two different application rates of urea	Inorganic nitrogen fertilizer application
Comparison of conventional tillage using moldboard plow to strip tillage.	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 sources(s) of interest (SOC stock change and/or N_2O and CH_4 change, as applicable), ideally using control plots to test the practice category.
- All validation dataset sources must be reported. The same measurement dataset sources can be used for validating multiple practice categories, when appropriate.
- Studies must report sufficient information to be modeled, i.e. providing enough information that model inputs have low uncertainty relative to modeled results, and the model can be appropriately initialized. 'Enough' information to initialize, or reinitialize in the case of dynamic models with state variables that require reinitialization during true-up, and run a model is model- and emissions source-specific. 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 Model Validation Report.

- Studies reporting the effects of changing multiple practices at once ("stacked" practice changes) may be used provided that (1) 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, and (2) each stacked practice study is used only once per Model Validation Report.

- 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 also be used for SOC validation. Newer methods for SOC stock monitoring are becoming available that can observe changes with greater precision at shorter time intervals. New and novel methods for SOC monitoring will be acceptable if there is peer-reviewed support of this practice or independent expert support (approved VVB). 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, but the model must also predict SOC stock change at the corresponding depth. Thus, a fully compiled dataset for evaluating model performance and uncertainty may contain different depths for SOC stock change measurements as long as the model is predicting SOC stock change at each corresponding depth.

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

- Datasets can be drawn from a benchmark database maintained by a third party, if approved by VVB. 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.

- Project developers 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 developers 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. The project developer should report the number of validation data measurements of each data type (SOC, N_2O and CH_4) for each project domain combination of practice category and crop functional group, and include a histogram showing the range of validation data values (e.g. measured SOC change). In the case 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 below) should demonstrate that it addresses the discrepancy.

Requirement 2: Specific Dataset Requirements to Validate Model

The specific 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, at least three declared soil textural classes must be represented, and the range in clay contents must span at least 15 percentage points. When the number of declared soil textural classes is less than three, all textural classes that do occur within the project's geographic extent must be included in the dataset, and there must be a range in clay contents spanning at least 15 percentage points. Once validated, a PC/CFG/ES combination will be approved for crediting within all declared climate zones/nationally defined agricultural land regions and for all declared soil textures.

The purpose of these minimums is to ensure testing for generalized model performance, i.e. that a model is not hyper-calibrated for a specific combination of factors that leads to poor model performance in other contexts. It is in a project'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, 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). If the available data fail to meet one of these minimums due to data scarcity, or fails while also exceeding the others in a way that supports a demonstrable test of generalized model performance, a case may be made for a valid exception to Requirement 2. For example, a case could be made if only two of three declared climate zones are included in the validation dataset because no data could be found, but five or more soil types are included (as opposed to three), and the furthest geographic extent between experimental sites is at least 500 km. Or, if only two of three declared soil types are included because no data could be obtained for the third, but five or more different soil types are included, with a span in clay content ≥30%. Any such cases should be addressed explicitly in the Model Validation Report and will need to be approved by the independent expert 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). If model parameter sets vary by climate zone/nationally defined agricultural land region this may require additional measurement datasets beyond the minimum described above to ensure all parameter sets are validated.

Special Rules for Practice Categories

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

If grazing practices have been validated on pasture, and a CFG has been validated for either the Cropping or Soil Disturbance practice categories, the model can be considered validated for grazing on residue for that CFG. For grazing practices, pasture can 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 can be considered an extra practice category eligible for crediting due to its effects in reducing methane emissions. Validation of the inorganic sulfur fertilizer application practice category would be analogous to the inorganic nitrogen fertilizer application practice category, and would encompass the same domain of practice effects to be used in validation (i.e. magnitude, form, timing, or method for nitrogen 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.

Box 3. Summary of requirements described in Section 5.2.3

Required for the Model Validation Report

- 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 CFG/PC/ES combo, for each emissions source. 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, or 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 Experimental location (only when split between calibration and validation)

Required Upon Request of the Verification Team

- 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
    - Mathematical transformations performed on measurement data
- Study-specific use of data to initialize and run the model, as well as a record for the filling of missing information using process described in Model Validation Report

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, N2O, or CH4 pools for the project domain defined in Section 5.2.2, using measured data that meet the requirements of Section 5.2.3. This is done using the calculation of bias, a simplified version of average relative error (FAO, 2019), 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 model underestimates the practice effect.

The calculation of bias is defined as:

Equation 1

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

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 should be 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). Pooled measurement uncertainty (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 5.3). Because not all studies will report measurement standard error, PMU may be computed using all studies used in a Model Validation Report using the same measurement technique. (Note that this implies studies evaluating different CFGs can be pooled together.) When PMU cannot be obtained, a default replacement value may be used for PMU that is based on typical measurement error for a given measurement technique, per independent third-party expert approval and VVB review.

Equation 2

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

Where

k	=	Number of observations examined across all studies
σ_j	=	Standard error of the j^{th} observed change in SOC, N_2O, or CH_4
n_j	=	number of replicate measurements used in the j^{th} observation

A model is judged as valid if mean model bias is less than PMU, and model predictive uncertainty is determined as described in section 5.2.5. However, 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 where a model may be performing reasonably well even if mean bias is greater than PMU, for example due to limited availability of measured datasets or poor reporting of measured uncertainties. A project developer is allowed to petition for validating the model for use, if it can be clearly justified that the model is showing reasonable overall performance given available measured data. Such a petition will need to be approved by an independent third-party expert and checked by the VVB, following this expert review.

In this model evaluation framework, large model biases result in large residuals. Following guidance for model predictive 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. High model prediction error will therefore be yielded in two circumstances- 1) through low precision of an accurate model or 2) high precision of an inaccurate model. Figure 5.3 and Figure 5.4 walk through an illustrative process to meet the requirements described above.

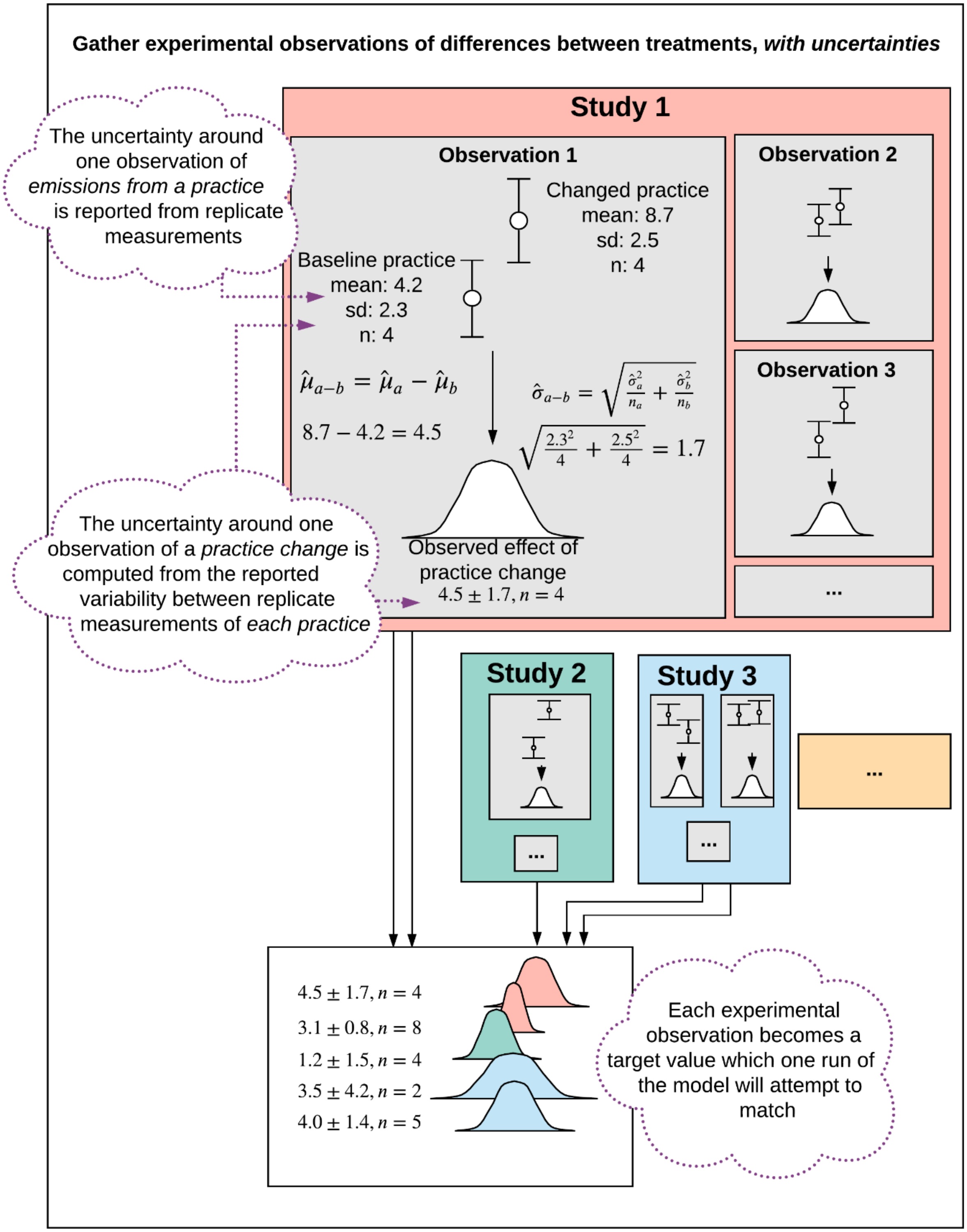

Figure 5.3: Visual summary of one possible approach to calculations for determining measurement uncertainty of an observed practice change effect.

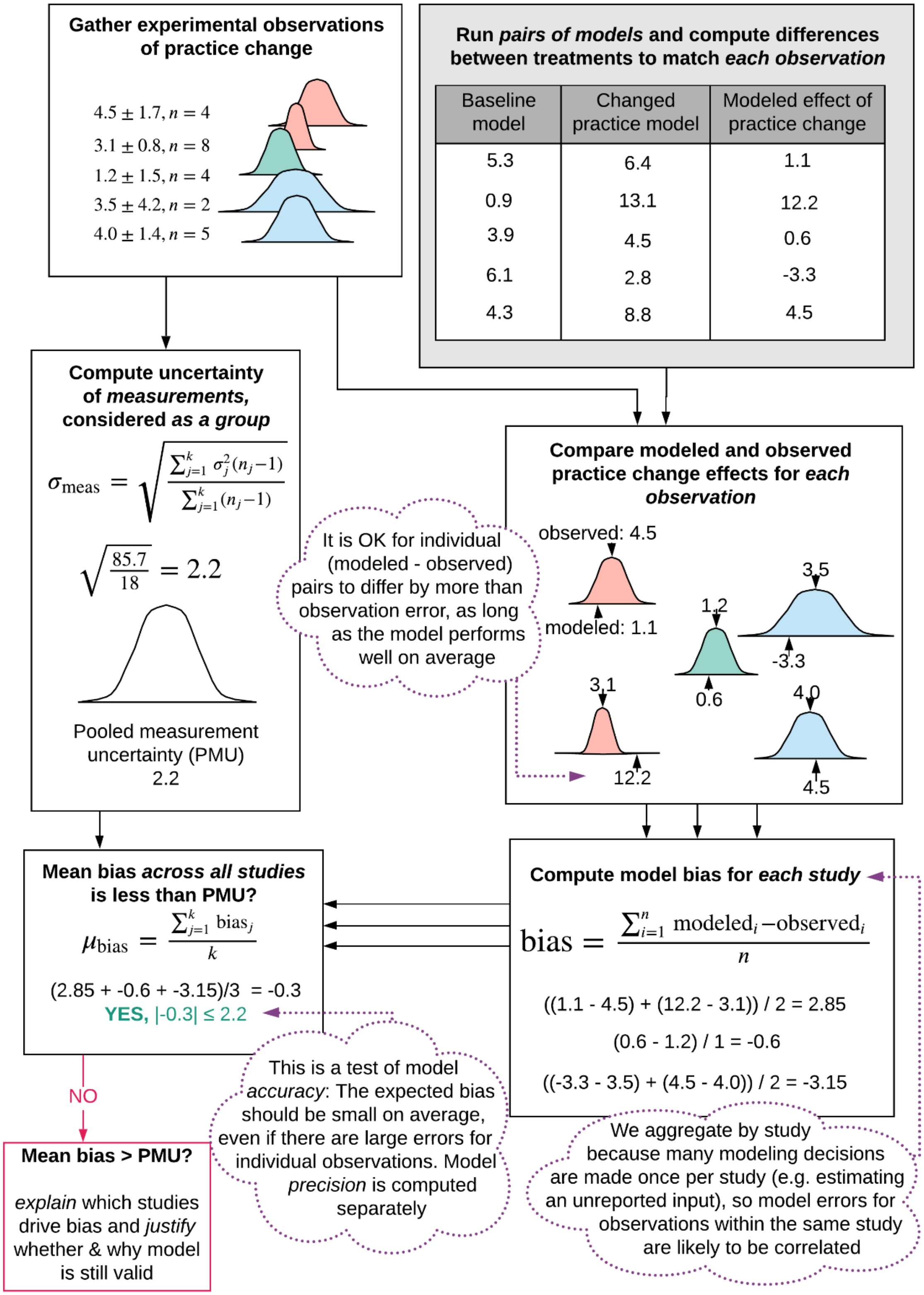

Figure 5.4: Visual summary of calculations for demonstrating that model bias is on a similar scale as measurement error

Box 4. Summary of requirements described in Section 5.2.4

Required for the Model Validation Report

- One complete example derivation of:
  - Calculation of model bias for a study, per Figure 5.3.
  - Calculation of PMU for a single measurement technique, per Figure 5.4.
- 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's validation dataset, ranked highest to lowest
  Average bias across all studies in a PC/CFG/ES's validation dataset.

Required Upon Request of the Verification Team

- 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

In order to evaluate the model for performance, the same datasets should be used to estimate the uncertainty of a model's 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 where there are few validation data (e.g., by using a weakly informative prior if using a Bayesian framework, Fig 5.1.B) as well as account for data variability (i.e. with a wider posterior when data are more variable if using a Bayesian framework). These features enable the model to adequately estimate the confidence in its predictions, as described next.

In the Model Validation Report, as a check that model uncertainty bounds have been appropriately set, measured versus modeled results should be compared for each PC/CFGES combination for changes in SOC, N_2O, and CH_4 (if relevant), and demonstrate a minimum confidence coverage of 90% for 90% prediction intervals (i.e., the 90% prediction intervals should contain the measured value for at least 90% of the validation data). It should be recognized that there may be circumstances where model uncertainty bounds are appropriately set even if 90 % conference coverage is not achieved, for example due to limited availability of measured datasets. A project developer is allowed to petition for validating the model for use with such error bounds, if it can be clearly justified that the model prediction error is appropriately set given available measured data (for example, error bounds that cover 6 out of 7 observations, or 7 out of 8 observations where missing one drops the confidence coverage below 90%). Such a petition will need to be approved by the independent third-party expert and reviewed by the VVB.

In the Model Validation Report, the following should be also 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.

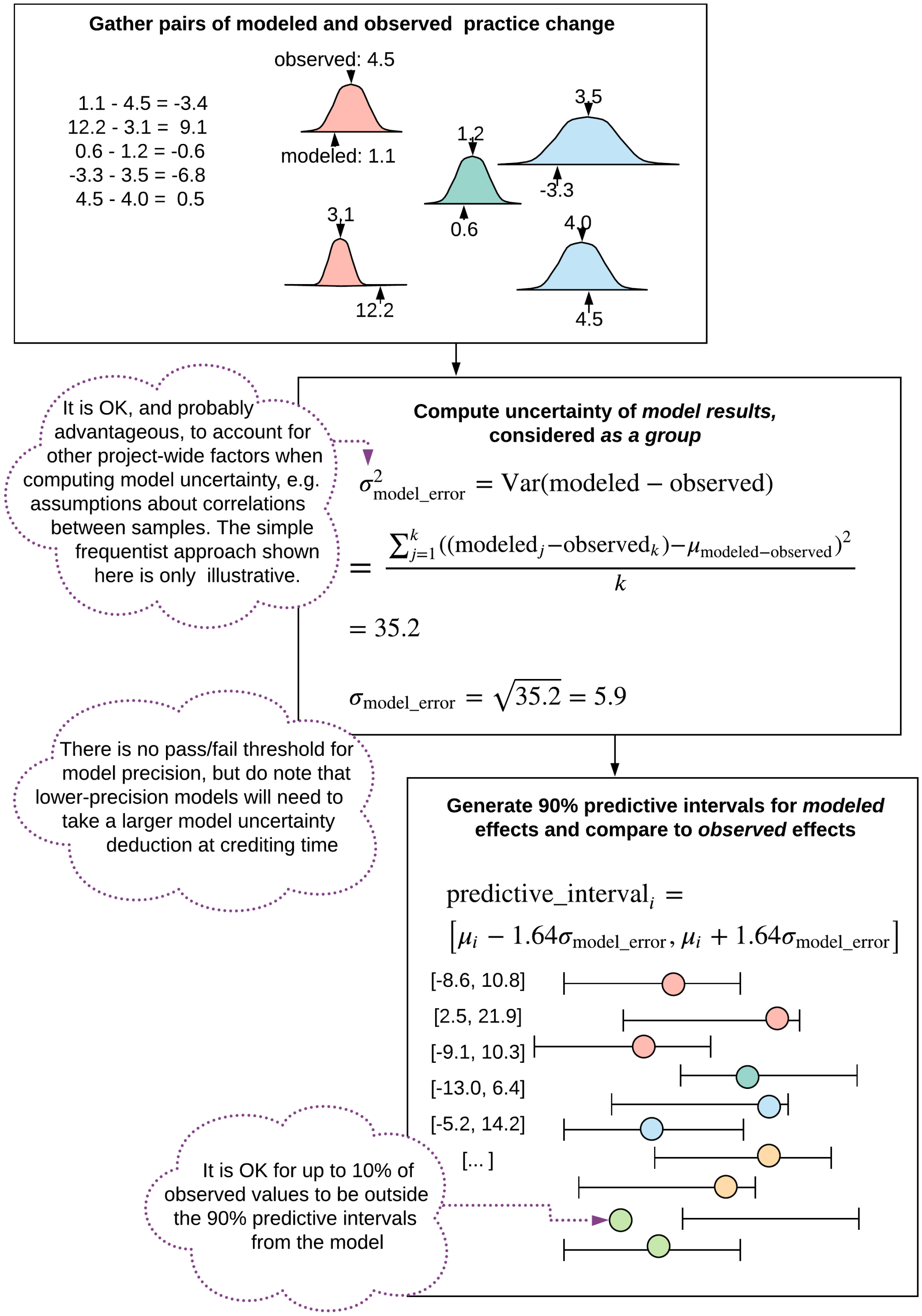

Figure 5.5 Illustrative example of one possible approach to computing model prediction error and demonstrating that model predictions are consistent with validation data.

Box 5. Summary of requirements described in Section Error! Reference source not found.

Required for the Model Validation Report

- For each PC/CFG combination and emissions source:
  - Graphs of measured versus modeled results demonstrating that the 90% prediction intervals contain the measured value at least 90% of the time, per Figure 5.5.
  - Scatterplot of the 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

A Model Validation Report following the above requirements and guidance must be submitted with each monitoring report. Model validation requirements must be satisfied and confirmed prior to the completion of verification activities. The Model Validation Report must be project specific, including demonstration of model validation for the project's domain and combinations of crop functional groups, practice categories, and emissions sources.

Model Validation Reports must be:

1. Independently assessed by a third-party expert service provider hired by the project developer and approved by Verra; or

2. Accepted for publication in one of the peer-reviewed journals listed below and reviewed by an independent third-party expert service provider hired by the project developer and approved by Verra. Where the peer-reviewed publication option is pursued the following also applies:

    a) It is acceptable that the paper has not yet been printed as long as it 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. In this circumstance, the project should submit the peer-reviewed publication and responses to all revisions that clearly demonstrate revisions do not impact model validation

    b) It is acceptable that model validation is 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). The model validation must demonstrate the model was found acceptable for use by the peer reviewers for a give biophysical domain and a given set of practices.

    c) Additionally, the biophysical domain and practices used in the publication must be shown to completely meet the same domain requirements laid out in Sections 5.5.2 and 5.2.3, as well as cover the practice categories and crop functional groups identified in Section 5.2.1.

d) The same datasets used in the peer-reviewed model validation should be used to calculate model prediction error used 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 the 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) Lastly, 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 Model Validation Report.

VVBs will be required to confirm the requirements of the *Model Calibration, Validation and Uncertainty Guidance for the Methodology for Improved Agricultural Land Management* are met. VVBs are not expected to review and confirm the successful validation of the model, as this step already requires independent review and assessment through third-party independent expert review, separate from the project verification. Rather, the VVB must assess and confirm whether the validation report referenced for use of models during the reporting period is, in fact, appropriate to the project domain. This includes assessing appropriate coverage of crop types, practices, and climate zones, for example. Each VVB must demonstrate, to Verra's satisfaction, that they include a team member in each given reporting period that is sufficiently knowledgeable regarding the use of the particular model used to quantify emission reductions and removals in that reporting period.

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 Model Validation Report can be used. If the project is expanded to new practice categories, new crop functional groups, new emissions sources, or the model is changed, the Model Validation Report needs to be:

- Revised, reviewed by an independent third-party expert and re-submitted, or
- Submitted and accepted for publication as a new journal article in one of the peer-reviewed journals listed below and reviewed by an independent third-party expert.

In both cases, the Model Validation Report or peer-reviewed publication must also be reviewed by an approved VVB to confirm its appropriateness for the project domain. All Model Validation Reports will be public documents.

Following are the pre-approved journals (note that a new journal may be added to the list per the requirements of the latest version of the VCS *Methodology Approval Process*):

- Agricultural and Forest Meteorology

- Agricultural Systems
- Agriculture, Ecosystems and Environment
- Agronomy Journal
- Atmospheric Environment
- Biogeochemistry
- Biogeosciences
- Ecological Applications
- Ecological Modeling
- Ecosystems
- Environmental Modelling and Software
- Environmental Pollution
- Field Crops Research
- Frontiers in Ecology and the Environment
- Geoderma
- Global Biogeochemical Cycles
- Global Change Biology
- Journal of Environmental Quality
- Journal of Geophysical Research - Biogeosciences
- Nutrient Cycling in Agroecosystems
- Plant & Soil
- PLoS ONE
- Science of the Total Environment
- Soil & Tillage Research
- Soil Biology & Biochemistry
- Soil Research
- Soil Science Society of America Journal
- Soil Use & Management
- Vadose Zone Journal

5.3 Substitution for Missing Crop Types

If during the calibration and validation process no sufficient data are available for 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. If an entire CFG is not validated, substitutions may be made that entail specific replacements be made for the baseline and project simulations. This method depends on the availability of alternative, conservative CFGs that meet all of the above criteria and have been validated; without any alternatives, no substitution can be made.

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 would mean emitting *fewer* GHG emissions than the missing crop, and 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 would mean emitting *more* GHG emissions than the missing crop, and should be clearly supported with peer-reviewed literature.

Note also that Quantification Approach 2 (Measure and Remeasure) is an available option in cases where the model has not been validated. However, currently quantification approach 2 cannot be used because a performance benchmark does not exist. Interested stakeholders would be responsible for developing the performance benchmark in accordance with VCS Guidance for Standardized Methods.

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	Equation 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 models used to estimate stock change/emissions that meets applicability conditions of this module may 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	Equation 1
Source of data	See Section 5.2.3 of this module.
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	Equation 1
Source of data	See Section 5.2.3 of this module
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	Equation 1
Source of data	See Section 5.2.3 of this module
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	Equation 2
Source of data	See Section 5.2.3 of this module
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 practice in the *j*th observation
Equations	Equation 2
Source of data	See Section 5.2.3 of this module
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 jth observation
Equations	Equation 2
Source of data	See Section 5.2.3 of this module
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	Equation 1
Source of data	See Section 5.2.3 of this module
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

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.

Climate Action Reserve. Expected adoption: June 10, 2020. Soil Enrichment Protocol. Available at http://www.climateactionreserve.org/how/protocols/soil-enrichment/.

Dietze, M. (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. 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., … Douglas, 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

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.

Mishurov, M., & Kiely, G. (2011). Gap-filling techniques for the annual sums of nitrous oxide fluxes. *Agricultural and Forest Meteorology*, *151*(12), 1763–1767.

Natural Resources Conservation Service. 2014 Farm Bill - Conservation Compliance Crop List. United States Department of Agriculture. Available at https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/programs/farmbill/?cid=stelprdb1262733.

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.

Wallach, D., Makowski, D., Wigington, J., Brun, F. (2014). Working with Dynamic Crop Models: Methods, Tools and Examples for Agriculture and Environment, second ed. Elsevier Science, Oxford, UK.

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. Available at 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.