Model Calibration, Validation, and Verification Guidance

For Soil Enrichment Projects

Draft for Public Comment

April 17, 2020

Expected date of publication: June 10, 2019

Table of Contents

1 Introduction 1
2 Model Validation 1
2.1 Declare Practice Effects Requiring Evaluation 1
2.2 Define the Project Domain 2
2.2.1 Declare Project Crop Types 2
2.2.2 Declare Project Land Resource Regions 2
2.2.3 Declare Project Soils 2
2.3 Gather Validation Data that Meet the Following Requirements 2
2.4 Demonstrate Lack of Bias or Conservative Bias for Each Practice Effect 4
3 Linking Validation Data to Model Prediction Error 4
4 Satisfying the Model Validation Requirements via Peer-Review 4
5 Model Calibration 5
5.1 Guidance on Model Calibration using Frequentist Approaches 5
5.2 Guidance on Bayesian Methods for Calibration, Validation, and Error 5
5.3 Substitution for Missing Crop Types 7
6 Verification of Model Usage 8
7 References 9

1 Introduction

A model must show a lack of bias or conservative bias when being used to estimate soil organic carbon (SOC) stock change and, if applicable to the project, flux change of N_2O and CH_4, to quantify soil enrichment projects. Model validation must be specific to the model being used in the project, as well as how the model is being used to make project quantifications (i.e., for what cropping system and biophysical conditions). Model validation must also make appropriate use of published experimental datasets to compare modeled predictions to real-world change. To facilitate this process, the steps laid out in Section 2 must be followed to evaluate model bias and demonstrate meeting model validation requirements.

A discussion regarding verification requirements for the proper implementation of this biogeochemical modelling guidance can be found in Section 4 of this document, as well as Section 8.3.3.1 of the Soil Enrichment Protocol (SEP).

All stakeholders making use of this guidance should contact the Reserve to ensure they are using the most up to date version of this guidance. Project developers and verifiers must use the version of this document that is in place at the commencement of the reporting period in question. In the cases where multiple reporting periods are being verified at once, project developers and verifiers should seek Reserve guidance and approval as to which version of this guidance should be applied to which reporting period.

2 Model Validation

2.1 Declare Practice Effects Requiring Evaluation

For every practice considered additional within the project, the model must be shown to have an unbiased or conservatively biased representation of the underlying biogeochemical process governing the effect of that practice. To do so, each practice must be binned into the following categories, and the associated practice *effect* evaluated using appropriate experimental data.

Table 2.1. Practice Categories and their Associated Practice Effects Requiring Biogeochemical Performance Evaluation

Practice Category	Practice Effect Requiring Evaluation
Nitrogen fertilizer application	Magnitude, form, or method for nitrogen fertilizer applied, with form encompassing inorganic and organic N fertilizers, and method encompassing surface or subsurface application
Water management/irrigation	Magnitude or timing of irrigation water applied
Tillage and/or residue management	Tillage encompassing soil disturbance, residue management encompassing soil exposure after harvest
Crop planting and harvesting (e.g., crop rotations, cover crops)	Variety of crops grown, which may include cover crops
Grazing practices	Any of the following: presence/absence of grazing, stocking density, forage type or quality

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

2.2 Define the Project Domain

For each practice effect declared in the project description, the model must be evaluated in terms of its bias in estimating emissions reductions. Evaluation of each effect begins with defining the project domain in terms of its biophysical attributes. Specifically, the project developer must declare the unique crop types, land resource regions, and soil attributes associated with each declared practice effect.

2.2.1 Declare Project Crop Types

Crop types for each practice effect must be declared. Crop types can be grouped into bins 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[1]),
- photosynthetic pathway (C3/C4/CAM),
- tree/shrub/herbaceous (trees and shrubs have woody plant growth, versus herbaceous species that do not grow woody plant material),
- flooded/not flooded

2.2.2 Declare Project Land Resource Regions

The full list of land resource regions (LRRs) associated with each practice effect must be declared.[2] A comparable framework for defining resource regions in geographies outside of the U.S. will be provided in a future version of this document.

2.2.3 Declare Project Soils

Soils are to be declared for each practice effect 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.

2.3 Gather Validation Data that Meet the Following Requirements

Requirement 1: Validation datasets for each declared practice effect from Step 1 must include measurements (declared in Requirements 2 and 3) 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. Some hypothetical examples of acceptable experimental treatments to evaluate practice effects are given in the following table:

---

[1] Resource can be found here: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/programs/farmbill/?cid=stelprdb1262733
[2] Resource can be found here: https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/technical/nra/nri/?cid=nrcs143_013721
[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

Table 2.2. Examples of Acceptable Experimental Treatments to be Used in Evaluating Practice Effects

Practice Effect	Experimental treatment
Magnitude, form, or method for nitrogen fertilizer applied, with form encompassing inorganic and organic N fertilizers, and method encompassing surface or subsurface application	Comparison of two different application rates of urea; comparison of inorganic N fertilizer to manure.
Tillage encompassing soil disturbance	Comparison of conventional tillage using moldboard plow to strip tillage.
Variety of crops grown	Comparison of single-crop rotation to double-crop rotation; comparison of no cover-crop to with cover crop.

Validation data must adhere to the following guidelines:

- Measured datasets must be drawn from peer-reviewed and published experimental datasets with measurements of SOC stock change (and annual/seasonal measures of N_2O and CH_4 change if applicable) using control plots to test the practice effect. All dataset sources must be reported. The same measurement dataset sources can be used for validating multiple practice effects, when appropriate.
- In the case of SOC stocks, sources that include repeat measurement through time must span at least a 5-year interim. Similarly, measures of paired fields using space-for-time substitution must approximate at least a 5-year interim.
- In the case of N_2O and CH_4 flux, any combination of measurements from chambers and/or eddy covariance flux towers are acceptable. To conform with typical reporting periods that span at least the seasonal timescale, event-based or sub-seasonal measurements of N_2O and CH_4 must be aggregated to the seasonal or annual timescale for use in validation using appropriate methods (e.g., Mishurov & Kiely, 2011; Turner et al., 2016).
- Datasets can be drawn from a benchmark database maintained by a third party, approved by the Global Soils Partnership (or comparable). 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.
- Project developers are expected to take reasonable steps to identify and use available datasets that meet the above criteria. If appropriate datasets commonly used in peer-reviewed model validation for a practice effect are found to be excluded, they must be added.

Requirement 2: Validating a practice effect can only be completed at the scale of an individual sample/field if there are measurements of SOC stock and annual/seasonal N_2O and CH_4 flux change (if applicable) meeting the above criteria that in total include the same soil textural class, or one that is within 30% of the same clay content, as well as the crop type and soil-climate zone relevant to that location.

Requirement 3: Validating a practice effect for the entire Project Domain can only be completed if there are measurements of SOC stock and annual/seasonal N_2O and CH_4 flux change (if applicable) that in total cover:

- At least three declared LRRs
- At least three declared soil textural classes
- A range in declared clay contents spanning at least 15%
- All declared crop types associated with that practice effect

2.4 Demonstrate Lack of Bias or Conservative Bias for Each Practice Effect

For each practice effect declared in Step 1, the model must be shown to be unbiased or conservatively biased in estimating the change in SOC, N_2O, or CH_4 pools for the project domain defined in Step 2, using measured data that meet the requirements of Step 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 must be calculated for each individual experimental study since different studies may use different temporal units of aggregation, soil depths, or measurement techniques. The calculation of bias is defined as:

$$\boldsymbol{bias} = \sum_{i=1}^{n} \boldsymbol{P_i} - \boldsymbol{O_i}$$

Where,

P_i	=	Predicted (modeled) value of change in SOC, N_2O, or CH_4 with the practice
O_i	=	Observed value of change in SOC, N_2O, or CH_4 with the practice

Bias indicates the average tendency of the modeled estimates to be larger or smaller than their observed counterparts (Moriasi et al., 2007). An unbiased model will have *bias* = 0.0. Positive values indicate model overestimation bias, meaning that the model overestimates the practice effect and thus the credits earned. A negative value indicates model underestimation bias, or an underestimation of the credits earned. Sufficient model validation requires the model to be unbiased or conservatively biased for each study examined, i.e., bias must be shown to be <= 0.0 in all cases.

3 Linking Validation Data to Model Prediction Error

Validation data should be used to estimate the uncertainty of a model's predictions, i.e., the model prediction error (Section D.2). The model prediction error calculation should be shown to penalize fewer data points (i.e., have a higher variance; see Equations D.4, D.5), such that the uncertainty deduction in credits is higher when fewer data are available.

4 Satisfying the Model Validation Requirements via Peer-Review

In lieu of explicitly evaluating bias as described above, successful model validation may be demonstrated through publication in a peer-reviewed journal. In this case, the same model version and model parameter values/parameter set values must be used in the peer-reviewed publication as are used in the project. The publication must demonstrate that separate datasets were used to for model calibration/parameterization and model validation (unless qualifying for a special exception; see Section 5.1). The model validation must demonstrate the model was found acceptable for use by the peer reviewers for a given biophysical domain and a given set of practices. Additionally, the biophysical domain and practices used in the publication must be shown to completely meet the same domain requirements laid out in Steps 2 and 3, as well as cover the practice effects identified in Step 1. The same datasets used in the peer-reviewed model validation should be used to calculate model prediction error used in the project.

5 Model Calibration

5.1 Guidance on Model Calibration using Frequentist Approaches

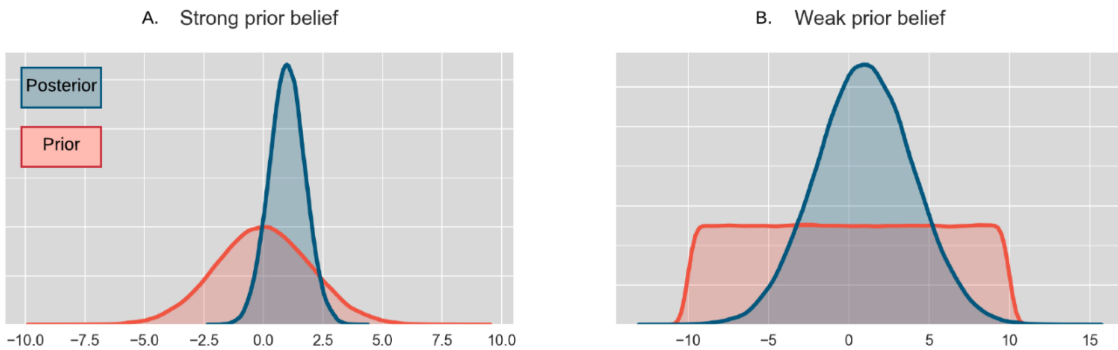
Calibration is the process of estimating model parameters. The protocol does not require or prescribe a model calibration procedure though calibration is encouraged as a way to improve model performance and thus reduce model prediction error.

Because biogeochemical models often contain a large number of parameters, different strategies can be employed to perform calibration. 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.

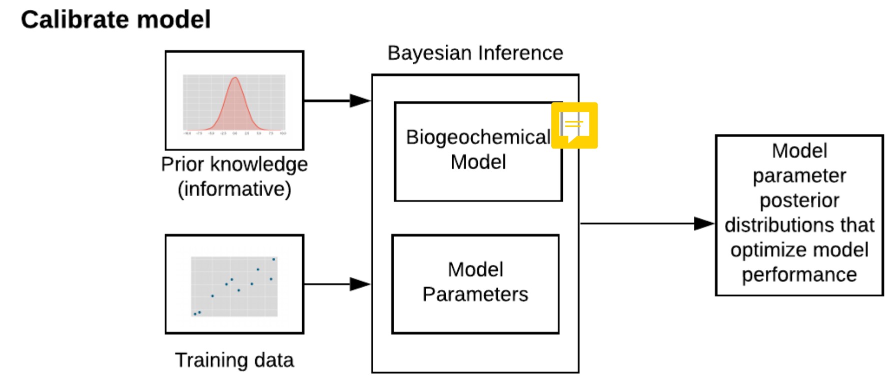
If calibration is performed using experimental datasets, there are no requirements on those datasets. However, data used in validation must be shown to be completely separate from data used in calibration through numerical comparison, e.g., the differences between data points are shown to be nonzero. The one exception to this requirement is when a project developer can demonstrate the proper use of, via publication in a peer reviewed journal, a rigorous calibration and validation process using multiple partitions of the same dataset, e.g., using k folding or similar.

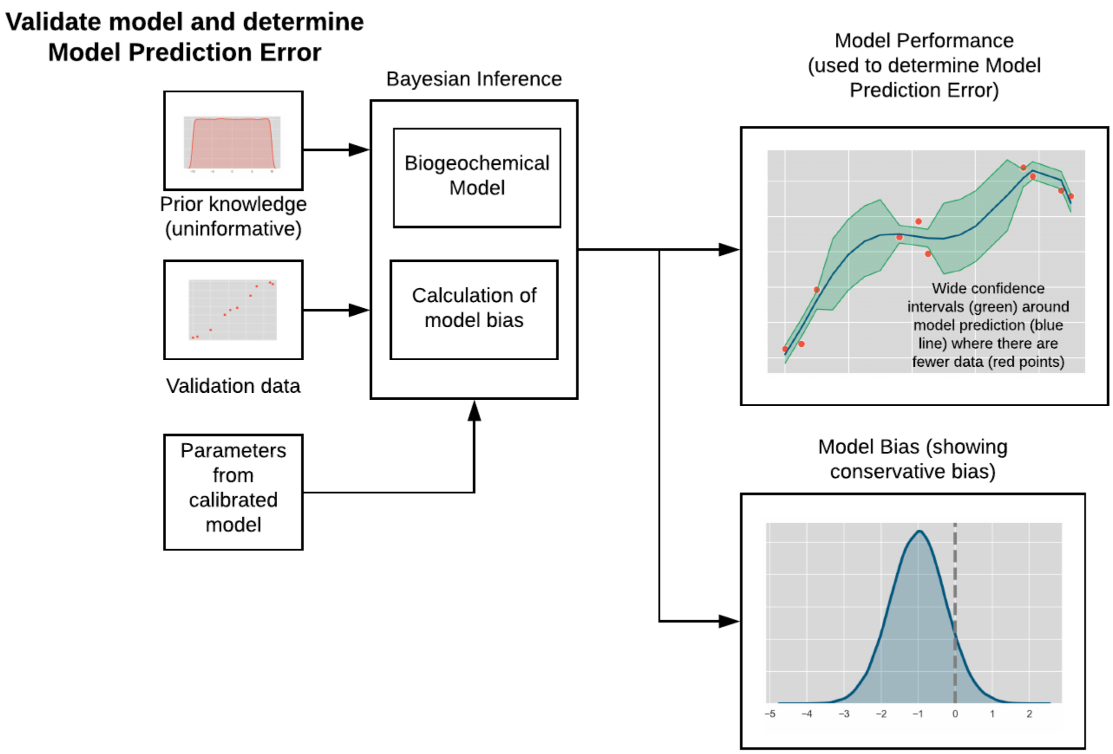
5.2 Guidance on Bayesian Methods for Calibration, Validation, and Error

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).



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 its central role in the developing field of Ecological Forecasting (Dietz, 2017), as well as in the Predictive Ecosystem Analyzer Project data-model integration system (pecanproject.org). It is possible to use Bayesian approaches for model validation and model prediction error, as well, as long as the datasets used for each process are kept separate (with the exception of cross-validation where partitions are made of the same dataset using k folding or similar, see Section 5.1). This type of framework is encouraged for model validation and model prediction error, as the confidence intervals around model predictions will be directly based on the availability of observed data. The following figure presents a conceptual workflow for a Bayesian approach to these analyses.





5.3 Substitution for Missing Crop Types

If during the calibration and validation process no sufficient data are available for a declared crop type, a substitution may be made that entails specific replacements be made for the baseline and with-project simulations. This method depends on the availability of alternative crop types for a given practice effect that meet all of the above criteria; without any alternatives no substitution can be made.

- Baseline:
  - Replace the missing crop type with an unfertilized perennial grass;
- Project:
  - Replace the missing crop type with an alternative crop type for which data are available that best matches the missing crop in terms of its attributes, i.e., N fixation, annual/perennial, photosynthetic pathway, plant form, and flooded/not flooded status. The acceptable alternative crop will have the most matching categories among all crop types available.
  - For multiple alternative crop types having the same number of matching attributes, the crop type that best accommodates the management practices of the missing crop should be selected.

6 Verification of Model Usage

Each verification team must include a person or persons who are expert in the particular biogeochemical model used to quantify emission reductions in that reporting period (if any). Guidance is provided in Section 2 for requirements that models must meet, in order to be eligible. Verifiers will be required to confirm the requirements of Section 2 of this document are met.

Expert guidance is needed to ensure the given biogeochemical model is appropriately validated, parameterized, and calibrated for each reporting period. If the project employs the use of a third-party expert to undertake validation, parameterization, calibration, and/or running a biogeochemical model in a given reporting period, then there will be no need for the verification team to independently verify such activities have been done appropriately, provided the verification team: confirms that the use of such third-party has been approved by the Reserve, that the party in question has the requisite expertise, that all requisite steps as set out in Section 2 of this document have been followed, and provided the expert provides the verification team with a sensitivity analysis regarding the requisite data inputs for the given model.

In other words, the verifier is simply required to confirm approval from the Reserve, confirm the qualification of the third-party, and confirm the requisite validation steps have been followed, but the verifier does not independently need to run the model themselves to confirm results appear reasonable. The verification team will still be required to confirm the reasonableness of all data input into the given biogeochemical model, following the requirements for baseline modelling in Section 3.4.1.1 of the SEP, and following expert guidance on the sensitivity of the given model to the requisite data inputs.

7 References

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. (2006). 2006 IPCC Guidelines for National Greenhouse Gas Inventories. Available at https://www.ipcc-nggip.iges.or.jp/public/2006gl/.

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.