Validation Report for DayCent-CR Version 1.1.0

Indigo Ag

Reviewed by:	Dr. Ankur Desai University of Wisconsin, Madison
First submission:	12/21/2023
Revised to address reviewer comments:	03/13/2024
Model requirements version:	Requirements and Guidance for Model Calibration, Validation, Uncertainty, and Verification for Soil Enrichment Projects, Version 1.1a
Prepared by:	Indigo Ag
Contact:	500 Rutherford Ave. Boston, MA 02129 +1 (844) 828-0240

Contents

To access Appendix A for academic research purposes, Indigo Ag (rpape@indigoag.com) should be contacted directly.

Figures

Tables

Acronyms and Abbreviations

API	application programming interface
C	carbon
C3A	C3 annual crop functional group
C3AN	C3 annual N-fixing crop functional group
C3PN	C3 perennial N-fixing crop functional group
C3S	C3 annual shrub crop functional group
C4A	C4 annual crop functional group
CAR	Climate Action Reserve
CFG	crop functional group
CH_4	methane
CO_2	carbon dioxide
CRNF	controlled release nitrogen fertilizer
CROP	cropping practices, planting and harvesting practice category
CTIC	Conservation Technology Information Center
CV	coefficient of variance
DISTURB	soil disturbance and/or residue management practice category
DREAM	DiffeRential Evolution Adaptive Metropolis
EENF	enhanced efficiency nitrogen fertilizer
ERS-ARMS	Economic Research Service Agricultural Resource Management Survey
GHG	greenhouse gas
GSA	global sensitivity analysis
IPCC	Intergovernmental Panel on Climate Change
LRR	Land Resource Region
MCMC	Markov Chain Monte Carlo
Model Requirements	Requirements and Guidance for Model Calibration, Validation, Uncertainty, and Verification for Soil Enrichment Projects, Version 1.1a, accessed on 17 Oct 2023
N	mineral nitrogen
N_2	elemental nitrogen
N_2O	nitrous oxide

NASS	National Agricultural Statistics Service
NFERT	inorganic nitrogen fertilizer application practice category
NI	nitrification inhibitors
NO_x	nitrogen oxides
OAT	One At a Time
ORG	organic amendments application practice category
PC	practice category
PMU	pooled measurement uncertainty
REML	restricted maximum likelihood
RMSE	root mean square error
SEP	Soil Enrichment Protocol
SOC	soil organic carbon
SD	standard deviation
SE	standard error
SWAT	Soil and Water Assessment Tool
USDA	United States Department of Agriculture

1 Report type

1.1 Report type

Type 1 (Project-specific, for project CAR1459, Indigo U.S. Project No. 1 [CAR1459])

1.2 Climate Action Reserve Soil Enrichment Protocol version

Version 1.1, accessed on 17 October 2023

1.3 Climate Action Reserve Soil Enrichment Protocol model requirements version

Requirements and Guidance for Model Calibration, Validation, Uncertainty, and Verification for Soil Enrichment Projects, Version 1.1a, accessed on 17 Oct 2023 (referred to hereafter as the "Model Requirements")

1.4 Model version

DayCent-CR Version 1.1.0.

This model version consists of the following components (collectively the model files). Each of these components are version-controlled independently from each other, but only the following component versions shall be considered the validated DayCent-CR Version 1.1.0:

1. Version 1.1.0 of the DayCent-CR model executable, compiled from revision 32aa16b of Indigo Ag's DayCent code repository.
2. Version 3.0 of the DayCent-CR model parameters, with fixed components from revision a7f2701 of Indigo Ag's DayCent validation data repository and probability distributions of calibrated components from revision 8326da36d5of Indigo Ag's cross-validation pipeline.

The code lineage was originally derived from the branch of DayCent maintained by the National Greenhouse Gas Inventory team and also used for the COMET-Farm system and has been modified since the previous validation report as described in Section 1.6.

During model simulations for project CAR1459, Indigo submits inputs to the model using the DayCent-CR application programming interface (API). This API was not used during calibration or validation and is not included in the model version described here. The validation described here should be applicable to any result obtained from DayCent-CR Version 1.1.0 whether it is run directly or accessed through any technically compatible version of the DayCent-CR API.

1.5 Version confirmation materials

The materials included in Table 1 have been provided for use by the reviewer of this report and project verifiers for CAR1459. All materials are version-tracked in their own repository separate from the model API, which may have independent version updates that do not change the validated model filesTable 1.

Table 1. Materials provided for review.
To access Appendices and Validation Supporting Files for academic research purposes, Indigo Ag should be contacted directly.

Material	Description
Supporting Materials 1	Documentation of validation and calibration datasets for each Crop Functional Group x Practice Change x Emission Source combination
Supporting Materials 2	Analysis of model performance for each Crop Functional Group x Practice Change x Emission Source combination
Appendix A	Documentation of calibrated parameter sets
Appendix B	Declaration of validated practice changes
Appendix C	MCMC sampler diagnostics
Appendix D	Comparison of thinned vs full posteriors of final calibration
Appendix E	Confidence interval width and coverage rates as a function of time
Appendix F	Proposal for disambiguating pooled measurement uncertainty (PMU)
DayCentCR_1.1.0_validation_supporting_files.zip	Copies of validation datasets and model run files used in the simulations for this report (in their initial state prior to model calibration), as well as code for running calibration and analyzing. See README.

1.6 Changes from previous model version

DayCent-CR versions 1.0 and 1.0.2 have been previously validated and approved for crediting of soil organic carbon (SOC) in CAR1459 (Indigo Ag, 2022). For the current validation of DayCent-CR Version 1.1.0, the following changes from version 1.0.2 have been made:

1. New algorithms have been added to simulate the effects of nitrification inhibitors (NIs) and controlled release nitrogen fertilizers (CRNFs), collectively known as enhanced efficiency nitrogen fertilizers (EENFs). Further information is provided in Section 2.
2. Daily modeled nitrous oxide (N_2O) emissions were capped to curtail the tendency of the existing denitrification algorithm to simulate unrealistically high short-term fluxes. Further information is provided in Section 2.
3. Simulation of nitrous oxide N_2O is calibrated and validated. See Section 3 for more information.
4. New sites have been added to the validation dataset, further expanding the domain of geographies, crop types, and practices covered by this validation. In particular, the new data include enough observations to validate select practice categories (PCs) for SOC changes in the C3 perennial N-fixing crop functional group (CFG) and select PCs for various CFGs for N_2O. See Section 4 for details.
5. Parameters selected by global sensitivity analysis for SOC did not change from the previous report, here one at a time sensitivity analysis was used to select additional parameters for joint calibration with SOC pool and N_2O flux observations as described in Section 3.
6. While using the same algorithms for calibration, the majority of implementation has been migrated from R to Python. In particular, in the previous report we used the implementation of the DREAM algorithm in the dream package for R (Vrugt et al., 2009), but in the present report we used the implementation of the DREAM algorithm in the SPOTPY packge for Python (Houska et al., 2015).
7. In the previous validation report (i.e., Indigo Ag, 2022) pooled measurement uncertainty (PMU) was calculated for each PC x CFG x Emission Source category; in this report, we calculated a 'global' PMU for each Emission Source category to reduce sampling error as described in Section 7.

1.7 Project team

Calibration, validation, and running of DayCent-CR for this project were performed by Indigo Ag, which is also the project developer of CAR1459. Two external collaborators assisted Indigo staff with interpretation of the literature and development of input files: Stephen Williams (retired) and Dr. Yao Zhang (Colorado State University). As required in Section 5 of the Model Requirements, Indigo Ag has the requisite expertise to calibrate and validate DayCent-CR for model performance and uncertainty. A summary of the project team's modeling qualifications is provided in Table 2.

Table 2. Project team modeling qualifications.

Team member	Degree	Modeling expertise (years of experience)		
		DayCent	Environmental mechanistic modeling	Statistical modeling
Chris Black	Ph.D.	11	14	11
Hamze Dokoohaki	Ph.D.	<1	12	12
Jeff Kent	Ph.D.	14	14	
David LeBauer	Ph.D.	<1	14	20
Michelle Schmidt	M.S.	<1	10	10
Brian Segal	Ph.D.	2	2	14
Stephen Williams (contractor)	M.S.	11	24	
Yao Zhang (Colorado State University)	Ph.D.	13	10	8

2 Model description

This report describes the validation of DayCent-CR version 1.1.0 for use in modeling changes in soil carbon and N_2O for crediting as part of CAR1459.

DayCent-CR is a process-based ecosystem biogeochemical model that simulates carbon and nitrogen dynamics in cropland and grassland systems and has been tailored for compliance with the requirements of the Climate Action Reserve Soil Enrichment Protocol (CAR SEP). The DayCent model (e.g., see Parton et al., 2001; Del Grosso et al., 2006; Del Grosso et al., 2012; Zhang et al., 2018) has been used extensively for more than two decades by researchers worldwide to simulate soil organic matter dynamics and soil trace gas N_2O and methane (CH_4) fluxes in a variety of managed ecosystems (cropland, grassland, savanna, forest). The model employs a daily time step and simulates plant processes (e.g., photosynthesis, phenology, dry matter allocation, senescence), soil water balance, soil temperature, soil organic matter dynamics for two plant litter and three soil organic matter pools, as well as mineral nitrogen transformations including elemental nitrogen (N_2), N_2O and nitrogen oxides (NO_x) emissions and CH_4 oxidation and emissions from soil. The model is used to estimate net carbon dioxide (CO_2), N_2O, and CH_4 emissions from soils in the U.S. national greenhouse gas (GHG) inventory submitted by the U.S. Environmental Protection Agency to the United Nations Framework Convention on Climate Change. The DayCent model is included within the COMET-Farm platform that implements United States Department of Agriculture's (USDA) entity-scale GHG inventory methods (Powers et al., 2014) and the model is implemented as part of the CAR's protocol for avoided conversion of grassland (http://www.climateactionreserve.org/how/protocols/grassland/).

The version of the model validated in this report is based on the latest version of the model developed to simulate soil biogeochemistry to 30-centimeter (cm) soil depth, with additional improvements to several soil and plant processes as documented in Gurung et al. (2020) and Mathers et al. (2023). The current version of the model (version 1.1.0) is structurally the same as documented in Mathers et al. (2023), with the following exceptions:

- New algorithms were added to simulate two types of EENFs: NIs and CRNFs. The effects of these products on N_2O emissions from croplands are quantified in a meta-analysis by Thapa et al. (2016).
    - NIs are a class of EENF products that are blended with nitrogen fertilizers and act in the soil to reduce rates of nitrification, resulting in reduced N_2O losses. The new algorithm represents these effects using a daily reduction factor on gross nitrification which is set at a value between 0 and 1 on the date of fertilizer application and increases toward 1 (and thus no effect) as a function of cumulative exposure to heat and moisture. The mathematical specification used is presented in Gurung et al. (2021).
    - CRNFs are a class of EENF products which typically use a semi-permeable coating to impede diffusion of nitrogen into the soil, resulting in a gradual release that is better matched to crop demand and reduced losses as N_2O. The new algorithm represents CRNF products by adding only small fractions of the total applied fertilizer to the simulated nitrogen pools, with the speed of release driven by cumulative exposure to heat and moisture. The mathematical specification used is presented in Gurung et al. (2021).
- Daily N_2O fluxes from denitrification were capped to maintain realistic rates of N_2O emission under peak conditions. While N_2O measurements commonly show brief periods of very high emissions, the empirical approximations used in DayCent's trace gas submodel occasionally predict fluxes that are not supported by measurements. Testing a simple cap on extreme simulated

pulse emissions resulted in reduced bias and root-mean-square error of N_2O predictions. While not published, a maximum daily flux was developed based on the 99th percentile of over seventy thousand GraceNet N_2O measurements representing 151 treatments across 18 sites (USDA, 2015). This cap is applied to the daily flux from denitrification, which represents most of the N_2O flux simulated by DayCent. Figure 1 shows total N_2O fluxes simulated with and without the daily cap active in the model, using default (uncalibrated) values for model parameters.

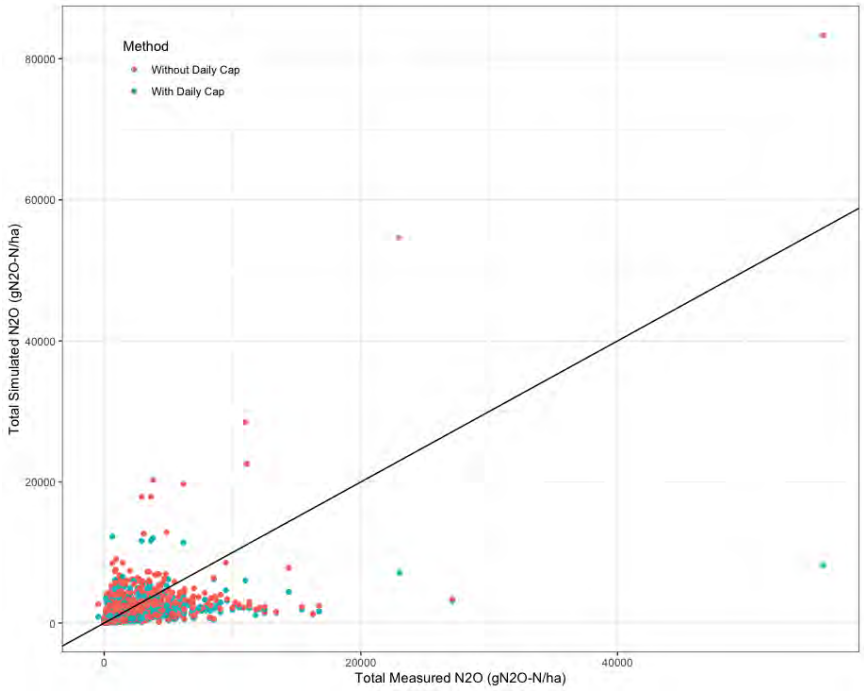

Figure 1. Total simulated N_2O fluxes from use of the capped (blue dots) and uncapped (red dots) model versions.

The model has been adapted to use initial estimates of SOC based on lab measurements of field sampled soils (see Section 3.2 for details) and soil organic nitrogen pools based on the modeled C:N ratios of each SOC pool as described in Mathers et al. (2023). This model is referred to as "DayCent-CR". This allows the model to operate in compliance with the CAR SEP Section 5, using the required direct measurements of SOC to initiate with-project and baseline simulations. In addition, the parameterization and validation of the model, using Bayesian techniques described herein, has been tailored specifically to the cropping domains defined in this report, and post-simulation quality checks have been implemented to ensure model applicability across axes not considered by the validation domain (see Section 11).

For clarity, here is an overview of key carbon and nitrogen quantities and their use in this report and in crediting.

1. **SOC Pool Size** or **SOC Stock** is the measurement of the mass of carbon in the top 30 cm of soil. This quantity is used to initialize the model and, with repeated measurements, to calculate its change over time.
2. **SOC Temporal Stock Change** is the change in SOC Stock over time. This quantity is used in model calibration.

3. **SOC Temporal Stock Treatment Difference** refers to the difference in SOC Temporal Stock Change between a practice change treatment and its baseline control. This quantity is used in validation.
4. **Soil Organic Nitrogen** is analogous to the SOC Pool Size and is measured and modeled but not directly used in calibration, validation like SOC. It is only used as a check on model performance, and to exclude sites with extreme values of this quantity.
5. **Seasonal N_2O Flux** is the interpolated and integrated sum of instantaneous N_2O fluxes over time, typically a growing season. This quantity is used alongside SOC Temporal Stock Change for calibration.
6. **N_2O Flux Treatment Difference** is the difference in N_2O flux between the treatment and baseline control scenario and is used alongside SOC Temporal Stock Change Treatment Difference in validation.
7. **Emissions Reduction** is quantity 3 and / or 6 (treatment differences) converted to equivalent reduction in atmospheric CO_2 (CO_2e); the sum of these can be used for credit generation.

In summary, the model is initialized using SOC Pool Size, calibrated to the change in SOC Pool Size over time and N_2O seasonal flux, and validated against the differences in these quantities between treatment and baseline control scenarios. For crediting, these impacts are converted to CO_2e.

3 Model calibration

Follows Model Requirements Section 2

3.1 Description of model calibration

DayCent-CR Version 1.1.0 was calibrated using an approach that is similar to empirical Bayes in some respects; our approach is not fully Bayesian due to the way the variance parameters are estimated. The model was calibrated using a joint Bayesian approach by simultaneously estimating the joint posterior parameters for SOC stock change and N_2O flux (Appendix A). The joint posterior of DayCent parameters was estimated using the DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt et al., 2011; Vrugt, 2016), which is a Markov Chain Monte Carlo (MCMC) algorithm. For each new set of parameter values proposed by the DREAM algorithm, normalized sum loglikelihood for SOC temporal stock change across all SOC sites and normalized sum loglikelihood for N2O flux across all N2O sites was estimated separately, and the total sum log likelihood was passed back to the optimizer enabling the simultaneous calibration of these two emission sources. The DREAM algorithm has been used by Zhang et al. (2020) to calibrate DayCent for crop growth/production. To calibrate DayCent-CR for modeling SOC temporal stock change, the likelihood function was used as presented in Mathers et al. (2023) to allow for heterogeneous residual variance. For modeling N_2O flux and flux change on a per-growing-season basis (see Section 6.3), homogeneous residual variance was assumed based on a scatterplot of residuals versus measurement interval. For both SOC and N_2O the likelihood function accounts for site (location) and year effects and estimates model error for predictions at new sites and is therefore suitable for the type of dataset used in this report i.e., data compiled from multiple experimental sites with repeated measurements that are correlated both in space and time.

In brief, the likelihood function assumes that the error follows a zero mean multivariate Gaussian distribution per Equation 1:

$$p(y_{\text{obs}} \mid \theta) = (2\,\pi)^{-n/2} \mid \Sigma \mid^{-1/2} \exp\left\{ -\frac{1}{2}(\hat{y}_{\text{mod}} - y_{\text{obs}})^{\intercal}\, \Sigma^{-1}\, (\hat{y}_{\text{mod}} - y_{\text{obs}}) \right\}$$

(Equation 1)

where θ is a vector of parameters that are used by DayCent to predict SOC and N_2O or that define the variance-covariance matrix Σ, $\hat{y}_{\text{mod}}$ and y_{obs} are natural log transformations of SOC and N_2O values (modeled and observed, respectively), and n is the number of observations. Both $\hat{y}_{\text{mod}}$ and Σ are functions of the parameters θ. The log-based transformation was defined as follows to allow including negative N_2O flux measurements in calibration:

$$Y_{transformed} = \log(1 - Y_{min} + Y)$$

(Equation 2)

where Y is the measured and modeled value of SOC and N_2O and Y_{min} was set to the minimum observation found in the dataset.

The variance-covariance matrix partitions model error into three components: variance between experimental sites σ^2_{site}, variance between years within sites $\sigma^2_{\text{site-year}}$, and unexplained residual variance σ^2_{resid}. When calibrating SOC stock, σ^2_{resid} was modeled as an exponential function of time:

$$\sigma^2_{\text{resid}} = \begin{cases} \sigma^2 \exp(2t\nu) & for\ SOC \\ \sigma^2 & for\ N_2 0 \end{cases}$$

(Equation 3)

where *t* is the number of years since the first measurement (the time at which SOC is reinitialized in DayCent). σ^2_{site}, $\sigma^2_{\text{site-year}}$, σ^2, and ν from Equation 3 are included in θ. These parameters were estimated by fitting the model residuals from each MCMC iteration using a linear random effect model with two levels of random effects (random intercept for site and random intercept for year nested within site) (Pinheiro et al., 2000) and in the case of SOC, an exponential residual variance model that is a function of years since the first measurement when modeling the residuals for SOC. See the supplement to Gurung et al. (2020) for additional discussion of this approach. To evaluate the likelihood for each MCMC iteration, SOC residual models were fit with the R package nlme (Pinheiro et al., 2022) using the lme function, while N_2O with homogenous variance model was fit with the lmer function from the lme4 package. The variance parameters were estimated via restricted maximum likelihood (REML) applied to the marginal model after adding in the Monte Carlo draws for the DayCent calibration parameters. This estimation procedure is similar to empirical Bayes (see Casella, 1985 and Carlin et al., 2009). However, in empirical Bayes, prediction and inference would be based on a single set of variance parameter estimates, whereas here prediction and inference are based on a distribution of variance parameter estimates. As a result, our approach is expected to capture more variability in the variance parameters than a traditional empirical Bayes analysis, but still less variability than a fully Bayesian analysis.

The exponential residual variance model (Equation 3) was chosen for SOC stock changes because it was straightforward to interpret and implement (it is one of the standard variance structures supported by the nlme package (Pinheiro et al., 2022)) and it performed well in practice (see Section 9 and Section 10). Note that at time 0 ($t = 0$), the residual variance becomes $\sigma^2_{\text{resid}} = \sigma^2 \exp(2t\nu) = \sigma^2 e\exp(0) = \sigma^2$. In other words, σ^2 is the residual variance at time zero. As shown in Appendix A, both σ and ν were estimated to be positive, so the residual variance never goes to zero, and increases as the time since SOC reinitialization increases. While exponential variance does not asymptote at long timescales in the way expected for true SOC dynamics, the increase in residual variance for the fitted model is modest over the time period in which the model will be deployed (per the CAR SEP, fields can generate credits for a maximum of 30 years). Please see Appendix E for diagnostics related to the impact of the exponential residual variance model on confidence interval width and coverage rates.

The calibration of DayCent-CR was implemented in Python version 3.9 (Python Software Foundation) using the SPOTPY library (Houska et al., 2015). The DREAM algorithm is described in detail by Vrugt et al. (2009). The calibration was run with 9 MCMC chains, all of which were run until the $\widehat{R}$ statistic of Gelman and Rubin (1992) dropped below 1.2 as recommended by Vrugt et al. (2009), suggesting convergence of the posterior distribution of model parameters. The first 50 percent of each chain was discarded as the "burn-in" period and the remaining 50 percent of each chain was used to summarize the posterior of the parameters θ. The post-burn-in simulations were thinned further to reduce computational burden during crediting, as described in the k-fold validation steps. Trace plots and plots of Gelman-Rubin $\widehat{R}$ statistics (i.e., parameter shrinkage factors) are provided in Appendix C.

Calibration and validation of the model were conducted simultaneously using a k-fold cross-validation procedure with k=5 folds. This is a statistical approach that ensures independence between calibration and validation datasets, as described on page 4 of the Model Requirements, and highlighted in the definition section for the term "Validation". In brief, the approach employed for this report consists of six major steps:

1. Study sites were first randomly divided into five non-overlapping disjoint groups (i.e. folds). If a given experiment was assigned to a fold, all the individual observations associated with that experiment were then assigned to that fold. To maintain balance between emission sources, fold assignment was done separately for SOC and N_2O so that each fold contained one-fifth of the SOC sites and one-fifth of the N_2O sites. Maps of the study sites used in this report are provided in Figure 2.
2. Second, for each of five data splits, one fold was reserved for validation and the remaining four folds were used for model calibration, giving an approximately 80-20 percent allocation in each data split between calibration and validation datasets, respectively.
3. Third, for each data split, Bayesian calibration was performed with DREAM as described above, resulting in a joint posterior distribution of model parameters estimated from the calibration data for that fold. As noted above, the first 50 percent of each chain was discarded as the "burn-in" period.
4. Fourth, out-of-sample predictions were made in each data split using the validation dataset and parameters from the joint posterior distribution that was calculated in step three from sites used for calibration. Out-of-sample predictions were then used to estimate the posterior predictive distributions of N_2O and SOC differences between the experimental treatments at the second time point, similar to the methods described in Gurung et al. (2020) and Mathers et al. (2023).
5. Fifth, model performance was quantified by computing model bias, root mean square error (RMSE), and 90 percent prediction interval coverage of the validation data, evaluating each metric separately for each fold and then calculating their means across all folds.
6. For the sixth and final step, the model was re-calibrated using the full dataset, and the resulting calibrated parameters retained to serve carbon credit predictions by saving 176 joint posterior draws evenly spaced over the post-burn-in period[1]. Whereas accurate estimates of tail probabilities are needed to assess validation criteria (i.e., coverage rates of 90 percent intervals), credit simulations require fewer posterior draws because only accurate estimates of variance are needed; because draws are by definition less likely to fall in the tails than near the center of the distribution, it takes more draws to obtain stable estimates of the 5th and 95th percentiles than to obtain stable estimates of variance (see Davison et al., 1997, Ch. 2.5.2 for related discussion in the context of bootstrap resampling). Furthermore, crediting is done at a large scale and DayCent simulations can be time-consuming, so computational efficiency is a key consideration. See Section 10 for a description of how the saved posteriors are used during crediting, and Appendix D for a comparison between thinned and full posteriors.

[1] Nineteen or twenty evenly spaced draws were taken from each of the nine chains, resulting in 176 total draws.

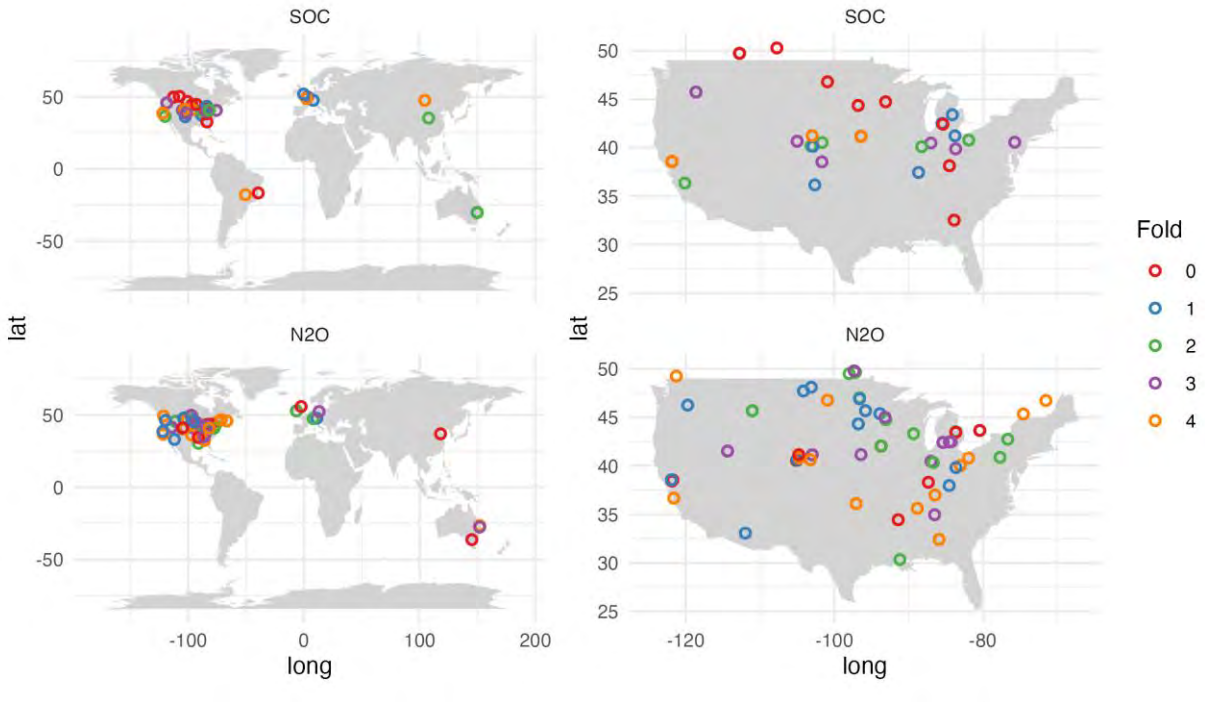

Figure 2. Locations of experimental sites with measurements of SOC (top) and N_2O (bottom) used for calibration and validation of DayCent-CR.

The prior distributions of parameters adjusted during the calibration process and summary statistics of marginal posterior distributions of model parameters for the final step using the full dataset are provided in Appendix A. For the full parameter set and auxiliary files needed to reproduce the validation, please see DayCentCR_1.1.0_validation_supporting_files.zip.

Choosing the final parameter set by recalibrating to the full dataset, as described in step 6 of the calibration procedure above, is common practice in fields making frequent use of statistical methods for cross-validation (Kuhn and Johnson, 2013; Roberts et al., 2017) because it provides a final parameterization that is maximally informed by all available training data. This approach complies with Section 2.3.1.2 of the Model Requirements ("the method of choosing the final parameter set must be a prespecified part of the cross-validation method", and "the parameter values identified as the final validated set…must be the ones used [for crediting]"), and cross-validation gives a reasonable estimate of the performance that can be expected from the final parameter set (a model fit to the full training set typically performs as well or better on new data than was observed on hold-outs from the training set during cross-validation (Roberts et al., 2017)). However, steps four and five of our cross-validation procedure inherently perform validation on five separate parameter sets, one per data split, that will differ slightly from the final joint posterior distribution created in step six for use during crediting, so care is needed to demonstrate that the final values and the cross-validation results are consistent with each other. To check this, a comparison between the parameter distributions obtained from cross-validation and from fitting the full dataset was completed (Appendices A and C), as well as between the distributions of model outputs (Section 10) to ensure differences between the validated and final parameterizations, particularly for the most sensitive parameters, would not materially change model results.

3.2 Model setup

For calibration and validation, DayCent-CR was run for all treatments and sites (See Supporting Materials 1 for site-level summaries and the full dataset in DayCentCR_1.1.0_validation_supporting_files.zip). The following describes the procedure used to simulate the experimental sites for the calibration and validation approach described above.

The model-driving input files for each site were created following the procedures described in Section 5. When site-specific data were available from the studies, these data were used as model inputs. Where site-specific data were not available from the experimental publication, soil data (texture and pH, which were then used to estimate other missing soil parameters) were taken from the gSSURGO database (Soil Survey Staff, 2022) and management information was estimated from typical agronomic practice in the region (see Section 5.2 and Section 5.3 for details). Because accurate representation of site-level weather conditions is crucial to ensuring model predictions match site-level agronomic histories, climate data (minimum and maximum daily temperature, precipitation) were taken from the highest-resolution available of the following sources, in this order of preference:

- On-site weather stations (when records were available for the historic period as well as the experimental period)
- PRISM database (United States only)[2]
- The nearest national weather station for non-U.S. sites (Barr´e et al., 2010; Environment and Climate Change Canada[3]; Australian Government; Bureau of Meteorology[4])
- Soil and Water Assessment Tool global weather data[5]
- Global Land Data Assimilation System[6]

DayCent-CR divides SOC into three conceptual pools that differ only in their turnover time and do not correspond to any physically measurable soil fractions. In order to estimate the proportions of the SOC pools, equilibrium simulations were conducted of native grassland (5,000 to 7,000 years) to bring the SOC pools to a steady state, followed by a simulation of historical agricultural management based on available data from the site or the region it is in, consistent with methods and data used in the U.S. National GHG Inventory (U.S. EPA, 2020). These historical periods before the experiments began were simulated using the default parameters in the DayCent-CR model. At the end of the historic period, the estimated proportions of SOC pools are used to fractionate the measured SOC at the beginning of the experiment to active, slow, and passive SOC pools in the model. After initialization of the SOC pools to match these proportions to the measured pool size, simulations of the experimental period were used to perform the calibration and validation process (see Section 3.1).

[2] http://prism.oregonstate.edu
[3] https://climate.weather.gc.ca/historical_data/search_historic_data_e.html
[4] http://www.bom.gov.au/climate/data/
[5] swat.tamu.edu/data/cfsr
[6] https://ldas.gsfc.nasa.gov/index.php/gldas

21 of the 48 experimental sites that generated SOC observations used in this analysis did not report SOC measurements at the beginning of the experiment. In these cases the entire history of the experiment was simulated, but the simulations were divided into two eras:

1. The period between experiment start and first SOC measurement was simulated as part of the historic period, then the simulation was stopped and model SOC was initialized to match the first SOC measurement as described above.
2. The period between first SOC measurement and experiment end was then simulated beginning from the reinitialized SOC values and the simulation result was used for calibration and validation.

This approach conforms to CAR SEP requirements that model simulations of SOC change for carbon credits must be initialized with in-field measurements of SOC. In other words, all reported experimental practices are modeled, but the model is calibrated and validated using equilibrium simulations, site history, and initial SOC measurements in the same way as this information would be used in a CAR SEP project, and calibration and validation are constrained to the time periods for which SOC observations are available. For some sites, initial SOC measurements were quite late relative to the full duration of the experiment (e.g., the Otis site, which started in 1966 but SOC was not measured until 2005). While this does leave portions of experimental history out of the calibration/validation exercise, initial SOC is a highly influential model driver and the error introduced by attempting to estimate SOC at experiment start time would be more detrimental to model performance than restricting validation of these sites to the period that is well constrained by measurements.

The same initialization procedure will apply to the use of the model in carbon crediting for a CAR SEP project, using site latitude and longitude, soil carbon measurements, and soil physical and chemical properties (described in Section 5). Comparable site-specific climate data (as demonstrated by peer-reviewed evidence in the CAR1459 Monitoring Plan) will be provided for all project simulations. Native grassland will be assumed for all the CAR SEP projects for the initial period simulated to reach a model steady-state (consistent with the U.S. National GHG Inventory and current implementation in COMET-Farm). The version of the model evaluated in this report requires the input of management information to begin in the year 2000. This means the model spin-up period, as described in Model Requirements Section 3.4.1.3, will extend from Jan 1, 2000 until the beginning of the required historic baseline period for a given location being simulated. All management information for the model spin-up period, required historic baseline period, and with-project periods must meet CAR SEP requirements and will be described in the CAR1459 Monitoring Plan.

3.3 Documentation of model parameter sets

DayCent-CR has hundreds of parameters and calibrating all of them simultaneously would be computationally impractical. Many of these model parameters have been previously tested and applied extensively without change, for example annually in U.S. National GHG inventory simulations (U.S. EPA, 2020), and not all model parameters have an impact on SOC and N_2O dynamics. Sixty-two (62) parameters were initially selected for consideration in the calibration exercise (Appendix A).The parameters included 27 directly related to SOC processes and to DayCent-CR's soil organic matter decomposition sub-routine, 1 parameter related to soil water ("FWLOSS(2)") which scales potential evapotranspiration, and 34 parameters related to N_2O. Candidate parameters for SOC calibration were selected because they control the decay rate of the SOC pools and carbon transfer efficiency between pools and directly affect the magnitude of SOC stocks and SOC stock differences. Similarly, candidate parameters for N_2O calibration were selected because they control nitrification, denitrification, water

balance, and other dynamics related to nitrogen cycling. While other parameters associated with other processes such as plant production may indirectly influence modeled SOC and N_2O dynamics, the selected parameters mediated these relationships. Consequently, those parameters were left as constants and assigned the default values used in COMET-Farm and the U.S. National GHG Inventory.

All parameters were assigned independent prior distributions provided in Appendix A. For the SOC calibration, the initial list of parameters and their prior ranges were taken from the values reported in Gurung et al. (2020). This initial list was then updated to include potentially influential water parameters and to align each parameter's prior range with values that are biogeochemically plausible for the conditions present in the project area.

The same parameters for SOC were calibrated that had been identified using global sensitivity analysis (GSA) in Mathers et al. (2023) and the past calibration report (Indigo Ag, 2022). In this analysis, 10 parameters were identified for SOC that each contributed more than 0.5 percent of variance (Appendix A). This inclusive cutoff was chosen to reduce the dependence of the GSA on the calibration dataset. Bayesian calibration was then performed on these 10 most influential parameters for SOC, and the rest of the parameters were fixed to their default values. The calibration also included REML estimates of four variance parameters (σ^2_{site}, $\sigma^2_{\text{site-year}}$, σ^2, and ν), as described in Section 3.1.

To identify additional parameters for calibration with N_2O flux observations, we used one at a time (OAT) sensitivity analysis. OAT was used instead of GSA because GSA results were inconsistent with expert knowledge of model structure, and results varied across the project domain. By contrast, OAT identified parameters that directly control N_2O across the project domain. We believe this difference is caused by the more thorough exploration of parameter space by GSA, which makes it more sensitive to outliers, nonlinear effects, parameter interactions and numerical artifacts (Razavi and Gupta, 2015). Appendix A presents the parameters used in the N_2O OAT, rationale for their prior distributions, and results of the sensitivity analysis.

One important consideration when choosing the parameters was to avoid selecting site- or environment-specific parameterizations, given that the goal of this exercise is to perform a joint calibration across all sites. This ensures that the model is generalizable and reduces the chance of overfitting a parameter to a specific site or site categories that are better represented in the calibration dataset.

Based on OAT, we added two additional parameters (damrmn(1) and VARAT[1,2](1,1)) to the original ten previously identified by GSA (Mathers et al., 2023; Indigo Ag, 2022). These parameters were found to be highly sensitive across sites, consistent with expectations based on model structure. OAT characterized the univariate influence of each parameter on the variance of modeled N_2O fluxes. To test the sensitivity of each parameter, we ran the model at fifty random draws from the prior parameter distribution while keeping all other parameters fixed at their defaults. Then we calculated the coefficient of variance (variance divided by the mean, CV) of predicted N_2O flux. Parameters were ranked by this CV. Although OAT cannot capture higher order interactions between parameters, it describes relative model output variability. Other than the selected parameters, N_2O showed < 1 percent sensitivity (CV) to the majority of the parameters.

Appendix A includes the following:

- Prior distributions for all DayCent parameters included in the Bayesian calibration, along with parameter descriptions and justification for prior specification.

- Summary statistics of the marginal posterior distributions for SOC parameters and random effects.
- Summary statistics for N_2O parameters and random effects.

Sampler diagnostics, including plots and statistical summaries of marginal posterior densities for each fold, along with trace plots, and Gelman-Rubin ($\hat{R}$) statistics, are available in Appendix C. For the full parameter set and auxiliary files needed to reproduce the validation, please see DayCentCR_1.1.0_validation_supporting_files.zip.

During the calibration process, instead of estimating the posterior distribution of model parameters for each Land Resource Region (LRR) separately, model parameters were treated as population-level variables and obtained using a single calibration for all LRRs. Because the joint posterior was used from this single calibration in crediting runs, the bias and uncertainty estimates presented here are generalizable to all crop types and management practices represented within the dataset used in this validation report.

3.4 Justification for splitting of experimental data

Because only a limited number of experiments have measured enough parameters over a long enough time span to parameterize soil biogeochemical models confidently, it is desirable to use studies from sites with the highest-quality measurements for both calibration and validation. To retain statistical independence of calibration and validation data (Model Requirements, Section 2), the calibration and validation were performed using a 5-fold cross-validation method following Section 2.3 of the Model Requirements. Cross-validation retains statistical independence of calibration and validation data by ensuring that each candidate model is never evaluated against the same data that trained it, but also retains efficiency by ensuring that every data point contributes to both the calibration and validation processes. Because of these properties, cross-validation is widely used for model evaluation in cases where the goal of calibration is to minimize prediction bias when data are limited.

To retain independence while dividing the available dataset into five folds, experimental sites were assigned into folds, taking into account the likelihood of high spatial and/or temporal correlation of repeated measurements from the same site. For sites where all experiments share a physical location and management history, all observations were assigned to the same fold. For sites with multiple experiments that are near each other but differ in timing or duration of experiment, crop type, or primary experimental goal (i.e., that differ at the level of CFG x PC combination, per Model Requirements, Section 2), the data from these experiments may be correlated in space (climate and soil factors, conditions during model spin up) but are likely uncorrelated in management. Therefore, these experiments were considered as separate "sites" and were each randomly allocated to folds. The intention of this approach was balancing the need for independent folds against the need to ensure that each fold contained approximately one-fifth of the data, as well as sufficient data from each crop and practice to be validated. To check for correlations not addressed by this approach, spatial variograms of model residuals after calibration were also created (see Appendix C). While there appears to be spatial correlation in SOC pool size, the range of spatial correlation was estimated to be 427 kilometers, and there does not appear to be spatial correlation in the SOC residuals, mean N_2O flux, or N_2O residuals.

4 Project domain

Follows Model Requirements, Sections 3.1 and 3.2

4.1 Practice categories

The project intends to credit practices (Appendix B) that are included in four PCs (Table 3).

Table 3. Practice categories included in the project domain.

PC	Abbreviation
Inorganic nitrogen fertilizer application	NFERT
Organic amendments application	ORG
Soil disturbance and/or residue management	DISTURB
Cropping practices	CROP

4.2 Crop functional groups

The project includes crops spanning five CFGs as provided in Table 4.

Table 4. CFGs included in the project domain.

CFG	Example crops	Lifecycle	Photosynthetic pathway[7]	Growth Form	Nitrogen fixing	Flooded / Not Flooded
C3A	Wheat, canola	Annual	C3	Herbaceous	No	Non-flooded
C3AN	Soy, annual alfalfa[8]	Annual	C3	Herbaceous	Yes	Non-flooded
C3PN	Perennial alfalfa, crown vetch	Perennial	C3	Herbaceous	Yes	Non-flooded
C3S	Cotton	Annual	C3	Shrub	No	Non-flooded
C4A	Corn, sorghum	Annual	C4	Herbaceous	No	Non-flooded

[7] Indicates the method of carbon fixation in plants (see Forseth, 2010)

[8] Note that crops with the genetic potential to grow perennially (e.g., alfalfa, vetch, clover) were included in this CFG when they were only grown for a single season, qualifying them as annuals per Section 3.2.1 of the Model Requirements.

4.3 Land resource regions

The project encompasses all LRRs in the continental U.S. (Table 5) and 8 IPCC climate zones (Table 6).

Table 5. LRRs occurring in the project area.

LRR	Name
A	Northwestern Forest, Forage, and Specialty Crop
B	Northwestern Wheat and Range
C	California Subtropical Fruit, Truck, and Specialty Crop
D	Western Range and Irrigated
E	Rocky Mountain Range and Forest
F	Northern Great Plains Spring Wheat
G	Western Great Plains Range and Irrigated
H	Central Great Plains Winter Wheat and Range
I	Southwest Plateaus and Plains Range and Cotton
J	Southwestern Prairies Cotton and Forage
K	Northern Lake States Forest and Forage
L	Lake States Fruit, Truck Crop, and Dairy
M	Central Feed Grains and Livestock
N	East and Central Farming and Forest
O	Mississippi Delta Cotton and Feed Grains
P	South Atlantic and Gulf Slope Cash Crops, Forest, and Livestock
R	Northeastern Forage and Forest
S	Northern Atlantic Slope Diversified Farming
T	Atlantic and Gulf Coast Lowland Forest and Crop
U	Florida Subtropical Fruit, Truck, and Specialty Crop

Table 6. Climate zones defined by IPCC (2019) appearing in the project.

IPCC climate zone	IPCC climate zone abbreviation
warm temperate dry	WTD
cool temperate dry	CTD
warm temperate moist	WTM
cool temperate moist	CTM
boreal dry	BD
boreal moist	BM
tropical moist	TrM
tropical dry	TrD

4.4 Soils

The project includes all 12 soil textures in the USDA soil texture classification along with the clay content at the midpoint of each texture class definition (Table 7).

Table 7. Names, abbreviations, and midpoint clay contents for USDA soil texture classes occurring in the project area.

Abbreviation	Texture class	Percent clay
Cl	Clay	70
ClLo	Clay loam	35
Lo	Loam	20
LoSa	Loamy sand	10
Sa	Sand	5
SaCl	Sandy clay	40
SaClLo	Sandy clay loam	30
SaLo	Sandy loam	10
Si	Silt	5
SiCl	Silty clay	45
SiClLo	Silty clay loam	35
SiLo	Silt loam	15

4.5 Emission sources

The model was validated for changes in SOC and N_2O emissions. Emissions of CH_4 are not included in this report.

4.6 Domain covered by this validation

The domains validated in this report are summarized in Table 8 and Table 9 for SOC and N_2O, respectively. The differences between Table 8 and Table 9 are due to a lack of available observations in Indigo's dataset. Table 10 and Table 11 summarize the biophysical attributes for each validated domain.

Following Model Requirements Section 3.3.1 paragraph 5 (allowing multiple CFGs to be aggregated when validating the ORG PC for SOC), a combined dataset for the ORG PC from all annual crops combined is included.

Model Requirements Section 3.3.1 paragraph 7 were followed allowing cropping systems that use irrigation as a background practice to not require validation of the WATER PC. This provision is met because multiple studies in the validation dataset use irrigation as a management practice. The range of precipitation regimes included in the validation dataset (SOC 17-163 cm yr^{-1}, N_2O 18-172 cm yr^{-1}), covering at least 3 LRRs or IPCC climate zones, are considered an adequate proxy for testing the effects of artificial rainfall supplied by irrigation.

To ensure accuracy in determining the range of climate conditions validated for each category, IPCC climate zones were assigned to each site through application of the decision tree provided in Figure 3A.5.2 in the *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*

(IPCC, 2019) using high spatial resolution climate data from TerraClimate[9] and WorldClim[10] 2.1. One site (Fort Valley) was manually reclassified from Warm Temperate Dry to Warm Temperate Moist on grounds that the evapotranspiration reported by TerraClimate was anomalously high and that papers from this site consistently refer to it as having a humid climate.

Table 8. Combinations of CFG and PC that are validated for SOC in this project.
+ is validated, - is not validated

	Crop Functional Group				
PC	**C3A**	**C3AN**	**C3PN**	**C3S**	**C4A**
CROP	+	+	+	+	+
DISTURB	+	+	-	+	+
NFERT	+	+	+	-	+
ORG	+	+	-	+ via Org x Annual	+

Table 9. Combinations of CFG and PC that are validated for N_2O in this project.
+ is validated, - is not validated

PC	Crop Functional Group				
	C3A	**C3AN**	**C3PN**	**C3S**	**C4A**
CROP	+	+	+	-	+
DISTURB	+	+	-	-	+
NFERT	+	+	-	-	+
ORG[a]	+	-	-	-	+

a. Note that CFGs have not been pooled for validation of ORG x N_2O. Although Model Requirements Section 3.3.1 paragraph 5 does not state any limits on which emission sources may use the pooled-CFGs rule, we believe it should be applied only when validating SOC changes. The underlying logic of the rule is that the amount of carbon in the added organic matter is the determining control on the effect of ORG on SOC, and that this will overwhelm any differences between annual CFGs. In contrast, the effect of organic amendments on N_2O flux could trigger crop-mediated feedbacks, e.g., the N_2O response to ORG cannot necessarily be assumed similar in legumes and non-legumes. Therefore, we interpret the "pooled annual CFGs for ORG" rule as a "pooled annual CFGs for ORG x SOC" rule and recommend that CAR consider stating this explicitly in future revisions of the Model Requirements.

[9] https://www.climatologylab.org/terraclimate.html
[10] https://worldclim.org/data/worldclim21.html

Table 10. Biophysical attribute ranges across which each PC/CFG was validated for SOC, meeting minimum requirements outlined in Model Requirements Section 3.3, Requirement 2.

All PC-CFG categories pass the "stacking" requirement (Model Requirements Section 3.3, Requirement 1) by containing at least one study that isolates the effect of the PC change being validated. See Supporting Materials 1 for counts of stacked and unstacked observations. **Clay range** refers to the difference in the maximum and minimum clay percentages among sites.

PC	CFG	# of sites	# of observations	LRRs	Climate zones	Countries	Soils	Clay range (%)
CROP	C3A	26	400	C, H, L, M, P, S	CTD, CTM, WTD, WTM	Australia, Canada, England, France, USA	Cl, ClLo, Lo, SaClLo, SaLo, SiCl, SiClLo, SiLo	54
CROP	C3AN	22	319	C, H, L, M, P, S	CTD, CTM, TrM, WTD, WTM	Australia, Brazil, Canada, England, France, USA	Cl, ClLo, Lo, SaLo, SiCl, SiClLo, SiLo	54
CROP	C3PN	7	91	S	CTD, CTM, TrM, WTD, WTM	Brazil, Canada, China, England, France, USA	Lo, SaLo, SiClLo, SiLo	20
CROP	C3S	6	162	C, P	TrM, WTD, WTM	Australia, Brazil, USA	Cl, ClLo, SaLo	54
CROP	C4A	20	230	C, H, L, M, P, S	CTD, CTM, TrM, WTD, WTM	Brazil, England, France, USA	Cl, Lo, SaLo, SiCl, SiClLo, SiLo	40
DISTURB	C3A	13	96	B, C, F, G, H, L, M, P	CTD, CTM, WTD, WTM	France, Switzerland, USA	Cl, ClLo, Lo, SaLo, SiClLo, SiLo	35
DISTURB	C3AN	11	75	C, L, M, N, P	CTD, CTM, TrM, WTD, WTM	Brazil, France, Switzerland, USA	Cl, ClLo, Lo, SaLo, SiClLo, SiLo	40
DISTURB	C3S	4	49	C, P	TrM, WTD, WTM	Australia, Brazil, USA	Cl, ClLo, SaLo	43
DISTURB	C4A	15	228	K, L, M, N, P	CTD, CTM, TrM,	Brazil, France, Switzerland, USA	Cl, Lo, SaLo, SiClLo, SiLo	40

PC	CFG	# of sites	# of observations	LRRs	Climate zones	Countries	Soils	Clay range (%)
					WTD, WTM			
NFERT	C3A	16	227	B, C, F, H, L, M, P, S	CTD, CTM, WTD, WTM	Canada, England, France, USA	Lo, SaLo, SiClLo, SiLo	25
NFERT	C3AN	8	83	C, L, M, P, S	CTM, WTD, WTM	France, USA	Lo, SaLo, SiClLo, SiLo	25
NFERT	C3PN	4	59	S	CTD, CTM, WTM	China, England, France, USA	Lo, SaLo, SiLo	19
NFERT	C4A	16	168	C, E, H, K, L, M, N, P, S	CTD, CTM, WTD, WTM	France, USA	ClLo, Lo, LoSa, SaLo, SiClLo, SiLo	25
ORG	Annuals	10	76	B, C, E, L, M	CTD, CTM, WTD	Canada, England, USA	ClLo, Lo, LoSa, SaLo, SiClLo, SiLo	29
ORG	C3A	10	125	B, C, M, S	CTD, CTM, WTD, WTM	Canada, England, Switzerland, USA	Cl, ClLo, Lo, SaLo, SiClLo, SiLo	34
ORG	C3AN	5	31	C, M, S	CTM, WTD, WTM	England, Switzerland, USA	Cl, Lo, SaLo, SiClLo, SiLo	34
ORG	C4A	7	37	C, E, L, M, S	CTD, CTM, WTD, WTM	England, USA	ClLo, Lo, LoSa, SaLo, SiClLo, SiLo	25

Table 11. Biophysical attribute ranges across which each PC/CFG was validated for N_2O, meeting minimum requirements outlined in Model Requirements Section 3.3, Requirement 2.

All PC-CFG categories pass the "stacking" requirement (Model Requirements Section 3.3, Requirement 1) by containing at least one study that isolates the effect of the PC change being validated. See Supporting Materials 1 for counts of stacked and unstacked observations. "Clay range" refers to the range of clay percentage among sites.

PC	CFG	# of sites	# of observations	LRRs	Climate zones	Countries	Soils	Clay range
CROP	C3A	13	103	E, F, G, H, K, L, M, N	CTD, CTM, WTM	Canada, USA	Cl, ClLo, Lo, SaClLo, SaLo, SiLo	51
CROP	C3AN	14	92	E, F, G, K, L, M, N	CTD, CTM, WTM	Canada, USA	Cl, ClLo, Lo, SaClLo, SaLo, SiLo	51
CROP	C3PN	6	86	E, F, K, M, S	CTD, CTM	Canada, USA	Cl, ClLo, Lo, SiLo	40
CROP	C4A	11	115	B, G, K, M, N	CTD, CTM, WTD, WTM	Canada, USA	Cl, ClLo, Lo, SaClLo, SiLo	48
DISTURB	C3A	11	61	C, E, F, H, L	CTD, CTM, WTD	Canada, China, Ireland, Switzerland, USA	Cl, ClLo, Lo, SaLo, SiLo	51
DISTURB	C3AN	7	25	C, E, L, M	CTD, CTM, WTD	Canada, Switzerland, USA	Cl, Lo, SiLo	43
DISTURB	C4A	11	78	C, G, K, L, M, R	CTD, CTM, WTD, WTM	Canada, China, USA	Cl, ClLo, Lo, SaLo, SiLo	46
NFERT	C3A	19	206	C, D, F, H, K, L, N, P	CTD, CTM, WTD, WTM	Australia, Canada, Germany, Ireland, USA	Cl, ClLo, Lo, LoSa, SaLo, SiClLo, SiLo	69
NFERT	C3AN	6	14	F, G, K, L, N, S	CTD, CTM, WTM	USA	ClLo, Lo, SiLo	15
NFERT	C4A	24	528	B, G, K, L, M, N, P, S	CTD, CTM, WTD, WTM	Australia, Canada, USA	Cl, ClLo, Lo, SaLo, SiLo	62

PC	CFG	# of sites	# of observations	LRRs	Climate zones	Countries	Soils	Clay range
ORG	C3A	5	39	D, P	CTD, CTM, WTM	Canada, Switzerland, USA	Cl, LoSa, SaLo, SiCl, SiLo	40
ORG	C4A	6	98	B, D, G, N	CTD, CTM, WTD, WTM	Canada, USA	Cl, ClLo, SiLo	63

5 Description of data requirements

Follows Model Requirements, Section 3.3.

To run DayCent-CR, the following information must be provided.

5.1 Site-specific model drivers

- Daily weather data for the site and time period to be simulated: precipitation, maximum and minimum temperature. DayCent-CR does not use the extra weather variables that are optional in other versions of DayCent (solar radiation, relative humidity, and windspeed); these variables are estimated using an internal calculation based on site latitude.
- Soil texture (sand, silt, clay), bulk density, pH, and hydraulic conductance for each soil horizon from the surface to the first fully root-restrictive layer. Because DayCent imposes strict internal restrictions on the relationships between these variables, Indigo follows the recommendation of the DayCent developers to compute bulk density and hydraulic conductance from soil texture using the equations of Saxton et al. (2006) even when direct measurements are available.
- Initial SOC stock in the 0 – 30 cm soil layer
- Depth to bedrock
- Site latitude

5.2 Management information

- Site history from before the experiment, for running modeled SOC pools to equilibrium: native vegetation type, approximate historic management. When not available, site history is inferred from local native vegetation types and regional historic agricultural records.
- Identities, including cultivar information when possible, of all crops in the rotation
- Planting dates and methods
- Tillage dates, types, and intensities: implements used, depth, number of passes
- Harvest and termination dates, methods, and types (e.g., grain, hay percent offtake, fruit, etc.)
- Residue management (e.g., burning, straw/stover removal)
- Nitrogen fertilization dates, types, amounts, NI and CRNF use, and application methods
- Herbicide dates and types
- Irrigation dates, types, amounts
- Organic matter addition dates, types (e.g., manure, green manure, compost, straw amendments, nitrogen fraction, C:N ratio, mass of the dry fraction)

5.3 Procedures for missing data

While most published experiments give sufficient detail on the experiment treatment management, pre-experiment details are often lacking. Whatever pre-experiment detail is provided in study documentation, or derived through communication with the experiment managers, is incorporated into model inputs for the simulation period leading up to the experiment. Sometimes more details can be gleaned from companion articles not emphasizing SOC or N_2O. When no other detail is available for the pre-experiment period information, the land use history most similar to the experiment itself is selected.

Where no specific information is available, as is often the case in the portion of the simulation period years and decades preceding the experimental start date, common regional practices are derived from available sources on crops grown, tillage and fertilizer inputs (NASS, ERS-ARMS, CTIC). Where more soil information is needed than provided in the associated publications and by the experimental

researchers, information is obtained from USDA Web Soil Survey for the soil series mentioned in the publications.

6 Description of validation data collection process

Follows Model Requirements, Section 3.3 Requirement 1

Studies used for model validation were either identified from a database of SOC and N_2O experiments that is contributed to and maintained by DayCent model developers from multiple research teams or obtained through publication search engines. The database tracks experiments found in peer-reviewed literature that report effects of management on SOC or N_2O. The database is used to develop a set of model inputs for parameterization and testing that have been updated and used continually alongside such projects as the U.S. National GHG Inventory (U.S. EPA, 2020), in which the DayCent model simulates U.S. agricultural GHG emissions for reporting to the United Nations Framework Convention on Climate Change.

The experiments selected have sufficient management information and reliable soil data to support model development and testing activities, i.e., all parameters listed above in Section 5 were reported, or could be inferred according to the procedures reported above in Section 5.3. The data compilation process focused on sites rather than individual publications because in many cases, especially for the longest-running studies that are of highest value for model validation, the SOC and/or N_2O measurements and the information needed to parameterize DayCent-CR for the study are reported in multiple separate publications from one site. Once a site was selected for inclusion in the database, all relevant publications for that experiment were found by searching for combinations of the name of the experiment or research station, key authors, and geographic descriptions (e.g., name of nearest town or of the institution sponsoring the research site), and by following citations in publications already identified for the site.

The identified studies are all from publicly reported long-term agricultural research sites where the effect of agronomic practices has been evaluated. Much effort by the DayCent model development team and Indigo has gone into assembling the relevant publications and databases associated with each experiment modeled. This includes all relevant datasets that the development team is currently aware of, through searching published literature, grey literature, and inquiries in research networks. Articles published any time before the end of 2022 were considered for inclusion. For this validation and calibration, data were evaluated from hundreds of papers and sites reporting SOC changes and/or N_2O emissions in cropland soils.

6.1 Data exclusion criteria

Sites were excluded when they failed to meet one or more of the following criteria:

- Sufficient information was provided to model the site accurately, or missing data could be inferred according to procedures reported above in Section 5.3.
- When a study was conducted outside the United States, it was in an IPCC climate zone that is present in the project domain.
- If SOC was measured:
    - The measured quantity was stock (not just concentration) of soil organic carbon, measured in units convertible to g C m^{-2} at each timepoint (typically in the form of percent carbon measured by dry combustion or wet digestion, plus a reliable measurement of soil bulk density by core or ring methods). Measurements of total soil carbon were used only if there was evidence that inorganic carbon (carbonates) were not present.

    - SOC was measured to a depth not less than 23 cm[11] in layers allowing reasonable approximation to 0-30 cm SOC stock: If no layer ended at exactly 30 cm, then the total SOC was required to be estimable by interpolation of bulk density and percent carbon across the depths that were reported. See DayCentCR_1.1.0_validation_supporting_files.zip for details of the transformations applied to each measurement. Most of the sites excluded from SOC validation were excluded because of too-shallow SOC measurements.
    - SOC was measured at least two times spanning a total interval of at least three years. If the first 30-cm SOC measurement was not taken at the onset of the experiment, only the data from timepoints after the first usable measurement were used.
- If N_2O was measured:
    - The measured quantity was net N_2O-N flux from soil to atmosphere, in units convertible to g N ha^{-1}, across the duration of the measurement (typically in the form of repeated static chamber surveys; flux tower data were also accepted).
    - N_2O fluxes were measured across at least one full growing season (planting to harvest).
    - Sampling frequency was sufficient to confidently estimate total flux for the season: Nominal sampling intervals of no more than 14 days, and preferably of 10 days or less. Data gaps in otherwise well-sampled years were evaluated case by case and were either filled by linear interpolation or fluxes were computed for a shortened season that started after or ended before the gap. Most of the sites excluded from N_2O validation were excluded because of low sample frequency.
    - Agronomic management data were reported in enough detail to confidently estimate total nitrogen inputs for each year measured (assuming fertilization "typical for the region" was not sufficient).

6.2 Temporal aggregation of N_2O data

Many N_2O studies reported year-round measurements but had much lower sampling frequencies during the fallow season. In these cases, seasonal flux was computed from the portion of the year that was well-sampled, and reported whole-year totals were not used. Studies that focused on short-duration events such as fluxes after thaw pulses, fertilization, or irrigation were included only if they also met the criteria above.

To demonstrate adequate model performance for N_2O at the seasonal scale, validation was done on changes in seasonal N_2O flux totals rather than changes in daily fluxes. Therefore, the site selection process excluded studies that focused solely on short duration N_2O events (e.g., measurements that track a thaw pulse but are then discontinued before the crop growing season) but included studies with short events in a full season of monitoring (often with increased sampling frequency around the event). After the full dataset was assembled, the observed N_2O fluxes were checked to ensure they covered the entire calendar year, as specified by Model Requirements Section 3.3 Requirement 1. This check was done both in aggregate and within climate zones to confirm adequate coverage of seasonal cycles that may vary with geography (Figure 3a) and within CFGs to confirm adequate coverage of management cycles that may vary with crop type (Figure 3b).

[11] 23 cm is equal to 9 inches and is a relatively common sample depth at sites where plow depth informed the soil sampling design.

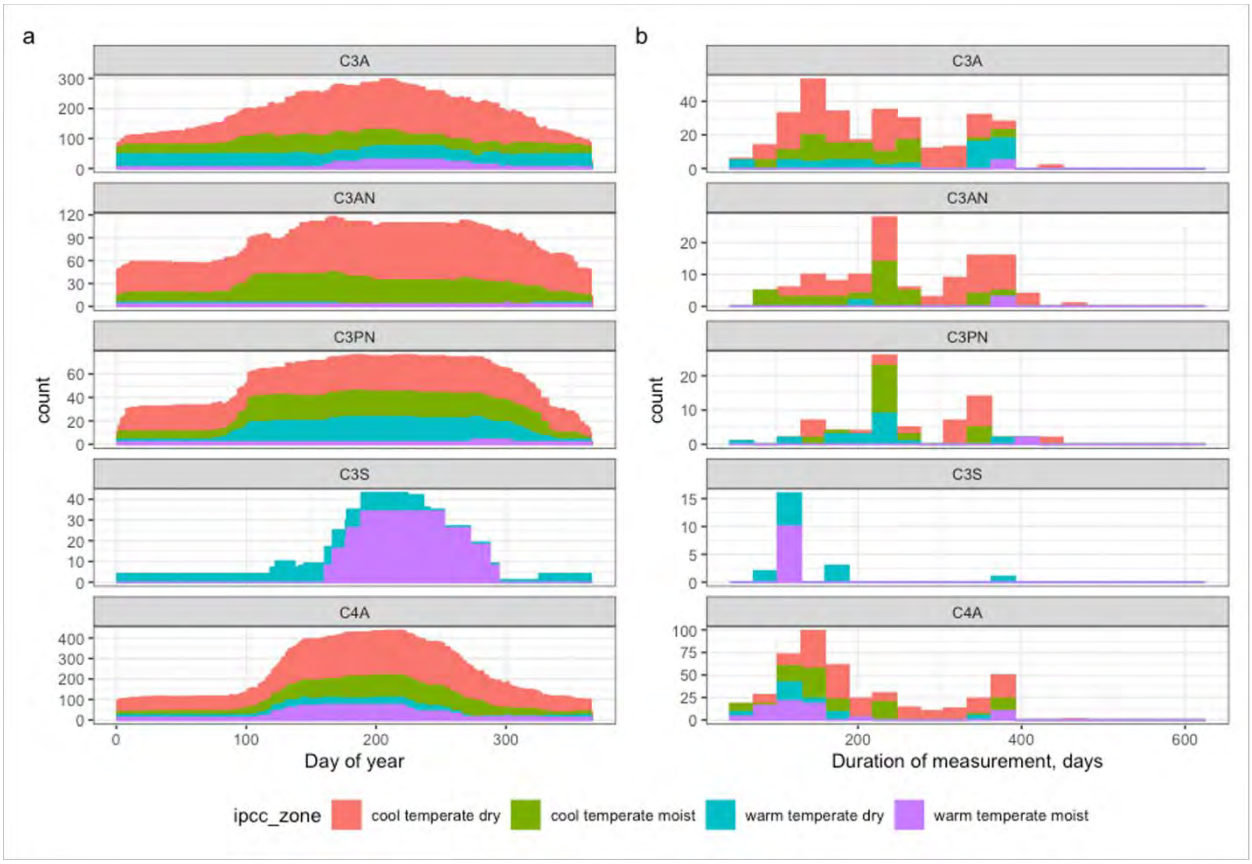

Figure 3. a) Number of N_2O observations that cover each day of year, by CFG and climate zone; b) Distribution of N_2O measurements by length of monitored season, with ~1-month bins, by CFG and climate zone. Histograms are stacked, so counts (y-axis) represent the total number of observations across all IPCC climate zones.

To aggregate N_2O fluxes from daily to seasonal scale, the sum of daily fluxes estimated by linear interpolation between observations beginning on the first measurement day and ending on the last measurement day were computed (Halvorson et al., 2016). Seasonal totals were not converted to estimates of annual fluxes, but instead were compared directly to the sum of daily modeled fluxes from the same range of days.

Although DayCent-CR does not predict negative N_2O fluxes, they do occur (Chapuis-Lardy et al., 2007). The N_2O flux dataset included 5 of 1,222 observations (0.4 percent) that were below zero indicating net N_2O uptake. All negative fluxes were retained because they are biophysically plausible and consistent with the distribution of N_2O measurements. All of the negative values were from unfertilized treatments and include:

- Four negative values that were > -20 g N ha^{-1} and statistically indistinguishable from zero.
- The minimum N_2O flux of -460 g N ha^{-1} occurred at the glenlea_Tgas site (Tenuta et al., 2016) and was collected without replication using the flux gradient approach but did not appear to be an outlier among other observations at the same site. The glenlea_Tgas site was one of only two in the validation dataset that used the flux gradient approach. The other site was elora (Wagner-Riddle et al., 2007) and had an average standard error of a similar magnitude (415 g N ha^{-1}, n = 15).

6.3 Uncertainty calculations

Where studies reported the uncertainty of their observations, the reported uncertainty values were extracted and used to compute pooled measurement uncertainty (PMU). The uncertainty of a given observation was recorded only if the original data were available or the publication reported a standard deviation (SD) or standard error (SE) for the treatment. When SD was reported, SE was calculated as $SE = \sqrt{\sigma_1^2/n_1}$. For a single study reporting N_2O flux (Zebarth et al., 2012) that reported uncertainty for the whole study but not for individual treatments, mean squared error across treatments was used as an estimate of variance within each treatment, and SE calculated as $SE = \sqrt{MSE/n_1}$ (Lajeunesse et al., 2013). Uncertainty estimates were not collected when they were reported solely as treatment differences including post-hoc ANOVA comparison tests. N_2O measurement uncertainties were extracted in a separate step after assembly of the mean flux dataset, so as a consistency check when calculating each N_2O measurement uncertainty, its associated mean seasonal flux was also recalculated. The resulting uncertainty estimate was used only if the mean was within 10 percent of the previously calculated mean seasonal flux in the validation dataset.

6.4 Category assignment

To assign datapoints to PC x CFG categories, we first paired individual measurements of emissions (cumulative N_2O flux or temporal change in SOC stock) into paired observations of the difference in emissions between two treatments measured at the same time, designating one treatment in each pair as the business-as-usual "baseline" and the other as the "changed practice." Only comparisons supported by the study design were used. We then assigned CFGs by considering all crops present during the measurement period in either treatment, and assigned PCs by considering all differences in management between the paired treatments. For example, when considering a 2x2 experiment comparing high and low fertilization on maize with and without a radish cover crop, we:

- Compared cover crops (yes vs no) within each fertilization level to obtain two observation pairs, both of which were assigned to both the CROP x C4A and CROP x C3A categories.
- Compared fertilization levels (high vs low) within each cover crop treatment to obtain two observation pairs: one FERT x C4A, one both FERT x C4A and FERT x C3A.

In the 2x2 case we did not consider the cross-treatment comparison between a high-N maize with no cover crop and low-N maize with cover crop. But if instead the experiment had compared "conventional" high-N-no-cover against "regenerative" low-N-with-cover with no ability to isolate the practices, this would be considered a stacked-practice (CROP and FERT) comparison under Section 3.3 Requirement 1 of the Model Requirements, and this one comparison would be included in all of the categories CROP x C4A, CROP x C3A, FERT x C4A, and FERT x C3A, after confirming as described above that each category also contained unstacked observations.

6.5 Validation dataset summary

Table 12 summarizes the information included in the validation dataset. Where information from multiple publications was combined for a single validation site, all publications used are included in the citation list for that site. When a study reported the effect of changing more than one practice at once with no ability to isolate the effects of each practice, the stacked observations were held out until the category contained at least one other validation study which reported the same effect in isolation. This was done to ensure that no category was validated solely against stacked practice studies, per Section 3.3 Requirement 1 of the Model Requirements.

Table 12. Summary of included SOC and N_2O data.

	SOC	**N_2O**
Number of sites usable for the calibration and validation process (see Figure 2)	48	79
Number of treatments	323	436
Number of measurements [a]	856	990
Number of practice change effect pairs [b]	1197	1381

[a.] A measurement that is a mean of replicates is counted as a single measurement

[b.] 1 SOC and 283 N_2O pairs were from PC x CFG combinations not validated in this report, and these pairs were included in the calibration runs (therefore allowing the final parameter set to be informed by these observations) but are not presented in this report.

For N_2O, an additional 48 treatments across 16 sites were identified from the literature but initial simulations produced simulation errors (e.g., nitrogen balance non-closure) from at least one observation. To align with the restrictions imposed on model use during crediting (Section 11), these treatments were not used for either calibration or validation and are not included in the observation counts in Table 12.

7 Bias evaluation

Follows Model Requirements Section 3.4

7.1 Calculating bias and pooled measurement uncertainty

In all categories, bias was computed for each study x PC x CFG combination as the mean difference between modeled and observed practice effects per Equation 4:

$$\text{bias} = \sum_{i=1}^{n} \frac{modeled_i - observed_i}{n}$$

(Equation 4)

where $observed_i$ is the observed difference and $modeled_i$ is the modeled difference in the relevant biogeochemical dynamic (SOC pool change or seasonal N_2O flux) between treatments in pair *i*, and n is the number of treatment pairs used from the study (per Equation 3.1 of the Model Requirements). When a study reported treatment pairs fitting multiple PC x CFG categories, only the observations matching the PC x CFG category of interest were included in the calculation.

In the case of SOC, while Equation 4 calculates bias at the second time point, it is identical to calculating bias in emission reductions between the first and second time point because measured and observed SOC at the first time point are always identical (the CAR SEP requires that modeled SOC be constrained to equal observed SOC at the first time point), and these values cancel out when subtracting $observed_i$ from $modeled_i$. When a treatment pair was measured at more than two time points, the starting year was taken as the first time point for all comparisons and each subsequent year was treated as a separate comparison. For example measurements in 2001, 2004, and 2007 are compared "3 and 6 years after 2001", not "3 years after 2001 and 3 years after 2004".

In the case of N_2O, the modeled and observed differences are calculated as the difference in cumulative flux, across the entire measured season, between the treatments in a pair. When a treatment pair was measured for more than one season, the starting date for later seasons was chosen to avoid overlap with the end date of earlier seasons.

Bias for each category was then computed as the unweighted mean of all per-study biases in that category, per Section 3.4 of the Model Requirements.

Bias was compared against the pooled measurement uncertainty (PMU) of the observed data. The PMU calculation prescribed by Section 3.4 Equation 3.2 of Model Requirements has been adapted as explained in Appendix F to handle the cases with unequal sample sizes (Equation 5):

$$\text{PMU} = \sqrt{\frac{\sum_{i=1}^{k} \sigma_i^2 (n_{i1} + n_{i2} - 2)}{\sum_{i=1}^{k} (n_{i1} + n_{i2} - 2)}}$$

(*Equation* 5)

where *k* is the number of observations with uncertainty reported, σ_i is the standard error of the i^{th} observation of differences between the treatments, and $n_{i1}$and n_{i2} are the number of replicates included in first and second study in the i^{th} treatment pair.

In the previous report (i.e., Indigo Ag, 2022), PMU was calculated for each CFG x PC x ES, but in this report a global PMU was calculated across all CFG x PC categories within each ES. This is primarily because there was no a-priori hypothesis that CFG or PC impacts measurement uncertainty. With as few as two sites reporting uncertainty in some CFG x PC groups, there was high variability in estimation of PMU, as would be expected from error due to small samples and without evidence that this was related to CFG or PC.

7.2 Example PMU calculation

In this example, we use a subset of the dataset including 14 pairs of observations representing changes in SOC under different cropping practice changes in corn (Table 13). These observations were reported alongside uncertainties for both treatments and therefore allow computing the standard error of the difference between treatments as $\sqrt{\sigma_1^2 + \sigma_2^2}$.

As described in Section 6, uncertainty was represented as standard error, transformed from SD where necessary. The needed summations over the product of degrees of freedom (df SE diff) and standard error (SE diff) to compute PMU are also shown in Table 13. This example calculation is shown for a single PC x CFG for brevity of the illustration.

Table 13. Computing pooled measurement uncertainty for CROP x corn from the observed standard errors of differences (SE in units of g C / m^2).

Site	n trt1	n trt2	SE trt1	SE trt2	SE diff	df SE diff	SE^2	SE^2 * df
hoytville	3	3	670	419	790	4	624,416	2,497,664
hoytville	3	3	419	37	421	4	176,904	707,616
hoytville	3	3	670	509	841	4	707,954	2,831,816
kbs	4	4	275	335	433	6	187,836	1,127,016
kbs	30	30	276	208	346	58	119,439	6,927,462
LaCage	4	4	420	660	782	6	611,993	3,671,958
LaCage	4	4	536	725	902	6	813,063	4,878,378
mead	4	4	455	288	539	6	289,982	1,739,892
wooster_SOC	3	3	171	215	275	4	75,460	301,840
wooster_SOC	3	3	154	80	174	4	30,102	120,408
wooster_SOC	3	3	127	215	250	4	62,350	249,400
wooster_SOC	3	3	179	80	196	4	38,455	153,820
wooster_SOC	3	3	127	171	213	4	45,369	181,476
wooster_SOC	3	3	179	154	236	4	55,743	222,972
sum	*NA*	*NA*	*NA*	*NA*	*NA*	118	*NA*	25,611,718
PMU	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	*NA*	466

7.3 Bias relative to PMU for SOC and N_2O

Table 14. Pooled measurement uncertainty of difference in SOC between treatments (g C m^{-2} and g N ha^{-1} across observation interval for SOC and N_2O, respectively).
n obs: Number of pairs of observations used in uncertainty computation. **n stacked**: number of observations taken from stacked PCs. **n sites**: Number of experimental sites the observation pairs were taken from. **Percent obs**: percentage of the observation pairs in the full dataset with uncertainty available. **Percent sites**: percentage of the sites in the full dataset with uncertainty available for at least one pair of observations. Number of sites, percent of observations, and percent of sites are not used in the PMU calculation but are presented to show the degree of data coverage.

Emissions source	n obs	n stacked	n sites	PMU	Percent obs	Percent sites
SOC	187	64	11	501	6%	23%
N_2O	675	84	27	829	23%	33%

7.4 Bias across all categories

Bias for each site and overall, organized by PC x CFG x Emission Source group are presented in Supporting Materials 2 and summarized in Table 15.

Table 15. Summary of bias calculations.

ES	PMU	PC	CFG	Bias	|Bias| < PMU
SOC	501	CROP	C3A	-54.5	TRUE
		CROP	C3AN	-54.3	TRUE
		CROP	C3PN	-62.5	TRUE
		CROP	C3S	-79.3	TRUE
		CROP	C4A	-31.2	TRUE
		DISTURB	C3A	-86.7	TRUE
		DISTURB	C3AN	-136	TRUE
		DISTURB	C3S	-183	TRUE
		DISTURB	C4A	-75	TRUE
		NFERT	C3A	11.2	TRUE
		NFERT	C3AN	-104	TRUE
		NFERT	C3PN	-184	TRUE
		NFERT	C4A	-98.2	TRUE
		ORG	ANNUALS	378	TRUE
		ORG	C3A	368	TRUE
		ORG	C3AN	217	TRUE
		ORG	C4A	59.4	TRUE
		ALL	ALL	28.6	TRUE
N_2O	829	CROP	C3A	-183	TRUE
		CROP	C3AN	114	TRUE
		CROP	C3PN	233	TRUE
		CROP	C4A	61.3	TRUE
		DISTURB	C3A	6.18	TRUE

ES	PMU	PC	CFG	Bias	|Bias| < PMU
		DISTURB	C3AN	-17.2	TRUE
		DISTURB	C4A	-211	TRUE
		NFERT	C3A	-349	TRUE
		NFERT	C3AN	-114	TRUE
		NFERT	C4A	231	TRUE
		ORG	C3A	-572	TRUE
		ORG	C4A	-468	TRUE
		ALL	ALL	55.8	TRUE

8 Model prediction error

Follows Model Requirements, Section 3.5

8.1 Description of calculation method

Model uncertainty bounds on the difference in emissions between the practice and the baseline scenarios were estimated using a Monte Carlo method as described in Gurung et al. (2020). In brief, the method takes draws from the posterior predictive distribution of the calibrated model (see Section 3). The posterior predictive draws account for uncertainty in DayCent calibration parameters, as well as errors in DayCent predictions due to variability across sites, across years within the same site, and unexplained errors. After all simulations are complete, the 90 percent posterior prediction intervals are calculated by taking the 5th and the 95th percentiles from the Monte Carlo simulations of the posterior predictive distribution, providing the central interval of the posterior prediction (Gelman et al., 2014, p. 33). The performance metric for acceptable model uncertainty is the percentage of measured observations from out-of-sample validation data that fall within the 90 percent posterior prediction interval.

Similar to the bias calculation described in Section 7, for each treatment pair the posterior prediction intervals are formed for the difference in SOC at the second time point and for the difference in cumulative N_2O flux across the measurement period. Coverage rates are then calculated as the proportion of these posterior prediction intervals that contain the observed difference emissions. As in Section 7.1, difference in SOC at the second time point is equivalent to emission reductions between the first and second time point, because modeled SOC is constrained to be equal to observed SOC at the first time point, so the values at the first time point cancel out when comparing two treatments.

Because the model is calibrated independently in each fold, and the folds have comparable predictor ranges (i.e. models calibrated with data from 4 folds are not extrapolating far outside their training data to validate the 5th hold-out fold), the average out-of-sample performance (i.e., bias and predictive interval coverage rates) across folds is a valid estimate of performance when the model is applied to new sites within the validated geographic, bioclimatic, and management domains outsides of the calibration dataset (Roberts et al., 2017).

The Bayesian approach used here complies with the Model Requirements criterion that the model uncertainty bounds of each prediction should account for cases “where there are few validation data” (Model Requirements, Section 3.5) and that they “account for data variability” (Model Requirements, Section 3.5). In particular, when data are more available and informative, the likelihood outweighs the prior and the choice of prior has diminishing effects on the posterior density. However, when there is not enough data or little information, the posterior tends to reproduce the prior. In this validation report weakly informative independent priors are used (as recommended in Model Requirements Section 3.5) that have a uniform distribution defined by their lower and upper bounds (see Appendix A and Section 3.3 for details). These uniform distributions are wide enough to expand beyond what is known or believed about the current understanding about the parameters’ range. For combinations of PC and CFG with little validation data or with observations that are highly variable, the method provides a conservative estimate of prediction error and can be improved in the future when additional datasets of higher quality are included.

8.2 Model prediction error across all categories

Supporting Materials 2 includes figures and tables of model prediction error for each PC x CFG x Emission Source group. For each group, there is a figure presenting measured versus modeled by fold and

by site; a figure presenting measured vs modeled overall, a histogram of residuals (predicted – observed), and a table including coverage overall and by fold. Table 16 summarizes the model prediction error.

Table 16. Model prediction error.

Emission Source	PC	CFG	RMSE	RMSE SD	Coverage (%)
SOC	CROP	C3A	1016	316	99.8
	CROP	C3AN	983	315	99.4
	CROP	C3PN	1161	356	100
	CROP	C3S	760	219	100
	CROP	C4A	1007	373	99.1
	DISTURB	C3A	935	299	100
	DISTURB	C3AN	971	403	100
	DISTURB	C3S	584	195	100
	DISTURB	C4A	1432	627	99.6
	NFERT	C3A	1117	446	99.6
	NFERT	C3AN	1101	396	100
	NFERT	C3PN	1250	270	100
	NFERT	C4A	1082	420	99.4
	ORG	ANNUALS	1562	1240	96.1
	ORG	C3A	1429	988	97.6
	ORG	C3AN	1116	280	100
	ORG	C4A	1104	257	100
	ALL	ALL	1016	316	99.3
N_2O	CROP	C3A	3997	3695	98.1
	CROP	C3AN	3558	3206	100
	CROP	C3PN	4878	3154	98.8
	CROP	C4A	5446	3537	93.9
	DISTURB	C3A	4041	4052	100
	DISTURB	C3AN	3434	4304	96
	DISTURB	C4A	4357	2500	100
	NFERT	C3A	3351	2106	100
	NFERT	C3AN	2586	1351	100
	NFERT	C4A	4070	1921	95.8
	ORG	C3A	4088	1918	97.6
	ORG	C4A	5774	2030	93.3
	ALL	ALL	4171	2695	97.4

9 Model validation outputs for use in CAR SEP uncertainty calculations

Follows Model Requirements, Section 3.5

When the model is used for crediting in project CAR1459 according to CAR SEP requirements, an uncertainty deduction will be computed using the methods described in CAR SEP Appendix D.2, using the final parameter set presented in Section 10. When running the model for crediting, the ensemble of simulations for a given datapoint will consist of 176 DayCent-CR simulations, which are then combined with draws from the random effect and residual variance distributions to give one posterior prediction for each unique combination of parameter states in the stored posterior. These are then summarized to quantify uncertainty in the estimate of total emission reductions for the project. These parameter sets are provided with supporting files in the file “thinned_posteriors_DayCentCR_1.1.0_val3.csv”. A comparison of full and thinned posteriors is provided in Appendix D.

10 Evaluation of final parameter set

After evaluating the model fitting procedure via 5-fold cross-validation, the final parameter set to be used for crediting was generated by applying the Bayesian calibration procedure to the entire dataset of observations with none held out. To obtain in-sample-predictions, random draws were taken of the random site and site-by-year random intercepts, which is aligned with our approach for making out-of-sample predictions; the best unbiased linear predictors were not used for the random intercepts.

The resulting posterior distributions from this final step are very similar to the distributions obtained during cross-validation and are saved for use when running the model for credits. Final parameters are summarized in Appendix A and parameters summarized by fold are in Appendix C. The performance of the model is reported here when fitting the validation data using the final parameter set (Table 17), but this is an evaluation against the training data and may not be representative of model performance at other sites. For an estimate of expected model performance at newly observed sites, the metrics computed from out-of-sample data during cross-validation are the correct metrics to use, and no other sections of this report are derived from models run with the final parameter set.

Table 17. Comparison of model performance across all PCs and CFGs between cross validation and final parameter sets. All statistics except coverage are in units of g C m^{-2} for SOC or g N ha^{-1} for N_2O.

	Bias	Coverage (%)	MSE (x 10^6)	RMSE
		SOC (g C m^{-2})		
Cross Validation	28.6	99.3	1.58	1140
Final Parameter Set	14.44	99.3	1.47	1125
		N_2O (g N ha^{-1})		
Cross Validation	55.8	97.4	24.7	4171
Final Parameter Set	76.7	97.3	23.0	4057

10.1 Model Performance with SOC parameter set

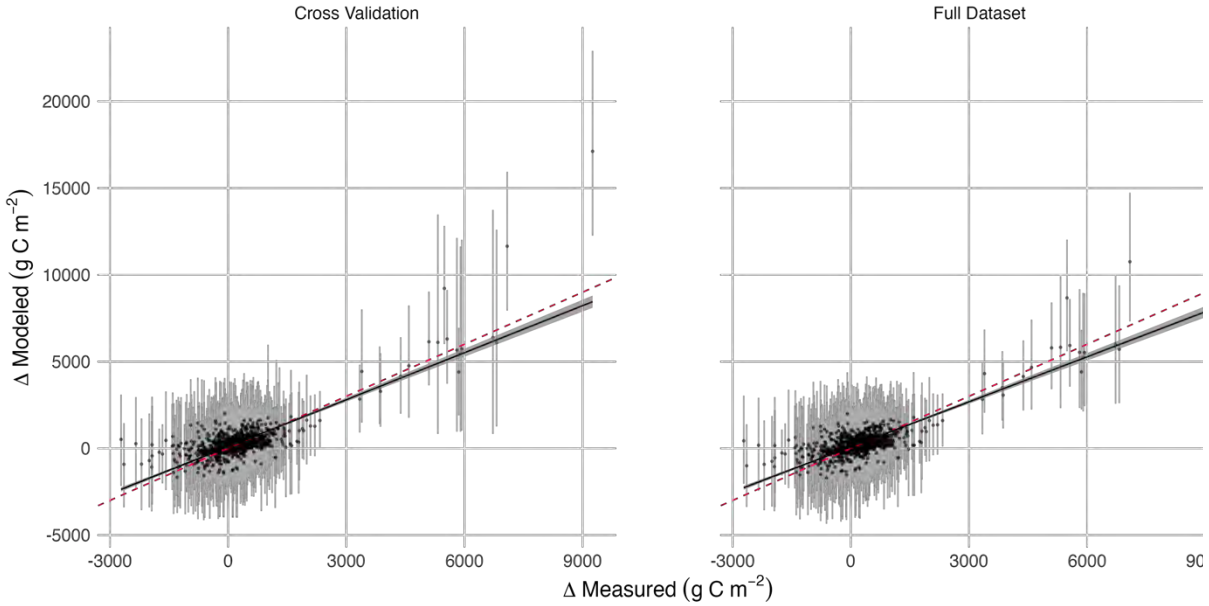

Figure 4. Model predictions versus measurements of SOC change in all practice categories and crop types, obtained during cross-validation (left image) and with final parameter set (right image). Error bars show 90 percent prediction intervals.

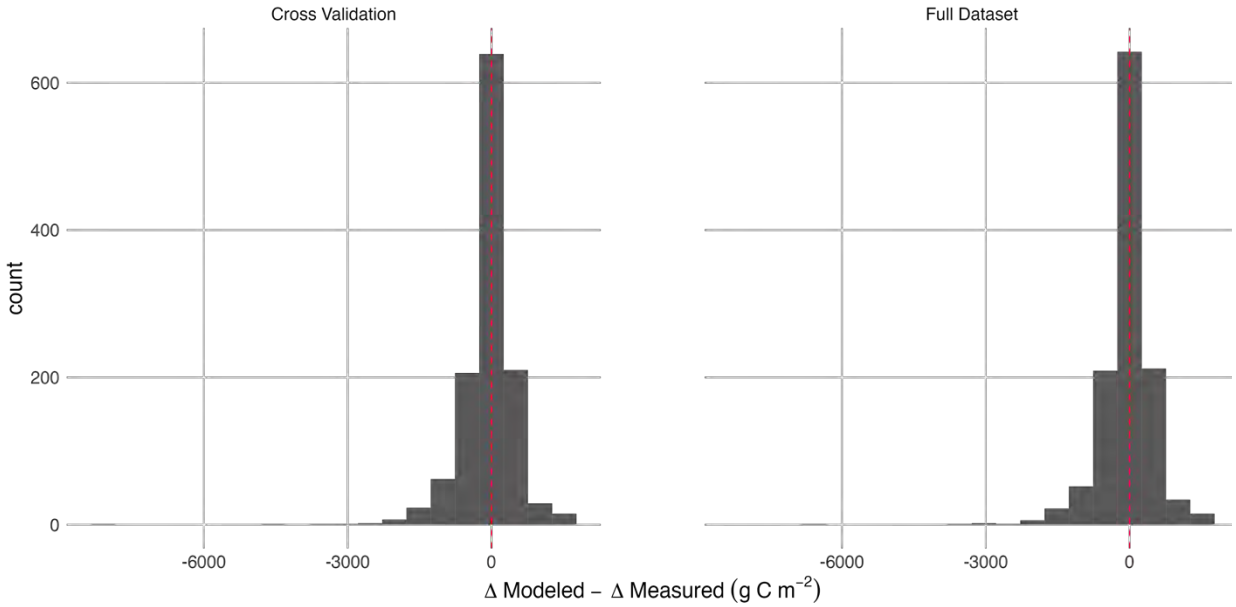

Figure 5. Histogram of model residuals (predicted - observed) for change in SOC in all studies used for model validation across all practices and crop types, obtained during cross-validation and with final parameter set.

Table 18. Comparison of model bias in each PC x CFG category during cross-validation and with the final parameter set.

PC	CFG	MSE x10^6 CV	MSE x10^6 Full	RMSE CV	RMSE Full	Full RMSE < CV RMSE?
CROP	C3A	1.1	1.1	1016	1025	FALSE
CROP	C3AN	1.1	1.1	983	986	FALSE
CROP	C3PN	1.5	1.3	1161	1105	TRUE
CROP	C3S	0.625	0.637	760	777	FALSE
CROP	C4A	1.2	1.1	1007	1002	TRUE
DISTURB	C3A	0.963	1.0	935	963	FALSE
DISTURB	C3AN	1.1	1.2	971	1005	FALSE
DISTURB	C3S	0.378	0.424	584	633	FALSE
DISTURB	C4A	2.4	2.4	1432	1434	FALSE
NFERT	C3A	1.5	1.3	1117	1076	TRUE
NFERT	C3AN	1.4	1.3	1101	1061	TRUE
NFERT	C3PN	1.6	1.5	1250	1195	TRUE
NFERT	C4A	1.4	1.3	1082	1089	FALSE
ORG	ANNUALS	4.0	2.6	1562	1348	TRUE
ORG	C3A	3.0	2.1	1429	1268	TRUE
ORG	C3AN	1.3	1.2	1116	1070	TRUE
ORG	C4A	1.3	1.2	1104	1067	TRUE
ALL	ALL	1.6	1.5	1140	1125	TRUE

10.2 Model Performance with N_2O parameter set

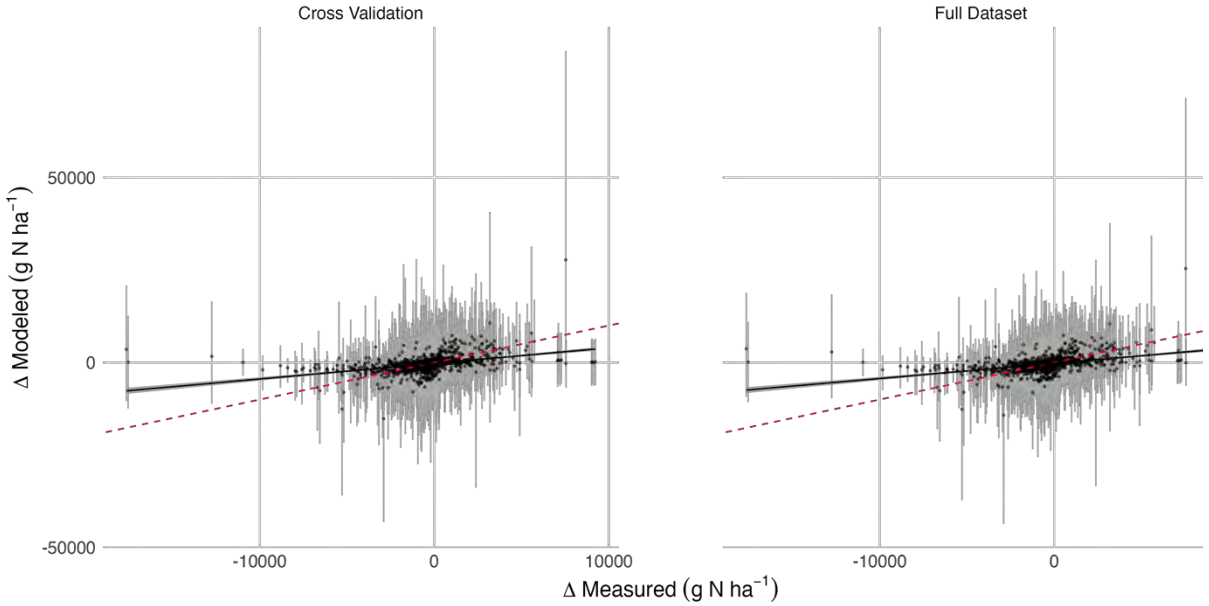

Figure 6. Model predictions versus measurements of N2O emission change in all practice categories and crop types, obtained during cross-validation (left image) and with final parameter set (right image). Error bars show 90 percent prediction intervals.

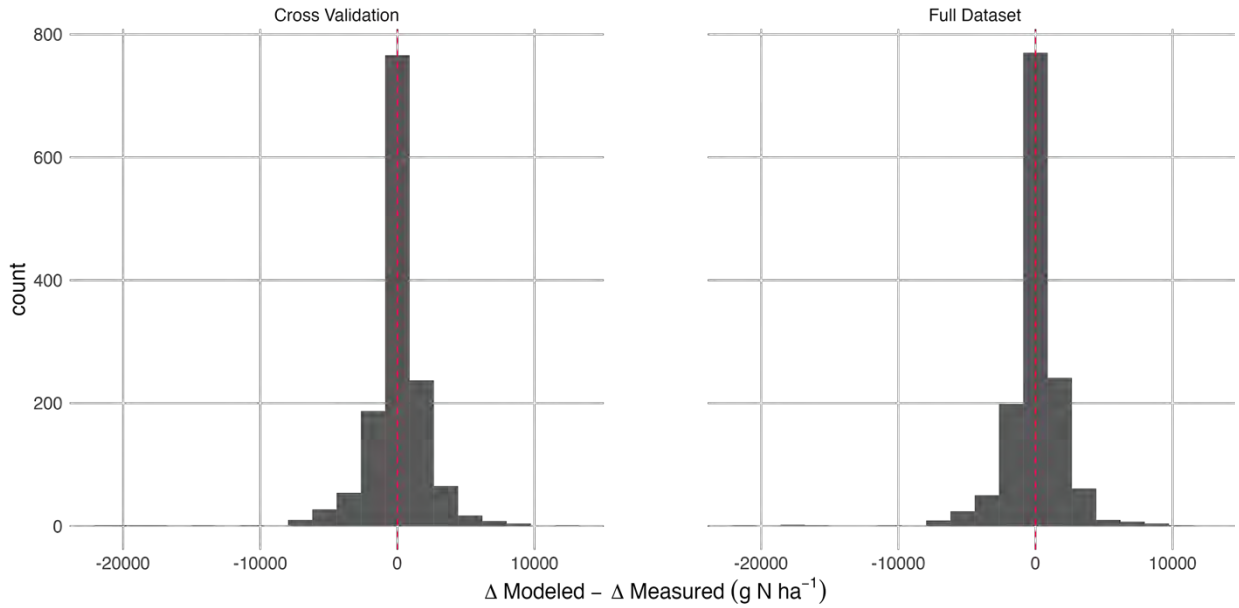

Figure 7. Histogram of model residuals (predicted - observed) for change in N_2O emission in all studies used for model validation across all practices and crop types, obtained during cross-validation and with final parameter set

Table 19. Comparison of model bias in each PC x CFG category during cross-validation and with the final parameter set.

PC	CFG	MSE x10^6 CV	MSE x10^6 Full	RMSE CV	RMSE Full	Full RMSE < CV RMSE?
CROP	C3A	29.5	30.0	3997	4005	FALSE
CROP	C3AN	22.8	19.2	3558	3405	TRUE
CROP	C3PN	33.6	33.3	4878	4826	TRUE
CROP	C4A	42.1	40.7	5446	5366	TRUE
DISTURB	C3A	32.5	33.1	4041	4065	FALSE
DISTURB	C3AN	29.6	23.3	3434	3213	TRUE
DISTURB	C4A	25.2	23.3	4357	4261	TRUE
NFERT	C3A	15.6	13.0	3351	3124	TRUE
NFERT	C3AN	8.4	8.3	2586	2561	TRUE
NFERT	C4A	20.2	18.7	4070	3955	TRUE
ORG	C3A	20.3	16.7	4088	3743	TRUE
ORG	C4A	37.4	36.4	5774	5714	TRUE
ALL	ALL	24.7	23.0	4171	4057	TRUE

11 Restrictions on application of model

N_2O flux is strongly driven by interactions between soil nitrogen pools and physical conditions such as temperature and soil water content, so large errors in simulation of these quantities are likely to mean modeled N_2O fluxes are unreliable (Parton et al., 2001). Therefore, as a quality control step, DayCent-CR's outputs are checked for acceptable nitrogen and water balance:

- Mineral soil nitrogen is typically a small proportion of total soil nitrogen (Schulten and Schnitzer, 1998), so modeled N_2O emissions will not be used for crediting when modeled mineral nitrogen exceeds 30 percent of total soil nitrogen, which exceeds the pool of even potentially mineralizable nitrogen in most soils (Nendel et al., 2019).
- Nitrogen losses to surface water and groundwater by runoff and leaching from cropland is typically between 5 to 50 kilograms per hectare in the United States (White et al., 2015). When DayCent-CR predicts nitrogen losses to water that exceed 100 kilograms per hectare the model will not be applied for N_2O crediting.
- Nitrogen cycling and transport is sensitive to hydrologic processes (e.g., soil moisture, infiltration), therefore modeled N_2O outcomes will not be applied when all water supplied through precipitation and/or irrigation is lost to evaporation and transpiration (i.e., no predicted runoff or infiltration) (Potter et al., 2006).

In previous validation reports, the model underestimated uncertainty under select combinations of conditions that were otherwise valid:

- DayCent 1.0 showed too little uncertainty for very large changes in SOC, and therefore restricted the valid range of the model to changes smaller than 5,000 g C m-2 (Indigo Ag, 2021).
- DayCent-CR 1.0.2 showed too little uncertainty at short timescales for only the ORG PC, and therefore applied a variance inflation factor of 1.36 to observations with organic amendments. (Indigo Ag, 2022).

In this report, uncertainty coverage of both SOC and N_2O appears adequate across both time (Appendix E) and effect size, so neither of these restrictions is needed for DayCent-CR 1.1.0.

It is important to note that, while the calibrated DayCent-CR model parameters generally fit well, some have edge-hitting tendencies and correlations that may limit the model's predictive accuracy, particularly when extrapolated beyond the calibration dataset. These behaviors indicate areas where the model could be improved, either by refining parameter priors or addressing potential structural limitations. As a result, predictions for SOC and N_2O should be limited to only CFG and practice changes that are directly validated in this report. Future model developments will aim to address these constraints to improve the model's robustness and expand its applicability.

12 References

Abdalla, M., Wattenbach, M., Smith, P., Ambus, P., Jones, M., & Williams, M. (2009). Application of the DNDC model to predict emissions of N2O from Irish agriculture. *Geoderma*, *151*(3), 327–337. https://doi.org/10.1016/j.geoderma.2009.04.021

Adviento-Borbe, M. A., Pittelkow, C. M., Anders, M., van Kessel, C., Hill, J. E., McClung, A. M., Six, J., & Linquist, B. A. (2013). Optimal Fertilizer Nitrogen Rates and Yield-Scaled Global Warming Potential in Drill Seeded Rice. *Journal of Environmental Quality*, *42*(6), 1623–1634. https://doi.org/10.2134/jeq2013.05.0167

Agresti, A., & Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician*, *52*(2), 119–126.

Allison, L. E. (1965). Organic carbon. In *Methods of soil analysis: Part 2 chemical and microbiological properties* (Vol. 9, pp. 1367–1378). John Wiley & Sons, Ltd.

Archer, D. W., & Halvorson, A. D. (2010). Greenhouse Gas Mitigation Economics for Irrigated Cropping Systems in Northeastern Colorado. *Soil Science Society of America Journal*, *74*(2), 446–452. https://doi.org/10.2136/sssaj2009.0080

Asgedom, H., Tenuta, M., Flaten, D. N., Gao, X., & Kebreab, E. (2014). Nitrous oxide emissions from a clay soil receiving granular urea formulations and dairy manure. *Agronomy Journal*, *106*(2), 732–744.

Autret, B., Mary, B., Chenu, C., Balabane, M., Girardin, C., Bertrand, M., Grandeau, G., & Beaudoin, N. (2016). Alternative arable cropping systems: A key to increase soil organic carbon storage? Results from a 16 year field experiment. *Agriculture, Ecosystems & Environment*, *232*, 150–164. https://doi.org/10.1016/j.agee.2016.07.008

Barré, P., Eglin, T., Christensen, B. T., Ciais, P., Houot, S., Kätterer, T., Oort, F. van, Peylin, P., Poulton, P., Romanenkov, V., & Chenu, C. (2010). Quantifying and isolating stable soil organic carbon using long-term bare fallow experiments. *Biogeosciences*, *7*(11), 3839–3850.

Barsotti, J. L. (2012). *Soil carbon and nitrogen and greenhouse gas emissions affected by sheep grazing under dryland cropping systems* [PhD Thesis]. Montana State University-Bozeman, College of Agriculture.

Barsotti, J. L., Sainju, U. M., Lenssen, A. W., Montagne, C., & Hatfield, P. G. (2013). Net Greenhouse Gas Emissions Affected by Sheep Grazing in Dryland Cropping Systems. *Soil Science Society of America Journal*, *77*(3), 1012–1025. https://doi.org/10.2136/sssaj2012.0386

Bista, P., Machado, S., Ghimire, R., Del Grosso, S. J., & Reyes-Fox, M. (2016). Simulating Soil Organic Carbon in a Wheat-Fallow System Using the Daycent Model. *Agronomy Journal*, *108*(6), 2554–2565. https://doi.org/10.2134/agronj2016.04.0202

Blevins, R. L., Thomas, G. W., Smith, M. S., Frye, W. W., & Cornelius, P. L. (1983). Changes in Soil Properties after 10 Years Continuous Non-Tilled and Conventionally Tilled Corn. *Soil and Tillage Research*, *3*(2), 135–146. https://doi.org/10.1016/0167-1987(83)90004-1

Campbell, B. D. (2012). *Carbon budgets and greenhouse gas emissions associated with two long-term tillage and crop rotation sites in Ohio* [The Ohio State University]. https://etd.ohiolink.edu/acprod/odb_etd/etd/r/1501/10?clear=10&p10_accession_num=osu1354559256

Campbell, C. A., VandenBygaart, A. J., Zentner, R. P., McConkey, B. G., Smith, W., Lemke, R., Grant, B., & Jefferson, P. G. (2007). Quantifying Carbon Sequestration in a Minimum Tillage Crop Rotation Study in Semiarid Southwestern Saskatchewan. *Canadian Journal of Soil Science*, *87*(3), 235–250. https://doi.org/10.4141/S06-018

Campbell, C. A., & Zentner, R. P. (1997). Crop production and soil organic matter in long-term crop rotations in the sub-humid northern Great Plains of Canada. In *Soil organic matter in temperate agroecosystems: Long-Term Experiments in North America* (pp. 317–334). CRC Press.

Carlin, B. P., & Louis, T. A. (2009). *Bayesian methods for data analysis* (3rd edition). CRC Press.

Casella, G. (1985). An introduction to empirical Bayes data analysis. *The American Statistician*, *39*(2), 83–87.

Chang, C., Janzen, H. H., & Cho, C. M. (1998). Nitrous Oxide Emission From Long-Term Manured Soils. *Soil Science Society of America Journal*, *62*(3), 677–682. https://doi.org/10.2136/sssaj1998.03615995006200030019x

Chantigny, M. H., Rochette, P., Angers, D. A., Goyer, C., Brin, L. D., & Bertrand, N. (2016). Nongrowing season N2O and CO2 emissions—Temporal dynamics and influence of soil texture and fall-applied manure. *Canadian Journal of Soil Science*, *97*(3), 452–464.

Chapuis-Lardy, L., Wrage, N., Metay, A., Chotte, J., & Bernoux, M. (2007). Soils, a sink for N2O? A review. *Global Change Biology*, *13*(1), 1–17. https://doi.org/10.1111/j.1365-2486.2006.01280.x

Christenson, D. R. (1997). Soil organic matter in sugar beet and dry bean cropping systems in Michigan. In *Soil Organic Matter in Temperate Agroecosystems: Long-Term Experiments in North America* (pp. 151–159). CRC Press.

Clapp, C. E., Allmaras, R. R., Layese, M. F., Linden, D. R., & Dowdy, R. H. (2000). Soil Organic Carbon and 13C Abundance as Related to Tillage, Crop Residue, and Nitrogen Fertilization under Continuous Corn Management in Minnesota. *Soil and Tillage Research*, *55*(3–4), 127–142. https://doi.org/10.1016/S0167-1987(00)00110-0

Clark, M. S., Horwath, W. R., Shennan, C., & Scow, K. M. (1998). Changes in Soil Chemical Properties Resulting from Organic and Low-Input Farming Practices. *Agronomy Journal*, *90*(5), 662–671. https://doi.org/10.2134/agronj1998.00021962009000050016x

Clayton, H., McTaggart, I. P., Parker, J., Swan, L., & Smith, K. A. (1997). Nitrous oxide emissions from fertilised grassland: A 2-year study of the effects of N fertiliser form and environmental conditions. *Biology and Fertility of Soils*, *25*(3), 252–260. https://doi.org/10.1007/s003740050311

Climate Action Reserve. (2022). *Soil Enrichment Protocol Version 1.1* (1.1). Climate Action Reserve. https://www.climateactionreserve.org/wp-content/uploads/2023/07/Soil-Enrichment-Protocol-V_1.1-final-1.pdf

Collins, H. P., Christenson, D. R., Blevins, R. L., Bundy, L. G., Dick, W. A., Huggins, D. R., & Paul, E. A. (1999). Soil Carbon Dynamics in Corn-Based Agroecosystems: Results from Carbon-13 Natural Abundance. *Soil Science Society of America Journal*, *63*(3), 584–591. https://doi.org/10.2136/sssaj1999.03615995006300030022x

Collins, H. P., Elliott, E. T., Paustian, K., Bundy, L. G., Dick, W. A., Huggins, D. R., Smucker, A. J. M., & Paul, E. A. (2000). Soil Carbon Pools and Fluxes in Long-Term Corn Belt Agroecosystems. *Soil Biology*, 12.

Collins, H. P., Streubel, J. D., Frear, C., Chen, S., Granatstein, D., Kruger, C., Alva, A. K., & Fransen, S. F. (2010). Application of AD dairy manure effluent to fields and associated impacts. *Cent Sustain Agric Nat Resour (CSANR), CSANR Res Rep*, 2010–001.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge University Press.

Del Grosso, S. J., Parton, W. J., Adler, P. R., Davis, S. C., Keough, C., & Marx, E. (2012). DayCent model simulations for estimating soil carbon dynamics and greenhouse gas fluxes from agricultural production systems. *Managing Agricultural Greenhouse Gases: Coordinated Agricultural Research through GRACEnet to Address Our Changing Climate*, 241–250.

Del Grosso, S. J., Parton, W. J., Mosier, A. R., Walsh, M. K., Ojima, D. S., & Thornton, P. E. (2006). DAYCENT National-Scale Simulations of Nitrous Oxide Emissions from Cropped Soils in the United States. *Journal of Environmental Quality*, *35*(4), 1451–1460. https://doi.org/10.2134/jeq2005.0160

Denef, K., Stewart, C. E., Brenner, J., & Paustian, K. (2008). Does Long-Term Center-Pivot Irrigation Increase Soil Carbon Stocks in Semi-Arid Agro-Ecosystems? *Geoderma*, *145*(1–2), 121–129. https://doi.org/10.1016/j.geoderma.2008.03.002

Dick, W. A., Edwards, W. M., & McCoy, E. L. (1997). Continuous application of no-tillage to Ohio soils: Changes in crop yields and organic matter-related soil properties. In *Soil Organic Matter in Temperate Agroecosystems: Long-Term Experiments in North America* (pp. 171–182). CRC Press.

Dick, W. A., Triplett, G. B., & Doren, D. M. V. (2013). *Triplett-Van Doren Long-term Tillage and Crop Rotation Data (1962-2012)*. Ohio Agricultural Research and Development Center. http://hdl.handle.net/1811/55716/

Dijkstra, F. A., Pendall, E., Mosier, A. R., King, J. Y., Milchunas, D. G., & Morgan, J. A. (2008). Long-term enhancement of N availability and plant growth under elevated CO2 in a semi-arid grassland. *Functional Ecology*, *22*(6), 975–982.

Ding, W., Cai, Y., Cai, Z., Yagi, K., & Zheng, X. (2007). Nitrous oxide emissions from an intensively cultivated maize–wheat rotation soil in the North China Plain. *Science of The Total Environment*, *373*(2), 501–511. https://doi.org/10.1016/j.scitotenv.2006.12.026

Dolan, M. S., Clapp, C. E., Allmaras, R. R., Baker, J. M., & Molina, J. A. E. (2006). Soil Organic Carbon and Nitrogen in a Minnesota Soil as Related to Tillage, Residue and Nitrogen Management. *Soil and Tillage Research*, *89*(2), 221–231. https://doi.org/10.1016/j.still.2005.07.015

Duan, Q. Y., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, *76*(3), 501–521. https://doi.org/10.1007/BF00939380

Dungan, R. S., Leytem, A. B., Tarkalson, D. D., Ippolito, J. A., & Bjorneberg, D. L. (2017). Greenhouse Gas Emissions from an Irrigated Dairy Forage Rotation as Influenced by Fertilizer and Manure Applications. *Soil Science Society of America Journal*, *81*(3), 537–545. https://doi.org/10.2136/sssaj2016.08.0254

Dusenbury, M. P., Engel, R. E., Miller, P. R., Lemke, R. L., & Wallander, R. (2008). Nitrous Oxide Emissions from a Northern Great Plains Soil as Influenced by Nitrogen Management and Cropping Systems. *Journal of Environmental Quality*, *37*(2), 542–550. https://doi.org/10.2134/jeq2006.0395

Elliott, E. T., Burke, I. C., Monz, C. A., Frey, S. D., Paustian, K. H., Collins, H. P., Paul, E. A., Cole, C. V., Blevins, R. L., Frye, W. W., Lyon, D. J., Halvorson, A. D., Huggins, D. R., Turco, R. F., & Hickman, M. V. (1994). Terrestrial Carbon Pools: Preliminary Data from the Corn Belt and Great Plains Regions. In J. W. Doran, D. C. Coleman, D. F. Bezdicek, & B. A. Stewart (Eds.), *SSSA Special Publications* (pp. 179–191). Soil Science Society of America and American Society of Agronomy. https://doi.org/10.2136/sssaspecpub35.c12

Ferchaud, F., Vitte, G., & Mary, B. (2016). Changes in soil carbon stocks under perennial and annual bioenergy crops. *GCB Bioenergy*, *8*(2), 290–306. https://doi.org/10.1111/gcbb.12249

Ferreira, A. C. B., Borin, A. L. D. C., Lamas, F. M., Bogiani, J. C., Silva, M. A. S., Silva Filho, J. L., & Staut, L. A. (2019). Soil carbon accumulation in cotton production systems in the Brazilian Cerrado. *Acta Scientiarum. Agronomy*, *42*.

Flessa, H., Dörsch, P., & Beese, F. (1995). Seasonal variation of N2O and CH4 fluxes in differently managed arable soils in southern Germany. *Journal of Geophysical Research: Atmospheres*, *100*(D11), 23115–23124.

Forseth, I. N. (2010). *The Ecology of Photosynthetic Pathways | Learn Science at Scitable*. https://www.nature.com/scitable/knowledge/library/the-ecology-of-photosynthetic-pathways-15785165/

Gao, X., Asgedom, H., Tenuta, M., & Flaten, D. N. (2015). Enhanced Efficiency Urea Sources and Placement Effects on Nitrous Oxide Emissions. *Agronomy Journal*, *107*(1), 265–277. https://doi.org/10.2134/agronj14.0213

Gelfand, I., Shcherbak, I., Millar, N., Kravchenko, A. N., & Robertson, G. P. (2016). Long-term nitrous oxide fluxes in annual and perennial agricultural and unmanaged ecosystems in the upper Midwest USA. *Global Change Biology*, *22*(11), 3594–3607. https://doi.org/10.1111/gcb.13426

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis* (3rd ed.). CRC press.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, *7*(4), 457–472.

Ghimire, R., Machado, S., & Bista, P. (2017). Soil pH, Soil Organic Matter, and Crop Yields in Winter Wheat-Summer Fallow Systems. *Agronomy Journal*, *109*(2), 706–717. https://doi.org/10.2134/agronj2016.08.0462

Ghimire, R., Machado, S., & Rhinhart, K. (2015). Long-Term Crop Residue and Nitrogen Management Effects on Soil Profile Carbon and Nitrogen in Wheat-Fallow Systems. *Agronomy Journal*, *107*(6), 2230–2240. https://doi.org/10.2134/agronj14.0601

Graler, B., Pebesma, E., & Heuvelink, G. (2016). Spatio-Temporal Interpolation using gstat. *The R Journal*, *8*(1), 204–218.

Grandy, A. S., Loecke, T. D., Parr, S., & Robertson, G. P. (2006). Long-Term Trends in Nitrous Oxide Emissions, Soil Nitrogen, and Crop Yields of Till and No-Till Cropping Systems. *Journal of Environmental Quality*, *35*(4), 1487–1495. https://doi.org/10.2134/jeq2005.0166

Guan, X.-K., Turner, N. C., Song, L., Gu, Y.-J., Wang, T.-C., & Li, F.-M. (2016). Soil carbon sequestration by three perennial legume pastures is greater in deeper soil layers than in the surface soil. *Biogeosciences*, *13*(2), 527–534. https://doi.org/10.5194/bg-13-527-2016

Guillaume, J., & Andrews, F. (2012). *dream: DiffeRential Evolution Adaptive Metropolis*. http://CRAN.R-project.org/package=dream

Gurung, R. B., Ogle, S. M., Breidt, F. J., Parton, W. J., Del Grosso, S. J., Zhang, Y., Hartman, M. D., Williams, S. A., & Venterea, R. T. (2021). Modeling nitrous oxide mitigation potential of enhanced efficiency nitrogen fertilizers from agricultural systems. *Science of The Total Environment*, *801*, 149342. https://doi.org/10.1016/j.scitotenv.2021.149342

Gurung, R. B., Ogle, S. M., Breidt, F. J., Williams, S. A., & Parton, W. J. (2020). Bayesian calibration of the DayCent ecosystem model to simulate soil organic carbon dynamics and reduce model uncertainty. *Geoderma*, *376*(June), 114529. https://doi.org/10.1016/j.geoderma.2020.114529

Halvorson, A. D., Del Grosso, S. J., & Stewart, C. E. (2016). Manure and Inorganic Nitrogen Affect Trace Gas Emissions under Semi-Arid Irrigated Corn. *Journal of Environmental Quality*, *45*(3), 906–914. https://doi.org/10.2134/jeq2015.08.0426

Halvorson, A. D., Follett, R. F., Reule, C. A., & Del Grosso, S. (2009). Soil organic carbon and nitrogen sequestration in irrigated cropping systems of the Central Great Plains. *Soil Carbon Sequestration and the Greenhouse Effect*, *57*, 141–157.

Halvorson, A. D., & Jantalia, C. P. (2011). Nitrogen Fertilization Effects on Irrigated No-Till Corn Production and Soil Carbon and Nitrogen. *Agronomy Journal*, *103*(5), 1423–1431. https://doi.org/10.2134/agronj2011.0102

Halvorson, A. D., & Schlegel, A. J. (2012). Crop rotation effect on soil carbon and nitrogen stocks under limited irrigation. *Agronomy Journal*, *104*(5), 1265–1273.

Halvorson, A. D., Stewart, C. E., & Del Grosso, S. J. (2016). Manure and Inorganic Nitrogen Affect Irrigated Corn Yields and Soil Properties. *Agronomy Journal*, *108*(2), 519–531. https://doi.org/10.2134/agronj2015.0402

Halvorson, A. D., Wienhold, B. J., & Black, A. L. (2002). Tillage, Nitrogen, and Cropping System Effects on Soil Carbon Sequestration. *Soil Science Society of America Journal*, *66*(3), 906–912. https://doi.org/10.2136/sssaj2002.9060

Hao, X., Chang, C., Travis, G. R., & Zhang, F. (2003). Soil Carbon and Nitrogen Response to 25 Annual Cattle Manure Applications. *Journal of Plant Nutrition and Soil Science*, *166*(2), 239–245. https://doi.org/10.1002/jpln.200390035

Hellebrand, H. J., Kern, J., & Scholz, V. (2003). Long-term studies on greenhouse gas fluxes during cultivation of energy crops on sandy soils. *Atmospheric Environment*, *37*(12), 1635–1644. https://doi.org/10.1016/S1352-2310(03)00015-3

Hellebrand, H., Scholz, V., & Kern, J. (2008). Fertiliser induced nitrous oxide emissions during energy crop cultivation on loamy sand soils. *Atmospheric Environment*, *42*(36), 8403–8411. https://doi.org/10.1016/j.atmosenv.2008.08.006

Hoben, J. P., Gehl, R. J., Millar, N., Grace, P. R., & Robertson, G. P. (2011). Nonlinear nitrous oxide (N2O) response to nitrogen fertilizer in on-farm corn crops of the US Midwest. *Global Change Biology*, *17*(2), 1140–1152. https://doi.org/10.1111/j.1365-2486.2010.02349.x

Horwath, W. R., & Burger, M. (2012). *Assessment of baseline nitrous oxide emissions in California cropping systems*. California Air Resources Board.

Houska, T., Kraft, P., Chamorro-Chavez, A., & Breuer, L. (2015). SPOTting model parameters using a ready-made python package. *PloS One*, *10*(12), e0145180.

Hulugalle, N. R., Weaver, T. B., Finlay, L. A., & Heimoana, V. (2013). Soil organic carbon concentrations and storage in irrigated cotton cropping systems sown on permanent beds in a Vertosol with restricted subsoil drainage. *Crop and Pasture Science*, *64*(8), 799. https://doi.org/10.1071/CP12374

Hunt, D., Bittman, S., Chantigny, M., & Lemke, R. (2019). Year-Round N2O Emissions From Long-Term Applications of Whole and Separated Liquid Dairy Slurry on a Perennial Grass Sward and Strategies for Mitigation. *Frontiers in Sustainable Food Systems*, *3*. https://www.frontiersin.org/articles/10.3389/fsufs.2019.00086

Hyatt, C. R., Venterea, R. T., Rosen, C. J., McNearney, M., Wilson, M. L., & Dolan, M. S. (2010). Polymer-Coated Urea Maintains Potato Yields and Reduces Nitrous Oxide Emissions in a Minnesota Loamy Sand. *Soil Science Society of America Journal*, *74*(2), 419–428. https://doi.org/10.2136/sssaj2009.0126

Indigo Ag. (2021). *Validation Report for DayCent-CR Version 1.0* (CAR1459). Climate Action Reserve. https://www.climateactionreserve.org/wp-content/uploads/2022/11/CAR1459_model_val_DayCentCR_1.0.pdf

Indigo Ag. (2022). *Validation Report DayCent-CR Version 1.0.2* (CAR1459). Climate Action Reserve. https://www.climateactionreserve.org/wp-content/uploads/2022/11/CAR1459_model_val_DayCentCR_1.0.2.pdf

IPCC. (2019). *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*. Intergovernmental Panel on Climate Change, Switzerland.

Ismail, I., Blevins, R. L., & Frye, W. W. (1994). Long-Term No-tillage Effects on Soil Properties and Continuous Corn Yields. *Soil Science Society of America Journal*, *58*(1), 193–198. https://doi.org/10.2136/sssaj1994.03615995005800010028x

Janzen, H. H., Johnston, A., Carefoot, J., & Lindwall, C. (1997). Soil organic matter dynamics in long-term experiments in Southern Alberta. In *Soil Organic Matter in Temperate Agroecosystems: Long-Term Experiments in North America* (pp. 283–296). CRC Press.

Jarecki, M. K., & Lal, R. (2005). Soil organic carbon sequestration rates in two long-term no-till experiments in Ohio. *Soil Science*, *170*(4), 280–291.

Jin, V. L., Schmer, M. R., Wienhold, B. J., Stewart, C. E., Varvel, G. E., Sindelar, A. J., Follett, R. F., Mitchell, R. B., & Vogel, K. P. (2015). Twelve Years of Stover Removal Increases Soil Erosion Potential

without Impacting Yield. *Soil Science Society of America Journal*, *79*(4), 1169–1178. https://doi.org/10.2136/sssaj2015.02.0053

Jin, V., & Varvel, G. (2018a). *All; 2001–2016. Ver. GPSR_NATRES. Fort Collins, CO: USDA-ARS Natural Resources Database. Project: NEMEIRR*. https://gpsr.ars.usda.gov/NEMEIRR

Jin, V., & Varvel, G. (2018b). *All; 1998–2016. Ver. GPSR_NATRES. Fort Collins, CO: USDA-ARS Natural Resources Database. Project: NEMERREM*. https://gpsr.ars.usda.gov/NEMERREM

Johnson, J. M. F., & Barbour, N. W. (2016). Nitrous Oxide Emission and Soil Carbon Sequestration from Herbaceous Perennial Biofuel Feedstocks. *Soil Science Society of America Journal*, *80*(4), 1057–1070. https://doi.org/10.2136/sssaj2015.12.0436

Johnson, J. M.-F., Archer, D. W., Weyers, S. L., & Barbour, N. W. (2011). Do mitigation strategies reduce global warming potential in the northern US corn belt? *Journal of Environmental Quality*, *40*(5), 1551–1559.

Johnston, A., Poulton, P., Coleman, K., Macdonald, A., & White, R. (2017). Changes in soil organic matter over 70 years in continuous arable and ley–arable rotations on a sandy loam soil in England. *European Journal of Soil Science*, *68*(3), 305–316. https://doi.org/10.1111/ejss.12415

Kelly, K. B., Phillips, F. A., Baigent, R., Kelly, K. B., Phillips, F. A., & Baigent, R. (2008). Impact of dicyandiamide application on nitrous oxide emissions from urine patches in northern Victoria, Australia. *Australian Journal of Experimental Agriculture*, *48*(2), 156–159. https://doi.org/10.1071/EA07251

Kennedy, T., Decock, C., & Six, J. (2013). Assessing drivers of N2O production in California tomato cropping systems. *Science of The Total Environment*, *465*, 36–47. https://doi.org/10.1016/j.scitotenv.2013.04.014

Kessavalou, A., Mosier, A. R., Doran, J. W., Drijber, R. A., Lyon, D. J., & Heinemeyer, O. (1998). Fluxes of Carbon Dioxide, Nitrous Oxide, and Methane in Grass Sod and Winter Wheat-Fallow Tillage Management. *Journal of Environmental Quality*, *27*(5), 1094–1104. https://doi.org/10.2134/jeq1998.00472425002700050015x

Khan, S. A., Mulvaney, R. L., Ellsworth, T. R., & Boast, C. W. (2007). The Myth of Nitrogen Fertilization for Soil Carbon Sequestration. *Journal of Environmental Quality*, *36*(6), 1821–1832. https://doi.org/10.2134/jeq2007.0099

Kong, A. Y. Y., Six, J., Bryant, D. C., Denison, R. F., & van Kessel, C. (2005). The Relationship between Carbon Input, Aggregation, and Soil Organic Carbon Stabilization in Sustainable Cropping Systems. *Soil Science Society of America Journal*, *69*(4), 1078–1085. https://doi.org/10.2136/sssaj2004.0215

Krauss, M., Ruser, R., Müller, T., Hansen, S., Mäder, P., & Gattinger, A. (2017). Impact of reduced tillage on greenhouse gas emissions and soil carbon stocks in an organic grass-clover ley—Winter wheat cropping sequence. *Agriculture, Ecosystems & Environment*, *239*, 324–333. https://doi.org/10.1016/j.agee.2017.01.029

Krauss, M., Wiesmeier, M., Don, A., Cuperus, F., Gattinger, A., Gruber, S., Haagsma, W. K., Peigné, J., Palazzoli, M. C., Schulz, F., Van Der Heijden, M. G. A., Vincent-Caboud, L., Wittwer, R. A., Zikeli, S., & Steffens, M. (2022). Reduced tillage in organic farming affects soil organic carbon stocks in temperate Europe. *Soil and Tillage Research*, *216*, 105262. https://doi.org/10.1016/j.still.2021.105262

Kuhn, M., & Johnson, K. (2013). Over-fitting and model tuning. In *Applied predictive modeling* (pp. 61–92). Springer.

Lajeunesse, M. J., Koricheva, J., Gurevitch, J., & Mengersen, K. (2013). Recovering missing or partial data from studies: A survey of conversions and imputations for meta-analysis. *Handbook of Meta-Analysis in Ecology and Evolution*, 195–206.

Lehman, R. M., & Osborne, S. L. (2013). Greenhouse gas fluxes from no-till rotated corn in the upper midwest. *Agriculture, Ecosystems & Environment*, *170*, 1–9. https://doi.org/10.1016/j.agee.2013.02.009

Leytem, A. B., Moore, A. D., & Dungan, R. S. (2019). Greenhouse Gas Emissions from an Irrigated Crop Rotation Utilizing Dairy Manure. *Soil Science Society of America Journal*, *83*(1), 137–152. https://doi.org/10.2136/sssaj2018.06.0216

Li, C., Six, J., Horwath, W. R., & Salas, W. (2014). *Calibrating, validating, and implementing process models for California agriculture greenhouse gas emissions*. California Environmental Protection Agency, Air Resources Board, Research ….

Liebig, M. A., Gross, J. R., Kronberg, S. L., & Phillips, R. L. (2010). Grazing Management Contributions to Net Global Warming Potential: A Long-term Evaluation in the Northern Great Plains. *Journal of Environmental Quality*, *39*(3), 799–809. https://doi.org/10.2134/jeq2009.0272

Liebig, M. A., Tanaka, D. L., & Gross, J. R. (2010). Fallow Effects on Soil Carbon and Greenhouse Gas Flux in Central North Dakota. *Soil Science Society of America Journal*, *74*(2), 358–365.

Liu, C., Zheng, X., Zhou, Z., Han, S., Wang, Y., Wang, K., Liang, W., Li, M., Chen, D., & Yang, Z. (2010). Nitrous oxide and nitric oxide emissions from an irrigated cotton field in Northern China. *Plant and Soil*, *332*(1/2), 123–134.

Lu, H. (2016). *The influence of land management and simulated climate change on N2O and CH4 exchange of montane grassland soils* [PhD Thesis]. Albert-Ludwigs-Universität Freiburg im Breisgau.

Maharjan, B., & Venterea, R. T. (2013). Nitrite intensity explains N management effects on N2O emissions in maize. *Soil Biology and Biochemistry*, *66*, 229–238. https://doi.org/10.1016/j.soilbio.2013.07.015

Mathers, C., Black, C. K., Segal, B. D., Gurung, R. B., Zhang, Y., Easter, M. J., Williams, S., Motew, M., Campbell, E. E., & Brummitt, C. D. (2023). Validating DayCent-CR for cropland soil carbon offset reporting at a national scale. *Geoderma*, *438*, 116647.

McSwiney, C. P., & Robertson, G. P. (2005). Nonlinear response of N2O flux to incremental fertilizer addition in a continuous maize (Zea mays L.) cropping system. *Global Change Biology*, *11*(10), 1712–1719. https://doi.org/10.1111/j.1365-2486.2005.01040.x

Mebius, L. J. (1960). A rapid method for the determination of organic carbon in soil. *Analytica Chimica Acta*, *22*, 120–124.

Mestelan, S. A. (2008). *Impact of Long-Term No till and Plow till on Soil Properties and Soil Nutrient Cycling* [Phd thesis, Ohio State University]. http://rave.ohiolink.edu/etdc/view?acc_num=osu1199221756

Migliorati, M. D. A., Scheer, C., Grace, P. R., Rowlings, D. W., Bell, M., & McGree, J. (2014). Influence of different nitrogen rates and DMPP nitrification inhibitor on annual N2O emissions from a subtropical wheat–maize cropping system. *Agriculture, Ecosystems & Environment*, *186*, 33–43.

Mikha, M. M., Vigil, M. F., & Benjamin, J. G. (2013). Long-Term Tillage Impacts on Soil Aggregation and Carbon Dynamics under Wheat-Fallow in the Central Great Plains. *Soil Science Society of America Journal*, *77*(2), 594–605. https://doi.org/10.2136/sssaj2012.0125

Mitchell, J. P., Shrestha, A., Horwath, W. R., Southard, R. J., Madden, N., Veenstra, J., & Munk, D. S. (2015). Tillage and Cover Cropping Affect Crop Yields and Soil Carbon in the San Joaquin Valley, California. *Agronomy Journal*, *107*(2), 588–596. https://doi.org/10.2134/agronj14.0415

Mitchell, J. P., Shrestha, A., Mathesius, K., Scow, K. M., Southard, R. J., Haney, R. L., Schmidt, R., Munk, D. S., & Horwath, W. R. (2017). Cover Cropping and No-Tillage Improve Soil Health in an Arid Irrigated Cropping System in California's San Joaquin Valley, USA. *Soil and Tillage Research*, *165*, 325–335. https://doi.org/10.1016/j.still.2016.09.001

Monreal, C. M., & Janzen, H. H. (1993). Soil Organic-Carbon Dynamics after 80 Years of Cropping a Dark Brown Chernozem. *Canadian Journal of Soil Science*, *73*(1), 133–136. https://doi.org/10.4141/cjss93-014

Morris, C. K. (2019). *Hydrologic Drivers of Nitrogen Cycling: Implications for an Earth System Model and Agricultural Management Practices*. Cornell University.

Mosier, A. R., Guenzi, W. D., & Schweizer, E. E. (1986). Soil Losses of Dinitrogen and Nitrous Oxide from Irrigated Crops in Northeastern Colorado. *Soil Science Society of America Journal*, *50*(2), 344–348. https://doi.org/10.2136/sssaj1986.03615995005000020018x

Mosier, A. R., Halvorson, A. D., Peterson, G. A., Robertson, G. P., & Sherrod, L. (2005). Measurement of Net Global Warming Potential in Three Agroecosystems. *Nutrient Cycling in Agroecosystems*, *72*(1), 67–76. https://doi.org/10.1007/s10705-004-7356-0

Mosier, A. R., Parton, W. J., Valentine, D. W., Ojima, D. S., Schimel, D. S., & Delgado, J. A. (1996). CH4 and N2O fluxes in the Colorado shortgrass steppe: 1. Impact of landscape and nitrogen addition. *Global Biogeochemical Cycles*, *10*(3), 387–399. https://doi.org/10.1029/96GB01454

Mosier, A. R., Parton, W. J., Valentine, D. W., Ojima, D. S., Schimel, D. S., & Heinemeyer, O. (1997). CH4 and N2O fluxes in the Colorado shortgrass steppe: 2. Long-term impact of land use change. *Global Biogeochemical Cycles*, *11*(1), 29–42.

Olson, K. R., Ebelhar, S. A., & Lang, J. M. (2010). Cover Crop Effects on Crop Yields and Soil Organic Carbon Content. *Soil Science*, *175*(2), 89–98. https://doi.org/10.1097/SS.0b013e3181cf7959

Osborne, S., & Lehman, R. (2018). *All; 2000-2015. Ver. GPSR_NATRES. Fort Collins, CO. USDA-ARS Natural Resources Database. Project: SDBRREAP*. https://gpsr.ars.usda.gov/SDBRREAP

Osterholz, W. R., Kucharik, C. J., Hedtcke, J. L., & Posner, J. L. (2014). Seasonal Nitrous Oxide and Methane Fluxes from Grain- and Forage-Based Production Systems in Wisconsin, USA. *Journal of Environmental Quality*, *43*(6), 1833–1843. https://doi.org/10.2134/jeq2014.02.0077

Parkin, T. B., & Hatfield, J. L. (2014). Enhanced Efficiency Fertilizers: Effect on Nitrous Oxide Emissions in Iowa. *Agronomy Journal*, *106*(2), 694–702. https://doi.org/10.2134/agronj2013.0219

Parkin, T. B., & Kaspar, T. C. (2006). Nitrous Oxide Emissions from Corn–Soybean Systems in the Midwest. *Journal of Environmental Quality*, *35*(4), 1496–1506. https://doi.org/10.2134/jeq2005.0183

Parkin, T. B., Kaspar, T. C., Jaynes, D. B., & Moorman, T. B. (2016). Rye cover crop effects on direct and indirect nitrous oxide emissions. *Soil Science Society of America Journal*, *80*(6), 1551–1559.

Parton, W. J., Holland, E. A., Del Grosso, S. J., Hartman, M. D., Martin, R. E., Mosier, A. R., Ojima, D. S., & Schimel, D. S. (2001). Generalized model for NO x and N2O emissions from soils. *Journal of Geophysical Research: Atmospheres*, *106*(D15), 17403–17419. https://doi.org/10.1029/2001JD900101

Paul, E. A., Collins, H. P., Paustian, K., Elliott, E. T., Frey, S., Juma, N., Janzen, H., Campbell, C. A., Zentner, R. P., Lafond, G. P., & Moulin, A. P. (2004). Management Effects on the Dynamics and Storage Rates of Organic Matter in Long-Term Crop Rotations. *Canadian Journal of Soil Science*, *84*(1), 49–61. https://doi.org/10.4141/S03-022

Phillips, R. L. (2007). Organic Agriculture and Nitrous Oxide Emissions at Sub-Zero Soil Temperatures. *Journal of Environmental Quality*, *36*(1), 23–30. https://doi.org/10.2134/jeq2006.0205

Phillips, R. L., Tanaka, D. L., Archer, D. W., & Hanson, J. D. (2009). Fertilizer Application Timing Influences Greenhouse Gas Fluxes Over a Growing Season. *Journal of Environmental Quality*, *38*(4), 1569–1579. https://doi.org/10.2134/jeq2008.0483

Pimentel, D., Hepperly, P., Hanson, J., Douds, D., & Seidel, R. (2005). Environmental, Energetic, and Economic Comparisons of Organic and Conventional Farming Systems. *BioScience*, *55*(7), 573. https://doi.org/10.1641/0006-3568(2005)055[0573:EEAECO]2.0.CO;2

Pinheiro, J. C., & Bates, D. M. (2000). Linear mixed-effects models: Basic concepts and examples. *Mixed-Effects Models in S and S-Plus*, 3–56.

Pinheiro, J. C., Bates, D. M., DebRoy, S., Sarkar, D., & R Core Team. (2022). *nlme: Linear and Nonlinear Mixed Effects Models*. https://CRAN.R-project.org/package=nlme

Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, *6*(1), 7–11.

Potter, S. R., Andrews, S., Atwood, J. D., Kellogg, R. L., Lemunyon, J., Norfleet, L., & Oman, D. (2006). *Model Simulation of Soil Loss, Nutrient Loss, and Change in Soil Organic Carbon Associated with Crop Production* (Conservation Effects Assessment Project (CEAP)). USDA Natural Resources Conservation Service. https://www.nrcs.usda.gov/publications/ceap-crop-2006-model-simulation-soil-nutrient-full.pdf

Powers, W., Auvermann, B., Cole, N. A., Gooch, C., Grant, R., Hatfield, J., Hunt, P., Johnson, K., Leytem, A., Liao, W., & Powell, J. M. (2014). Quantifying greenhouse gas sources and sinks in animal production systems. In M. Eve, D. Pape, M. Flugge, R. Steele, D. Man, M. Riley-Gilbert, & S. Biggar (Eds.), *Quantifying Greenhouse Gas Fluxes in Agriculture and Forestry: Methods for Entity-Scale Inventory*. U.S. Department of Agriculture.

Powlson, D. S., Bhogal, A., Chambers, B. J., Coleman, K., Macdonald, A. J., Goulding, K. W. T., & Whitmore, A. P. (2012). The potential to increase soil carbon stocks through reduced tillage or organic material additions in England and Wales: A case study. *Agriculture, Ecosystems & Environment*, *146*(1), 23–33.

Python Language Reference, version 3.9. (n.d.). Python Software Foundation. http://www.python.org

R Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rasmussen, P. E., & Smiley, R. W. (1997). Soil carbon and nitrogen change in long-term agricultural experiments at Pendleton, Oregon. In *Soil Organic Matter in Temperate Agroecosystems: Long-Term Experiments in North America* (pp. 353–360). CRC Press.

Razavi, S., & Gupta, H.V. (2015). What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity. *Earth and Environmental Systems Models, Water Resour. Res.*, *51*, 3070– 3092. https://doi.org/10.1002/2014WR016527

Research, R. (2014). *Dataset: Broadbalk soil organic carbon content 1843-2010*. Electronic Rothamsted Archive, Rothamsted Research. http://www.era.rothamsted.ac.uk/dataset/rbk1/01-BKSOC1843/

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Rochester, I. J. (2011). Sequestering carbon in minimum-tilled clay soils used for irrigated cotton and grain production. *Soil and Tillage Research*, *112*(1), 1–7.

Rodell, M., Houser, P. R., Jambor, U. E. A., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., & Bosilovich, M. (2004). The global land data assimilation system. *Bulletin of the American Meteorological Society*, *85*(3), 381–394.

Rosner, B. A. (2006). Fundamentals of Biostatistics. In *Thomson Brooks/Cole* (6th ed.).

Roy, A. K., Wagner-Riddle, C., Deen, B., Lauzon, J., & Bruulsema, T. (2014). Nitrogen application rate, timing and history effects on nitrous oxide emissions from corn (Zea mays L.). *Canadian Journal of Soil Science*, *94*(4), 563–573. https://doi.org/10.4141/cjss2013-118

Ruan, L., & Robertson, G. P. (2017). Reduced Snow Cover Increases Wintertime Nitrous Oxide (N_2O) Emissions from an Agricultural Soil in the Upper U.S. Midwest. *Ecosystems*, *20*(5), 917–927. https://doi.org/10.1007/s10021-016-0077-9

Sainju, U. M., Caesar-TonThat, T., Lenssen, A. W., & Barsotti, J. L. (2012). Dryland Soil Greenhouse Gas Emissions Affected by Cropping Sequence and Nitrogen Fertilization. *Soil Science Society of America Journal*, *76*(5), 1741–1757. https://doi.org/10.2136/sssaj2012.0076

Sainju, U. M., Stevens, W. B., Caesar-TonThat, T., & Liebig, M. A. (2012). Soil Greenhouse Gas Emissions Affected by Irrigation, Tillage, Crop Rotation, and Nitrogen Fertilization. *Journal of Environmental Quality*, *41*(6), 1774–1786. https://doi.org/10.2134/jeq2012.0176

Sainju, U. M., Whitehead, W. F., & Singh, B. P. (2005). Carbon accumulation in cotton, sorghum, and underlying soil as influenced by tillage, cover crops, and nitrogen fertilization. *Plant and Soil*, *273*(1), 219–234.

Sansoulet, J., Pattey, E., Kröbel, R., Grant, B., Smith, W., Jégo, G., Desjardins, R. L., Tremblay, N., & Tremblay, G. (2014). Comparing the performance of the STICS, DNDC, and DayCent models for predicting N uptake and biomass of spring wheat in Eastern Canada. *Field Crops Research*, *156*, 135–150.

Savage, K., Phillips, R., & Davidson, E. (2014). High temporal frequency measurements of greenhouse gas emissions from soils. *Biogeosciences*, *11*(10), 2709–2720. https://doi.org/10.5194/bg-11-2709-2014

Saxton, K. E., & Rawls, W. J. (2006). Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Science Society of America Journal*, *70*(5), 1569–1578. https://doi.org/10.2136/sssaj2005.0117

Scheer, C., Del Grosso, S. J., Parton, W. J., Rowlings, D. W., & Grace, P. R. (2014). Modeling nitrous oxide emissions from irrigated agriculture: Testing DayCent with high-frequency measurements. *Ecological Applications*, *24*(3), 528–538. https://doi.org/10.1890/13-0570.1

Schmer, M. R., Jin, V. L., Wienhold, B. J., Varvel, G. E., & Follett, R. F. (2014). Tillage and Residue Management Effects on Soil Carbon and Nitrogen Under Irrigated Continuous Corn. *Soil Science Society of America Journal*, *78*(6), 1987–1996. https://doi.org/10.2136/sssaj2014.04.0166

Schulten, H.-R., & Schnitzer, M. (1998). The chemistry of soil organic nitrogen: A review. *Biology and Fertility of Soils*, *26*(1), 1–15. https://doi.org/10.1007/s003740050335

Schwager, E. A., VanderZaag, A. C., Wagner-Riddle, C., Crolla, A., Kinsley, C., & Gregorich, E. (2016). Field Nitrogen Losses Induced by Application Timing of Digestate from Dairy Manure Biogas Production. *Journal of Environmental Quality*, *45*(6), 1829–1837. https://doi.org/10.2134/jeq2016.04.0148

Senapati, N., Hulugalle, N. R., Smith, P., Wilson, B. R., Yeluripati, J. B., Daniel, H., Ghosh, S., & Lockwood, P. (2014). Modelling soil organic carbon storage with RothC in irrigated Vertisols under cotton cropping systems in the sub-tropics. *Soil and Tillage Research*, *143*, 38–49.

Shelton, R. E., Jacobsen, K. L., & McCulley, R. L. (2018). Cover Crops and Fertilization Alter Nitrogen Loss in Organic and Conventional Conservation Agriculture Systems. *Frontiers in Plant Science*, *8*. https://www.frontiersin.org/articles/10.3389/fpls.2017.02260

Sistani, K. R., Jn-Baptiste, M., Lovanh, N., & Cook, K. L. (2011). Atmospheric Emissions of Nitrous Oxide, Methane, and Carbon Dioxide from Different Nitrogen Fertilizers. *Journal of Environmental Quality*, *40*(6), 1797–1805. https://doi.org/10.2134/jeq2011.0197

Skinner, R. H., Corson, M. S., & Rotz, C. A. (2008). Comparison of two pasture growth models of differing complexity. *Agricultural Systems*, *99*(1), 35–43.

Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *MMCE*, *4*, 407–414.

Soil Survey Staff. (2022). *Gridded Soil Survey Geographic (gSSURGO) Database for the United States of America and the Territories, Commonwealths, and Island Nations served by the USDA-NRCS Service*. United States Department of Agriculture, Natural Resources Conservation. https://gdg.sc.egov.usda.gov/

Station, K. B. (2021). *Main Cropping System Experiment (MCSE) downloadable tables 164 and 308*. Kellogg Biological Station, Long-Term Ecological Research. https://lter.kbs.msu.edu/datatables/

Stewart, C. E., Follett, R. F., Pruessner, E. G., Varvel, G. E., Vogel, K. P., & Mitchell, R. B. (2015). Nitrogen and Harvest Effects on Soil Properties under Rainfed Switchgrass and No-till Corn over 9 Years: Implications for Soil Quality. *GCB Bioenergy*, *7*(2), 288–301. https://doi.org/10.1111/gcbb.12142

Tarre, R., Macedo, R., Cantarutti, R. B., Rezende, C. de P., Pereira, J. M., Ferreira, E., Urquiaga, S., & Boddey, R. M. (n.d.). *The effect of the presence of a forage legume on nitrogen and carbon levels in soils under Brachiaria pastures in the Atlantic forest region of the South of Bahia, Brazil*.

Teepe, R., Brumme, R., & Beese, F. (2000). Nitrous oxide emissions from frozen soils under agricultural, fallow and forest land. *Soil Biology and Biochemistry*, *32*(11), 1807–1810. https://doi.org/10.1016/S0038-0717(00)00078-X

Tenuta, M., Gao, X., Flaten, D. N., & Amiro, B. D. (2016). Lower Nitrous Oxide Emissions from Anhydrous Ammonia Application Prior to Soil Freezing in Late Fall Than Spring Pre-Plant Application. *Journal of Environmental Quality*, *45*(4), 1133–1143. https://doi.org/10.2134/jeq2015.03.0159

Thapa, R., Chatterjee, A., Awale, R., McGranahan, D. A., & Daigh, A. (2016). Effect of Enhanced Efficiency Fertilizers on Nitrous Oxide Emissions and Crop Yields: A Meta-analysis. *Soil Science Society of America Journal*, *80*(5), 1121–1134. https://doi.org/10.2136/sssaj2016.06.0179

Thapa, R., Chatterjee, A., Johnson, J. M. F., & Awale, R. (2015). Stabilized Nitrogen Fertilizers and Application Rate Influence Nitrogen Losses under Rainfed Spring Wheat. *Agronomy Journal*, *107*(5), 1885–1894. https://doi.org/10.2134/agronj15.0081

Thornton, F. C., Shurpali, N. J., Bock, B. R., & Reddy, K. C. (1998). N2O and no emissions from poultry litter and urea applications to Bermuda grass. *Atmospheric Environment*, *32*(9), 1623–1630. https://doi.org/10.1016/S1352-2310(97)00390-7

Thornton, F. C., & Valente, R. J. (1996). Soil Emissions of Nitric Oxide and Nitrous Oxide from No-till Corn. *Soil Science Society of America Journal*, *60*(4), 1127–1133. https://doi.org/10.2136/sssaj1996.03615995006000040024x

Tian, Z., Wang, J. J., Liu, S., Zhang, Z., Dodla, S. K., & Myers, G. (2015). Application effects of coated urea and urease and nitrification inhibitors on ammonia and greenhouse gas emissions from a subtropical cotton field of the Mississippi delta region. *Science of The Total Environment*, *533*, 329–338. https://doi.org/10.1016/j.scitotenv.2015.06.147

U.S. EPA. (2020). *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2018*. https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2018

USDA Agricultural Research Service. (2015). *GRACEnet (Greenhouse gas Reduction through Agricultural Carbon Enhancement network)* [dataset]. https://doi.org/10.15482/USDA.ADC/1235557

Ussiri, D. A. N., Lal, R., & Jarecki, M. K. (2009). Nitrous oxide and methane emissions from long-term tillage under a continuous corn cropping system in Ohio. *Soil and Tillage Research*, *104*(2), 247–255. https://doi.org/10.1016/j.still.2009.03.001

Varvel, G. E. (2006). Soil Organic Carbon Changes in Diversified Rotations of the Western Corn Belt. *Soil Science Society of America Journal*, *70*(2), 426–433. https://doi.org/10.2136/sssaj2005.0100

Veenstra, J. J., Horwath, W. R., Mitchell, J. P., & Munk, D. S. (2006). Conservation Tillage and Cover Cropping Influence Soil Properties in San Joaquin Valley Cotton-Tomato Crop. *California Agriculture*, *60*(3), 146–153. https://doi.org/10.3733/ca.v060n03p146

Venterea, R. T., Dolan, M. S., & Ochsner, T. E. (2010). Urea Decreases Nitrous Oxide Emissions Compared with Anhydrous Ammonia in a Minnesota Corn Cropping System. *Soil Science Society of America Journal*, *74*(2), 407–418. https://doi.org/10.2136/sssaj2009.0078

Vitosh, M. L., Davis, J. F., & Knezek, B. D. (1973). Long-Term Effects of Manure, Fertilizer, and Plow Depth on Chemical Properties of Soils and Nutrient Movement in a Monoculture Corn System. *Journal of Environmental Quality*, *2*(2), 296–299. https://doi.org/10.2134/jeq1973.00472425000200020029x

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273–316.

Vrugt, J. A., & Ter Braak, C. J. F. (2011). DREAM: an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrology and Earth System Sciences*, *15*(12), 3701–3713.

Vrugt, J. A., Ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, *10*(3), 273–290.

Wagner-Riddle, C., Furon, A., Mclaughlin, N. L., Lee, I., Barbeau, J., Jayasundara, S., Parkin, G., Von Bertoldi, P., & Warland, J. (2007). Intensive measurement of nitrous oxide emissions from a corn–soybean–wheat rotation under two contrasting management systems over 5 years. *Global Change Biology*, *13*(8), 1722–1736. https://doi.org/10.1111/j.1365-2486.2007.01388.x

Wegner, B. R., Chalise, K. S., Singh, S., Lai, L., Abagandura, G. O., Kumar, S., Osborne, S. L., Lehman, R. M., & Jagadamma, S. (2018). Response of soil surface greenhouse gas fluxes to crop residue removal and cover crops under a corn-soybean rotation. *Journal of Environmental Quality*, *47*(5), 1146–1154.

White, M., Harmel, D., Yen, H., Arnold, J., Gambone, M., & Haney, R. (2015). Development of Sediment and Nutrient Export Coefficients for U.S. Ecoregions. *JAWRA Journal of the American Water Resources Association*, *51*(3), 758–775. https://doi.org/10.1111/jawr.12270

Wilson, T. M., McGowen, B., Mullock, J., Arnall, D. B., & Warren, J. G. (2015). Nitrous Oxide Emissions from Continuous Winter Wheat in the Southern Great Plains. *Agronomy Journal*, *107*(5), 1878–1884. https://doi.org/10.2134/agronj15.0096

Yanai, Y., Hirota, T., Iwata, Y., Nemoto, M., Nagata, O., & Koga, N. (2011). Accumulation of nitrous oxide and depletion of oxygen in seasonally frozen soils in northern Japan – Snow cover manipulation experiments. *Soil Biology and Biochemistry*, *43*(9), 1779–1786. https://doi.org/10.1016/j.soilbio.2010.06.009

Yu, T., Lin, F., Liu, X., & Wang, X. (2020). Recovery Role in Soil Structural, Carbon and Nitrogen Properties of the Conversion of Vegetable Land to Alfalfa Land in Northwest China. *Journal of Soil Science and Plant Nutrition*, *20*(3), 1366–1377. https://doi.org/10.1007/s42729-020-00218-w

Zebarth, B. J., Snowdon, E., Burton, D. L., Goyer, C., & Dowbenko, R. (2012). Controlled release fertilizer product effects on potato crop response and nitrous oxide emissions under rain-fed production on a medium-textured soil. *Canadian Journal of Soil Science*, *92*(5), 759–769. https://doi.org/10.4141/cjss2012-008

Zhang, Y. A., Suyker, A., & Paustian, K. (2018). Improved crop canopy and water balance dynamics for agroecosystem modeling using DayCent. *Agronomy Journal*, *110*(2), 511–524.

Zhang, Y., Arabi, M., & Paustian, K. (2020). Analysis of parameter uncertainty in model simulations of irrigated and rainfed agroecosystems. *Environmental Modelling & Software*, *126*(January), 104642–104642. https://doi.org/10.1016/j.envsoft.2020.104642

Zhang, Y., Suyker, A., & Paustian, K. (2018). Improved Crop Canopy and Water Balance Dynamics for Agroecosystem Modeling Using DayCent. *Agronomy Journal*, *110*(2), 511–524. https://doi.org/10.2134/agronj2017.06.0328

B Declaration of Practices

Table B1: Practice effects considered additional in this project and the associated practice categories

Practice Effect	**Practice Category**
New cover crop adoption	CROP
Longer duration of cover crops through delayed termination	CROP
Longer duration of cover crops through earlier planting	CROP
Adding a legume species to existing cover crop mix	CROP
Adding new crops to rotation	CROP
Tillage reduction through number of passes	DISTURB
Tillage reduction through delayed tilling	DISTURB
Tillage equipment/intensity change	DISTURB
N fertilizer reduction	NFERT
Change in N fertilizer application method	NFERT
Change in N fertilizer form	NFERT
Change in N fertilizer product with stabilizers or inhibitors	NFERT
Change in N fertilizer application timing	NFERT
Substitute synthetic N with organic amendments	NFERT and ORG

Appendix C: Sampler Diagnostics

Here we present figures and tables that illustrate the MCMC algorithim used to fit parameters during the five-fold cross-validation.

For each fold, the following figures and tables are shown:

- Plots of marginal densities for each model parameter, colored by chain.
- Trace plots for each parameter, colored by chain.
- Plots of marginal densities of variance parameters.
- Plot of Gelman-Rubin statistic, $\hat{R}$.
- Summary statistics of model parameter marginal posterior distributions.
- Summary statistics of variance parameter marginal posterior distributions.

Then we present variograms of between-site correlations of SOC pools N2O fluxes, and model residuals.

Contents

C.1 Fold 1

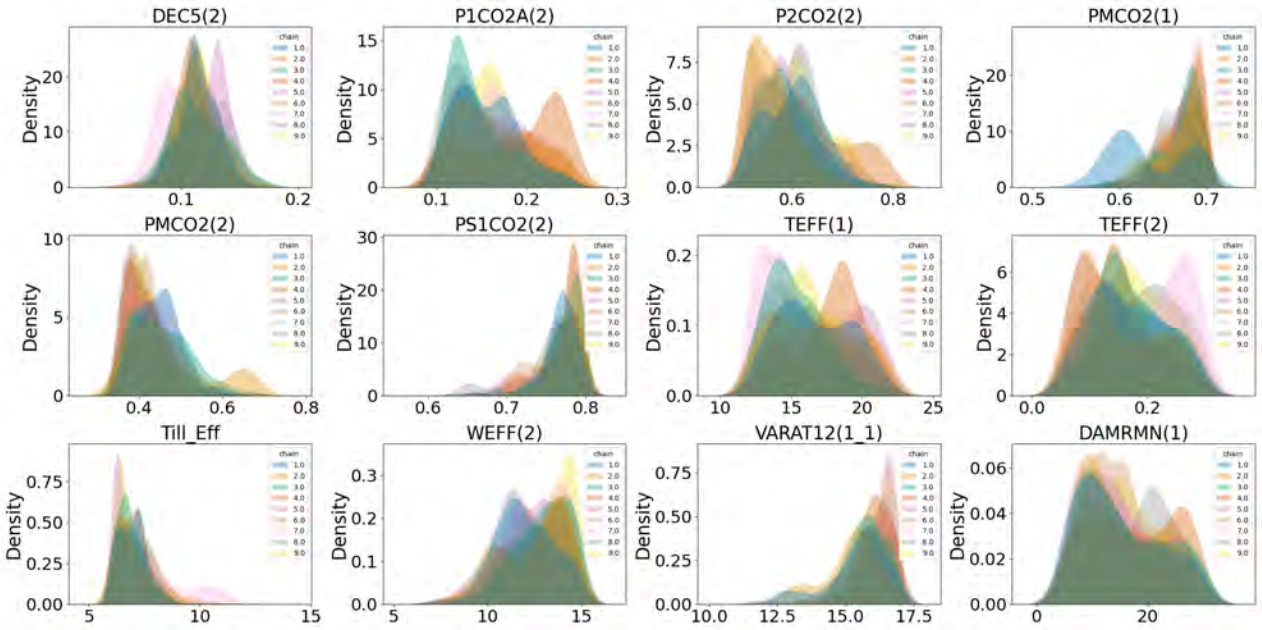

Figure C1: Marginal densities colored by Monte Carlo chain for fold 1.

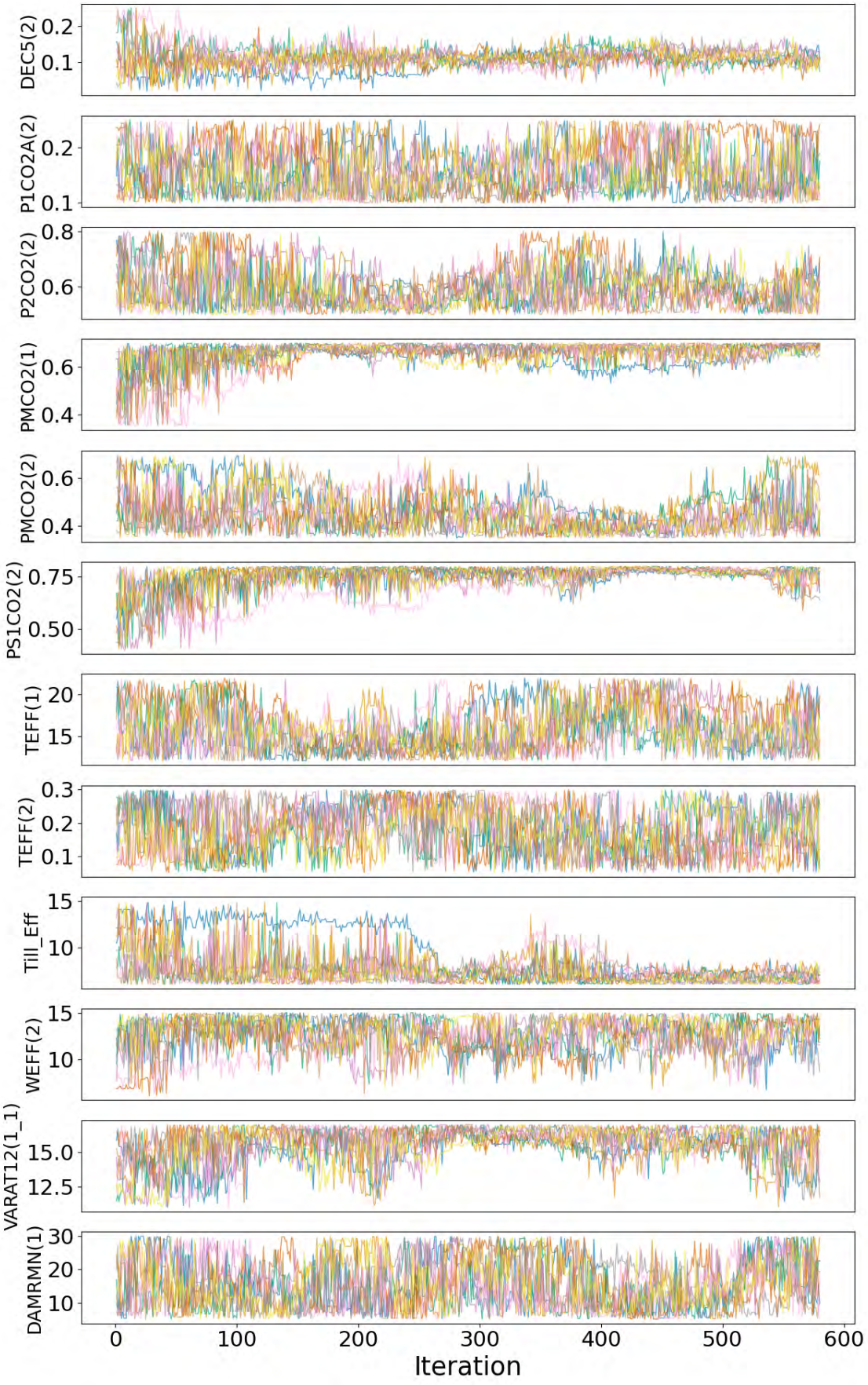

Figure C2: Traceplots for fold 1 (DayCent calibration parameters).

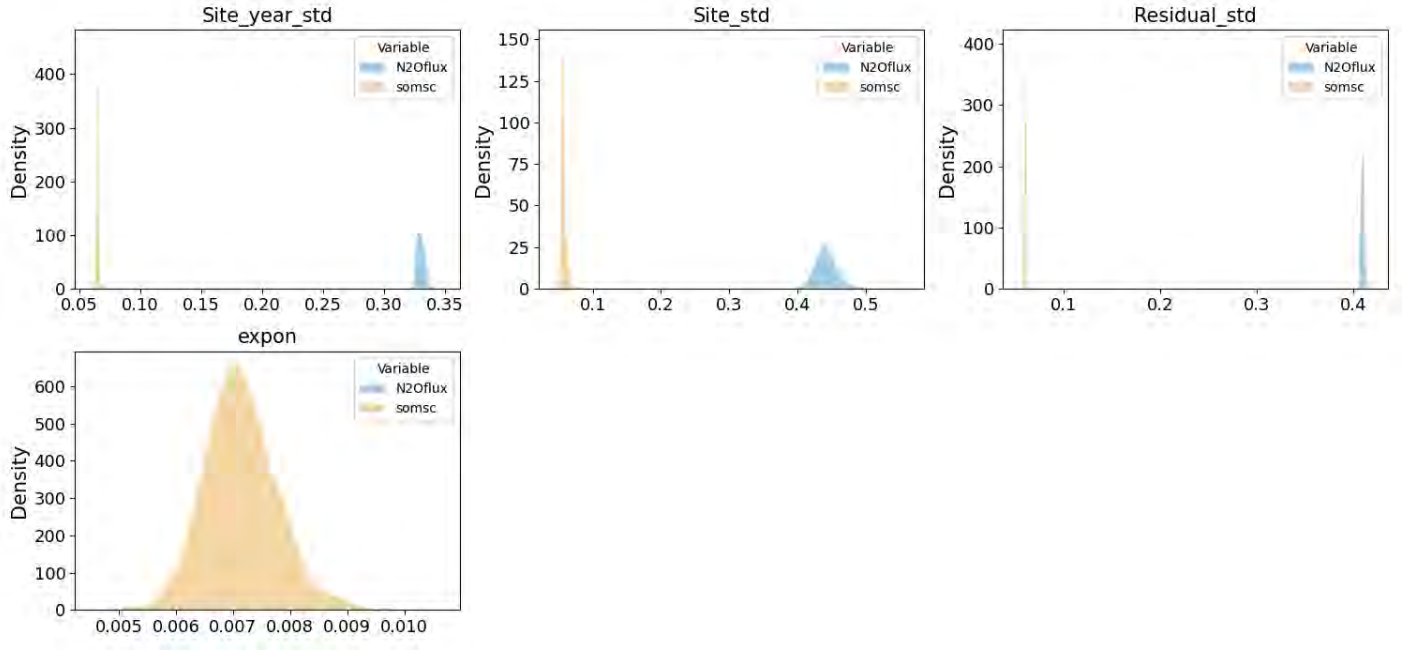

Figure C3: Marginal densities of variance parameters colored by Monte Carlo chain for fold 1.

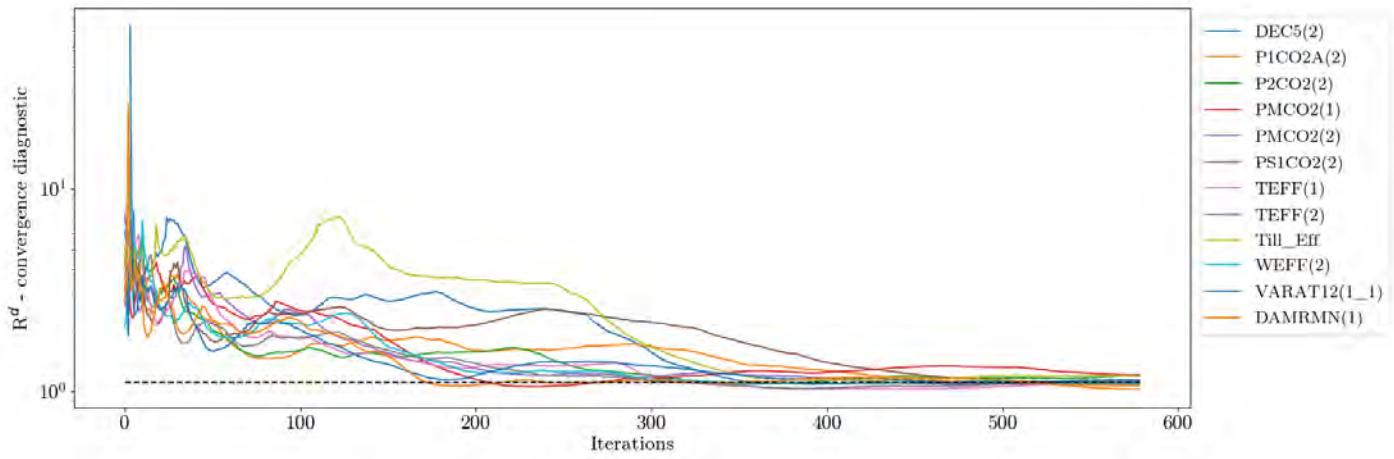

Figure C4: Gelman-Rubin (R) convergence criteria (i.e. shrinkage factors) for the DayCent calibration parameters and fold 1.

Table C1: Summary statistics of marginal posterior distributions of the calibrated DayCent-CR model parameters for fold 1 with 5th percentile, mean, median, 95th percentile and the standard deviation.

Parameter	5th percentile	95th percentile	Mean	median	Standard deviation
DAMRMN	6.4627	26.6329	15.522	14.5692	6.5035
DEC_5_2	0.0859	0.1434	0.115	0.1152	0.0176
P1CO2A_2	0.1141	0.2249	0.1606	0.1554	0.036
P2CO2_2	0.5165	0.6674	0.5867	0.5836	0.0477
PMCO2_1	0.625	0.6975	0.6675	0.6686	0.0243
PMCO2_2	0.3657	0.555	0.4442	0.4332	0.0617
PS1CO2_2	0.7235	0.7949	0.7687	0.7749	0.0253
TEFF1	13.0135	19.9753	16.2024	16.0369	2.1623
TEFF2	0.0741	0.2673	0.1685	0.1641	0.0621
Till_Eff	6.1044	7.6912	6.8217	6.7815	0.518
VARAT12(1,1)	13.6337	16.8401	15.5738	15.7918	1.0062
WEFF2	10.1177	14.562	12.625	12.7649	1.4261

Table C2: Random effects for the calibration dataset.

Parameter	Variable	5th percentile	95th percentile	Mean	Median	Standard deviation
$\sigma_{\text{site-year}}$	N2Oflux	0.325055	0.335709	0.329984	0.329779	0.003351
σ_{site}	N2Oflux	0.416062	0.471354	0.441674	0.440403	0.017058
σ	N2Oflux	0.406224	0.412065	0.408898	0.408818	0.001762
$\sigma_{\text{site-year}}$	SOC	0.062794	0.066108	0.064338	0.064208	0.001232
σ_{site}	SOC	0.053054	0.066261	0.057494	0.056064	0.005988
σ	SOC	0.057852	0.061259	0.059484	0.059441	0.001064
ν	SOC	0.006158	0.008207	0.007133	0.0071	0.000643

C.2 Fold 2

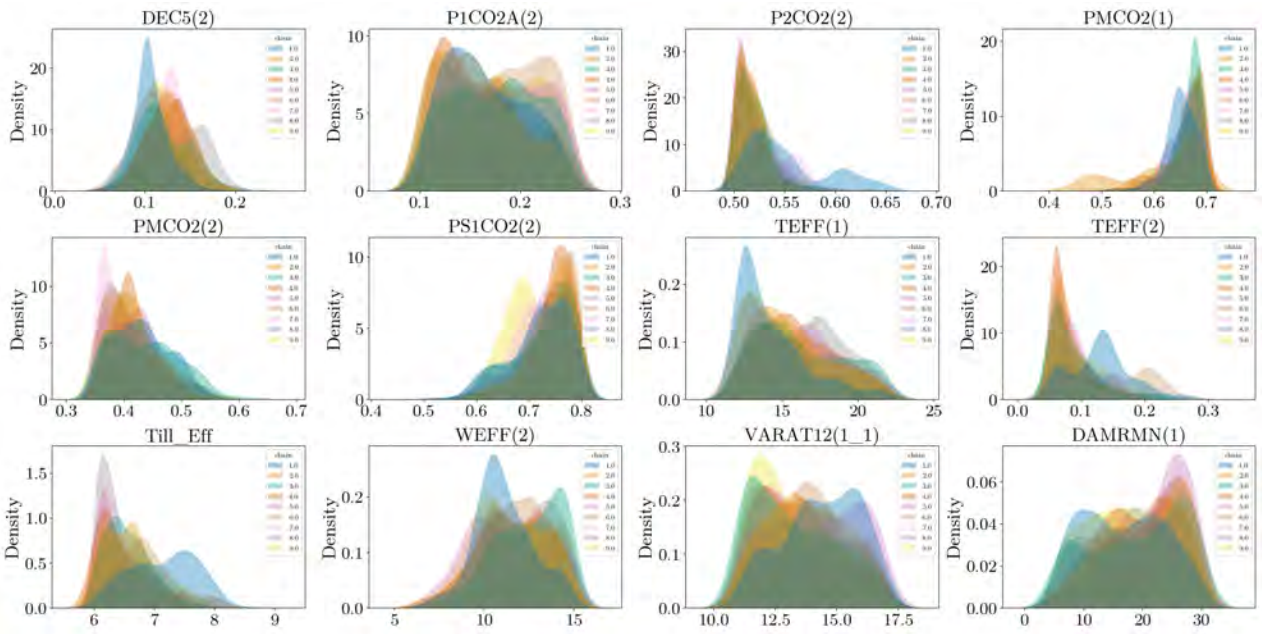

Figure C5: Marginal densities colored by Monte Carlo chain for fold 2.

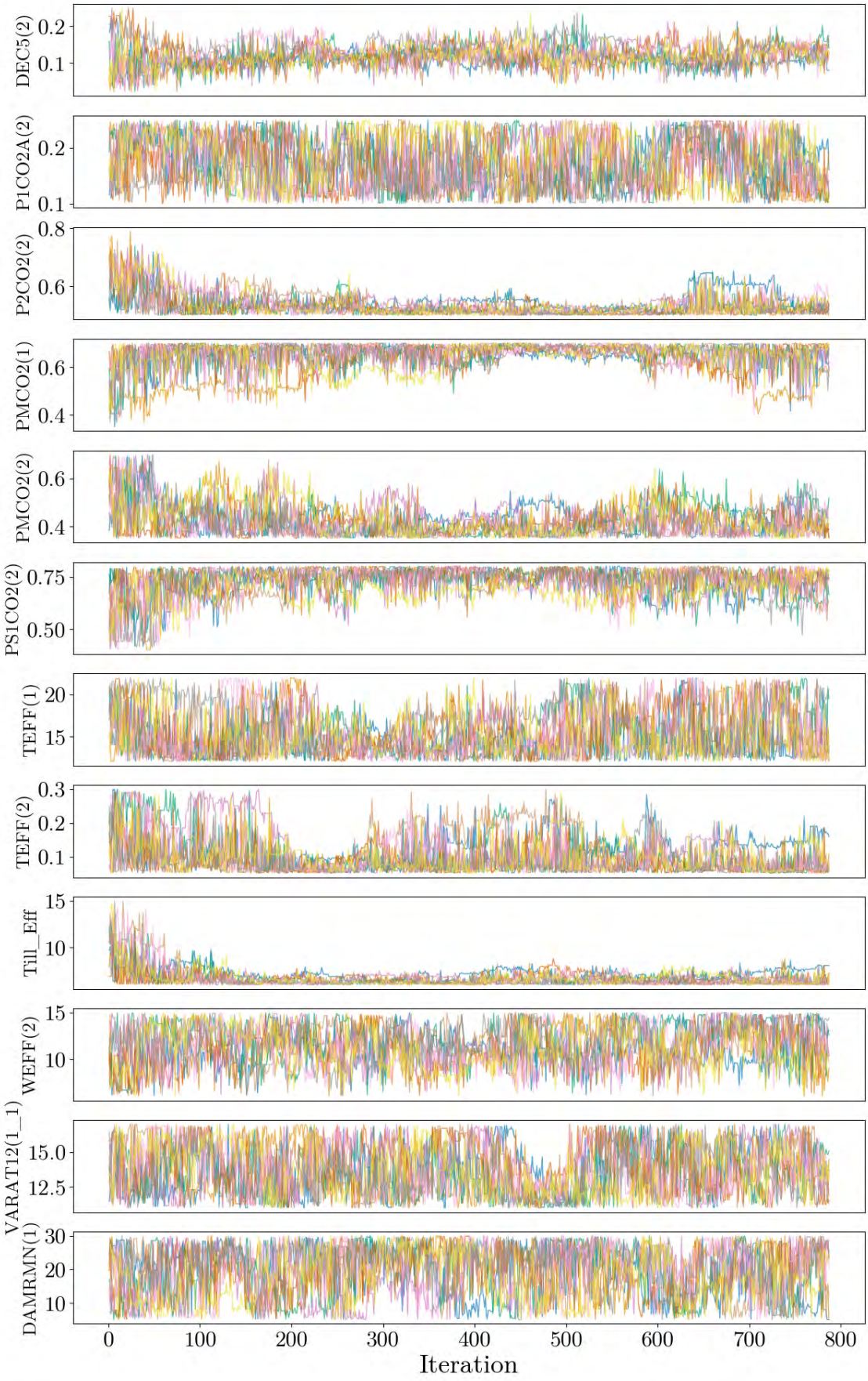

Figure C6: Traceplots for fold 2 (DayCent calibration parameters).

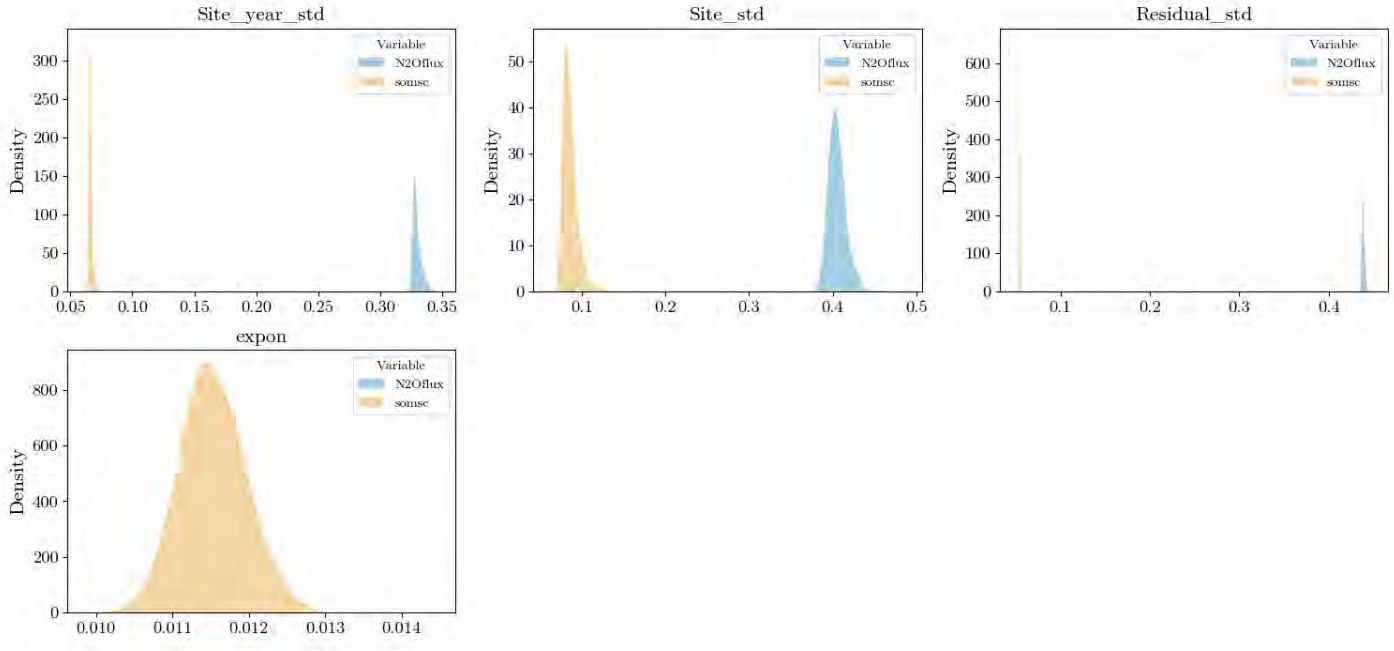

Figure C7: Marginal densities of variance parameters colored by Monte Carlo chain for fold 2.

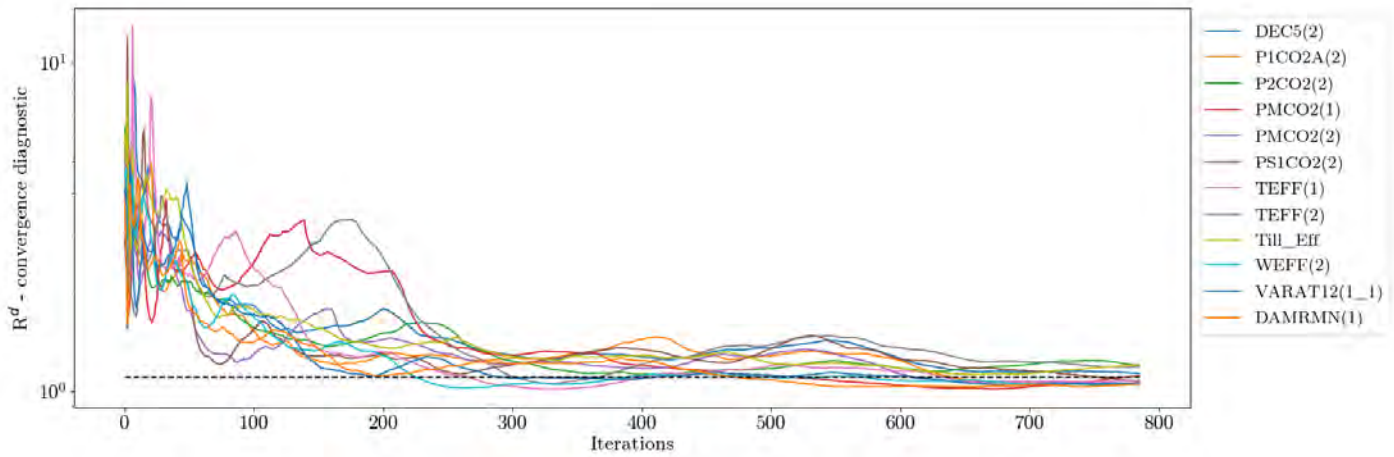

Figure C8: Gelman-Rubin (R) convergence criteria (i.e. shrinkage factors) for the DayCent calibration parameters and fold 2.

Table C3: Summary statistics of marginal posterior distributions of the calibrated DayCent-CR model parameters for fold 2 with 5th percentile, mean, median, 95th percentile and the standard deviation.

Parameter	5th percentile	95th percentile	Mean	Median	Standard deviation
DAMRMN	7.4344	28.3121	18.5014	18.6014	6.7373
DEC_5_2	0.0818	0.1577	0.1205	0.1216	0.024
P1CO2A_2	0.1117	0.2424	0.175	0.1733	0.0422
P2CO2_2	0.5026	0.5803	0.5331	0.5295	0.0265
PMCO2_1	0.5586	0.6886	0.6377	0.6476	0.0446
PMCO2_2	0.3571	0.5099	0.4225	0.4132	0.0485
PS1CO2_2	0.6279	0.7899	0.7201	0.7285	0.0543
TEFF1	12.3879	20.98	15.9303	15.3337	2.8008
TEFF2	0.0542	0.144	0.0879	0.0801	0.0304
Till_Eff	6.0702	7.3071	6.5715	6.5234	0.423
VARAT12(1,1)	11.4736	16.4448	13.9196	13.9727	1.5893
WEFF2	8.613	14.6181	11.8761	11.9663	1.8787

Table C4: Random effects for the calibration dataset.

Parameter	Variable	5th percentile	95th percentile	Mean	Median	Standard deviation
$\sigma_{\text{site-year}}$	N2Oflux	0.325129	0.336646	0.329356	0.328425	0.003574
$\sigma_{\text{site-}}$	N2Oflux	0.389779	0.424757	0.40511	0.403882	0.010825
σ	N2Oflux	0.435847	0.441353	0.438255	0.438052	0.00169
$\sigma_{\text{site-year}}$	SOC	0.06397	0.068515	0.06587	0.065659	0.001454
$\sigma_{\text{site-}}$	SOC	0.074559	0.102356	0.08557	0.083871	0.009063
σ	SOC	0.053042	0.055316	0.054012	0.053921	0.000711
ν	SOC	0.010833	0.012281	0.011528	0.011509	0.00044

C.3 Fold 3

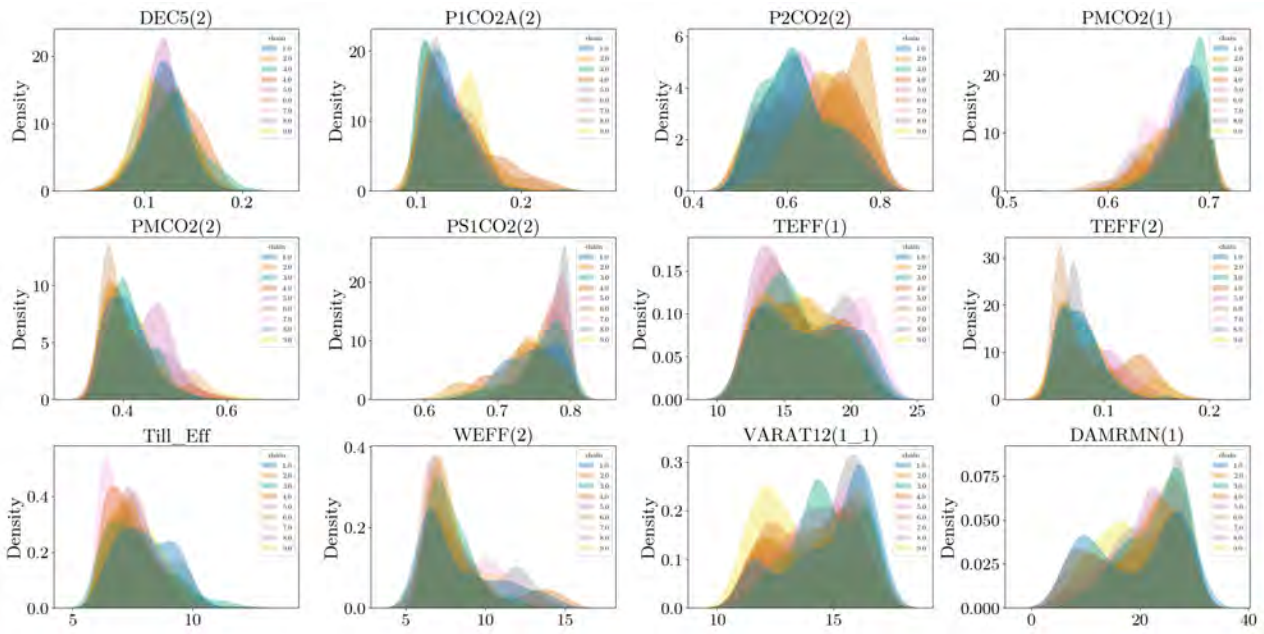

Figure C9: Marginal densities colored by Monte Carlo chain for fold 3.

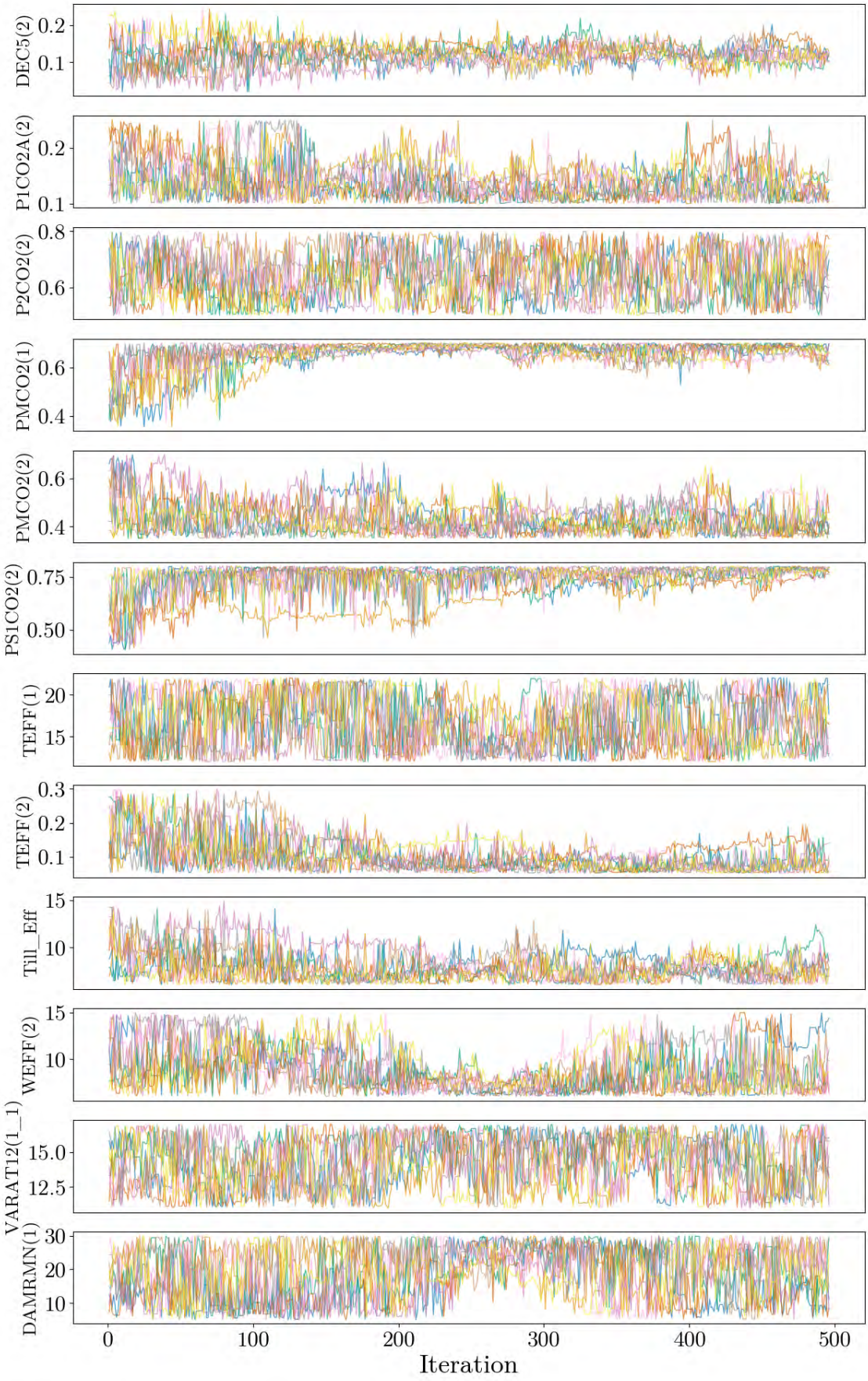

Figure C10: Traceplots for fold 3 (DayCent calibration parameters).

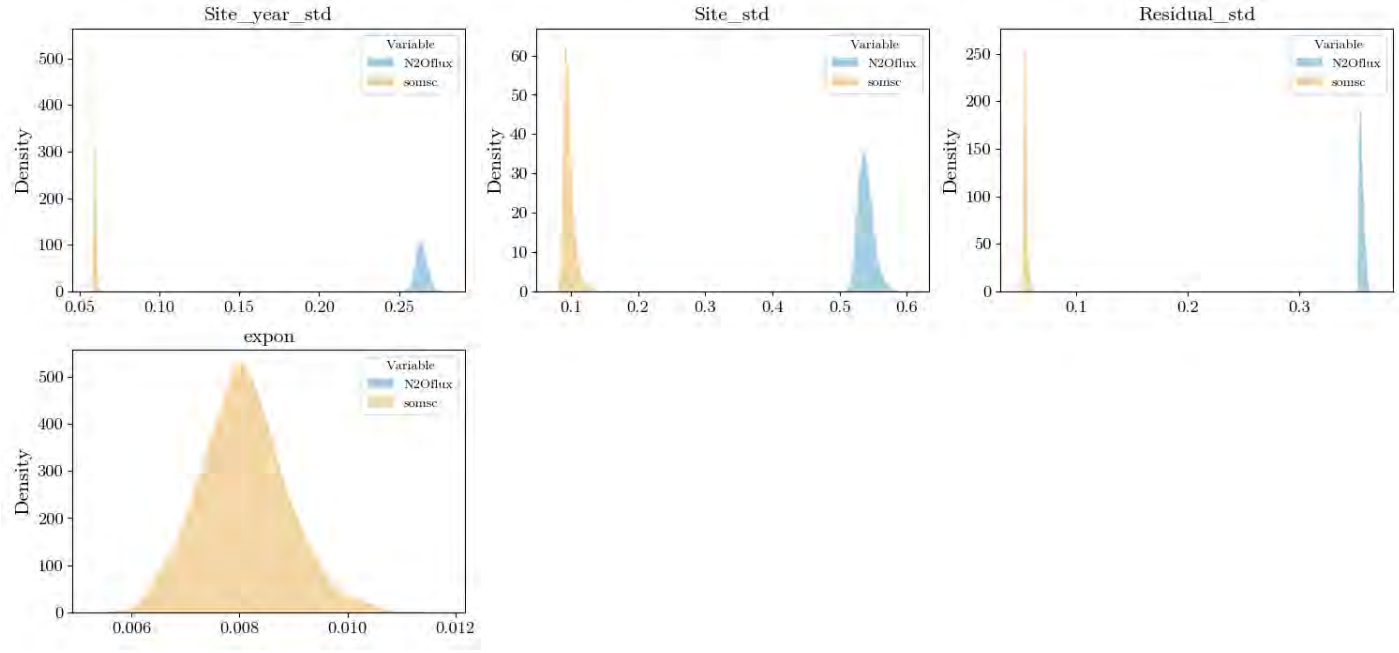

Figure C11: Marginal densities of variance parameters colored by Monte Carlo chain for fold 3.

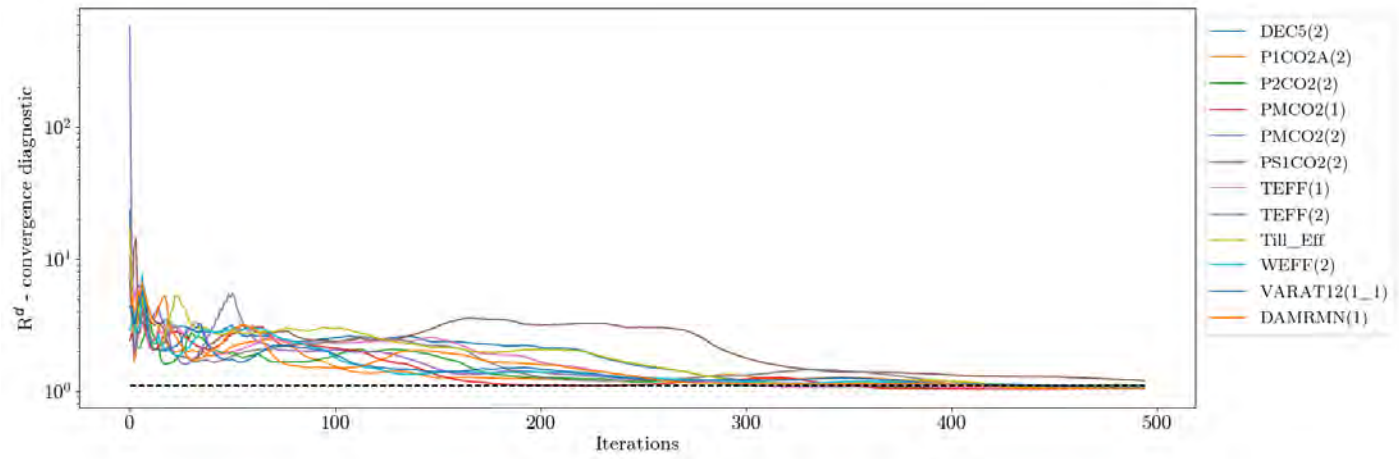

Figure C12: Gelman-Rubin (R) convergence criteria (i.e. shrinkage factors) for the DayCent calibration parameters and fold 3.

Table C5: Summary statistics of marginal posterior distributions of the calibrated DayCent-CR model parameters for fold 3 with 5th percentile, mean, median, 95th percentile and the standard deviation.

Parameter	5th percentile	95th percentile	Mean	Median	Standard deviation
DAMRMN	8.651	28.499	19.1439	19.5106	6.3542
DEC_5_2	0.0879	0.1525	0.1195	0.1193	0.0213
P1CO2A_2	0.1039	0.1798	0.1354	0.1317	0.0246
P2CO2_2	0.5331	0.7642	0.6494	0.6491	0.0739
PMCO2_1	0.6269	0.6974	0.6676	0.6718	0.0234
PMCO2_2	0.3582	0.5183	0.4272	0.4187	0.0536
PS1CO2_2	0.7227	0.7959	0.7663	0.7698	0.0241
TEFF1	12.6543	21.0392	16.6792	16.7136	2.6511
TEFF2	0.0546	0.1193	0.081	0.0764	0.0216
Till_Eff	6.3225	9.3565	7.6482	7.5313	0.9455
VARAT12(1,1)	11.5944	16.7136	14.1809	14.2527	1.6186
WEFF2	6.2722	11.9343	8.5229	8.1274	1.9181

Table C6: Random effects for the calibration dataset.

Parameter	Variable	5th percentile	95th percentile	Mean	Median	Standard deviation
$\sigma_{\text{site-year}}$	N2Oflux	0.258152	0.270373	0.263844	0.263643	0.003717
σ_{site}	N2Oflux	0.522189	0.560075	0.538748	0.53743	0.01174
σ	N2Oflux	0.35261	0.359333	0.35534	0.354947	0.002108
$\sigma_{\text{site-year}}$	SOC	0.058034	0.06097	0.059197	0.058953	0.000968
σ_{site}	SOC	0.087471	0.112421	0.096831	0.095078	0.008519
σ	SOC	0.051957	0.057201	0.054395	0.05432	0.001622
ν	SOC	0.006771	0.009473	0.008076	0.008033	0.000809

C.4 Fold 4

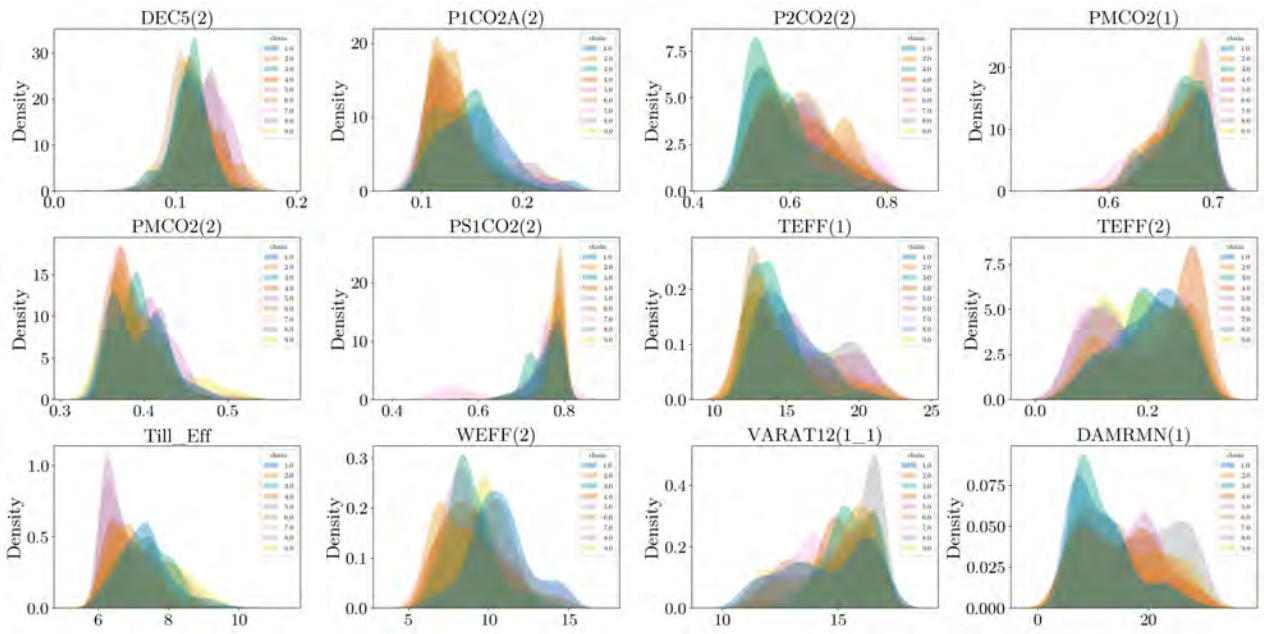

Figure C13: Marginal densities colored by Monte Carlo chain for fold 4.

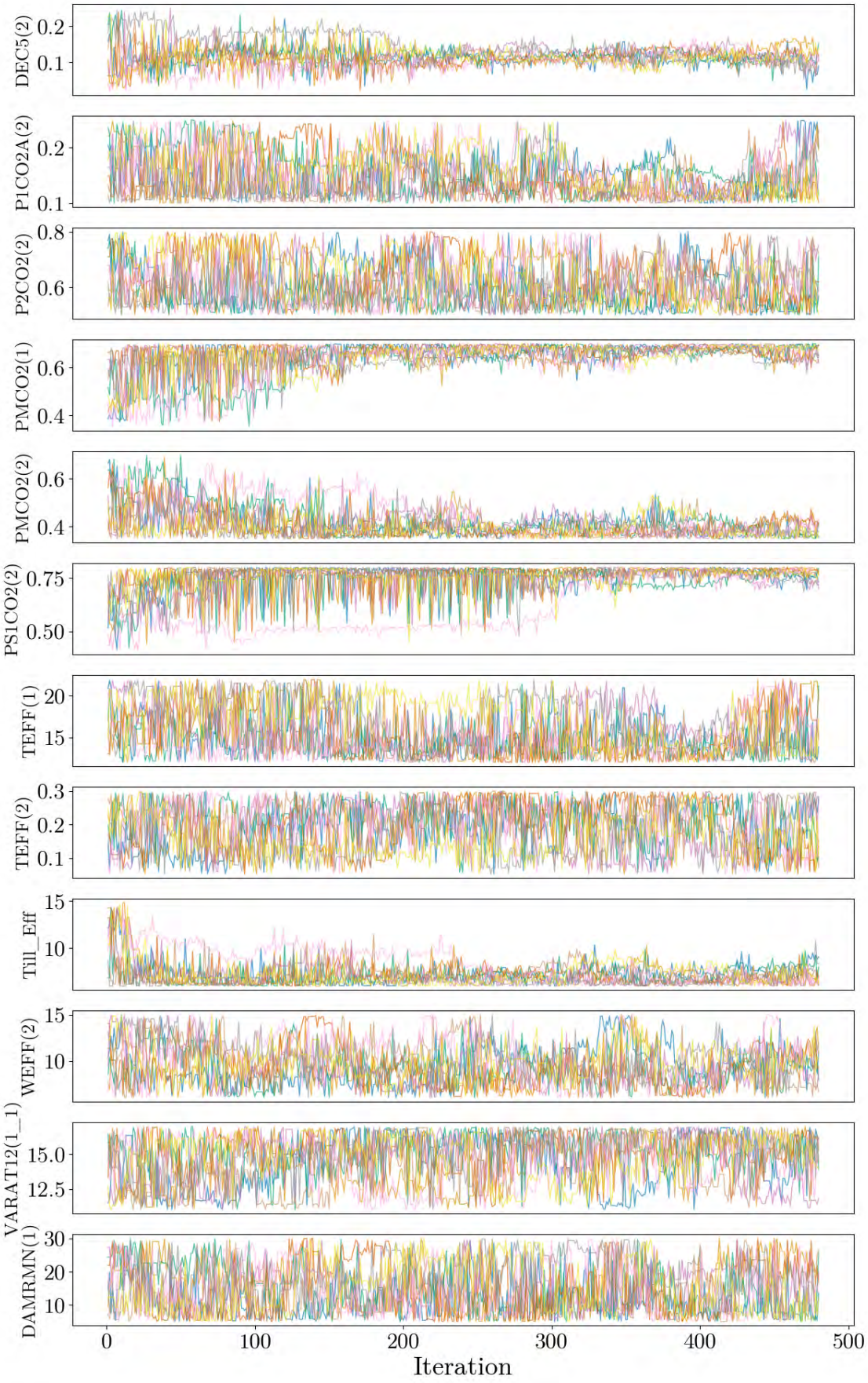

Figure C14: Traceplots for fold 4 (DayCent calibration parameters).

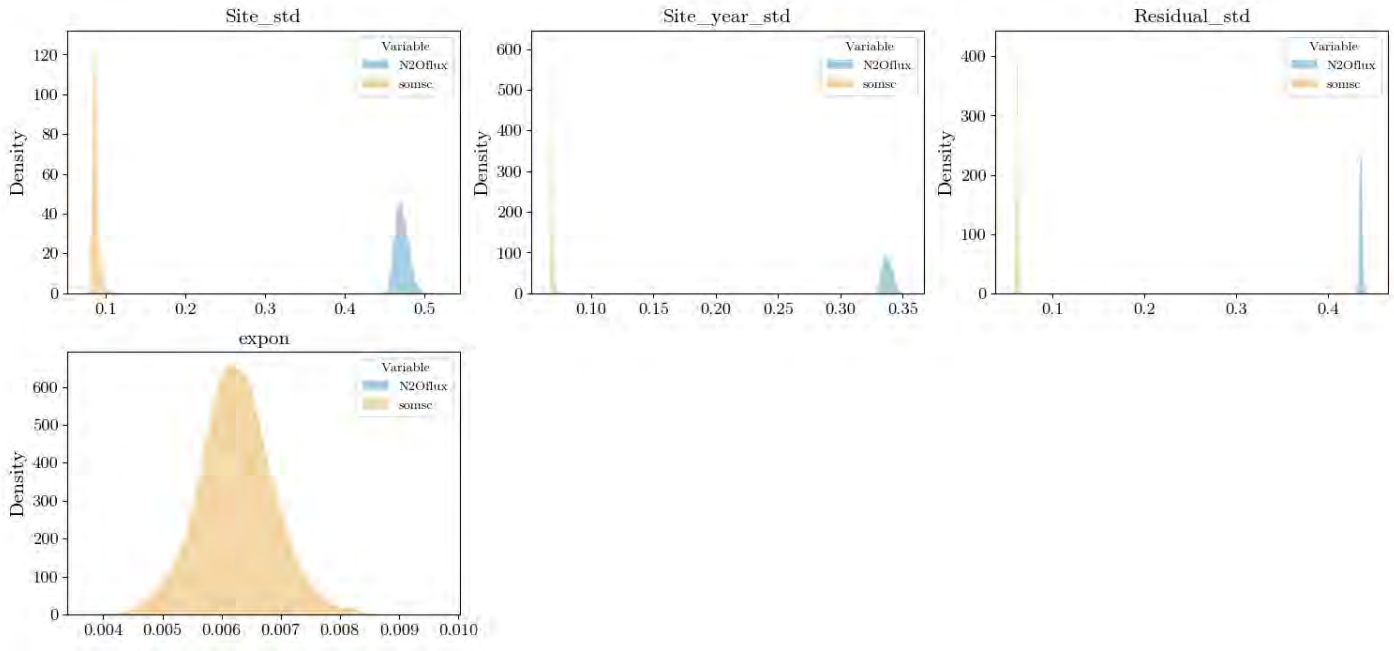

Figure C15: Marginal densities of variance parameters colored by Monte Carlo chain for fold 4.

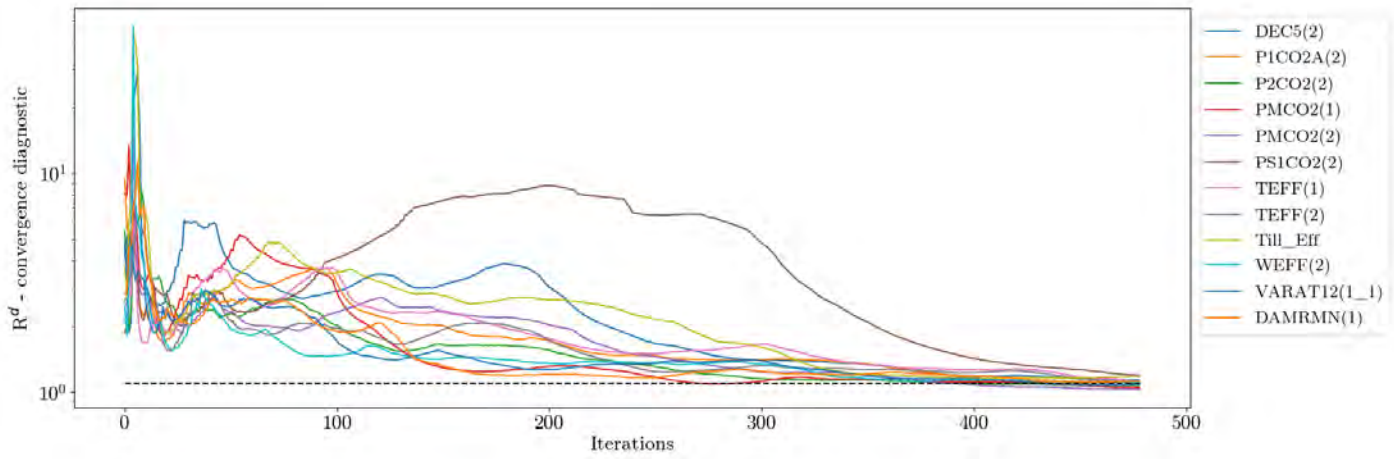

Figure C16: Gelman-Rubin (R) convergence criteria (i.e. shrinkage factors) for the DayCent calibration parameters and fold 4.

Table C7: Summary statistics of marginal posterior distributions of the calibrated DayCent-CR model parameters for fold 4 with 5th percentile, mean, median, 95th percentile and the standard deviation.

Parameter	5th percentile	95th percentile	Mean	Median	Standard deviation
DAMRMN	6.1831	24.9235	14.3041	14.0528	6.0002
DEC_5_2	0.0889	0.1422	0.1165	0.1163	0.0173
P1CO2A_2	0.103	0.1919	0.1365	0.128	0.0292
P2CO2_2	0.5171	0.7242	0.6051	0.591	0.0666
PMCO2_1	0.6366	0.6955	0.6715	0.676	0.0199
PMCO2_2	0.3527	0.4477	0.3921	0.3848	0.0323
PS1CO2_2	0.7316	0.7959	0.7699	0.773	0.0212
TEFF1	12.2527	19.2856	14.9725	14.4064	2.3161
TEFF2	0.0825	0.2848	0.1878	0.189	0.0673
Till_Eff	6.1585	8.1109	7.0263	6.953	0.6633
VARAT12(1,1)	12.7113	16.7688	15.0089	15.1893	1.3474
WEFF2	6.7579	12.0662	9.2586	9.1526	1.6918

Table C8: Random effects for the calibration dataset.

Parameter	Variable	5th percentile	95th percentile	Mean	Median	Standard deviation
σ_{site}	N2Oflux	0.459323	0.486587	0.47178	0.470911	0.00871
$\sigma_{\text{site-year}}$	N2Oflux	0.331578	0.343953	0.337443	0.337189	0.003919
σ	N2Oflux	0.432903	0.438151	0.435351	0.435274	0.001647
σ_{site}	SOC	0.081736	0.095246	0.086957	0.086028	0.005023
$\sigma_{\text{site-year}}$	SOC	0.067835	0.071062	0.068885	0.068519	0.001442
σ	SOC	0.059938	0.063396	0.061497	0.061418	0.001035
ν	SOC	0.005219	0.007287	0.006244	0.006228	0.000631

C.5 Fold 5

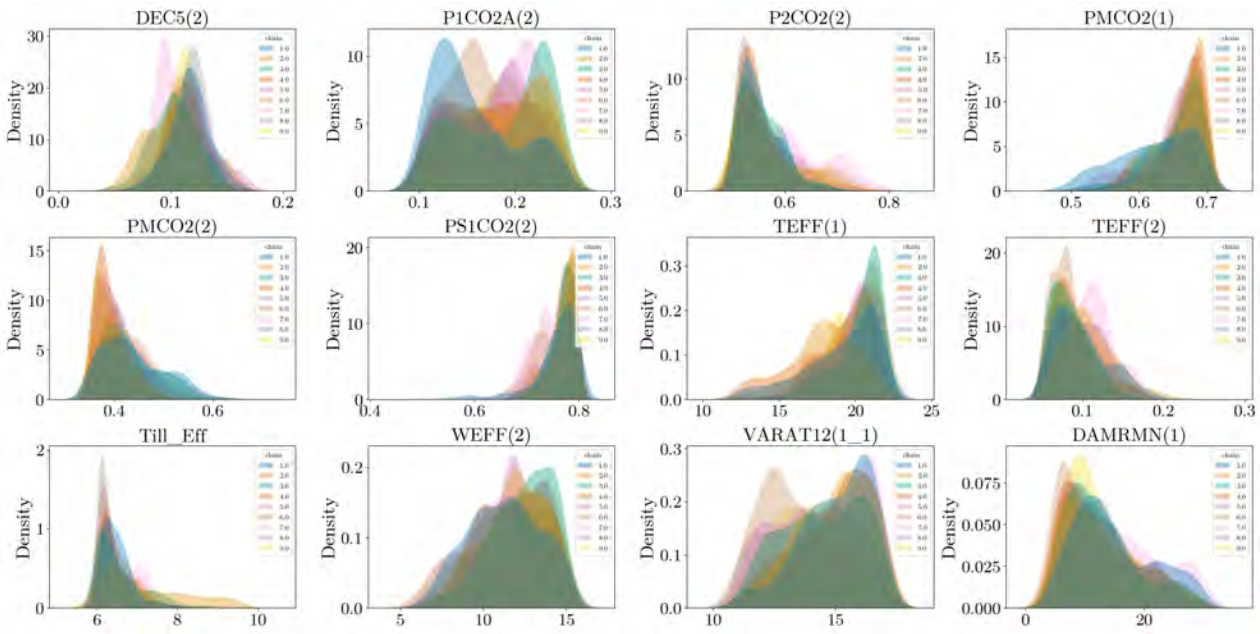

Figure C17: Marginal densities colored by Monte Carlo chain for fold 5.

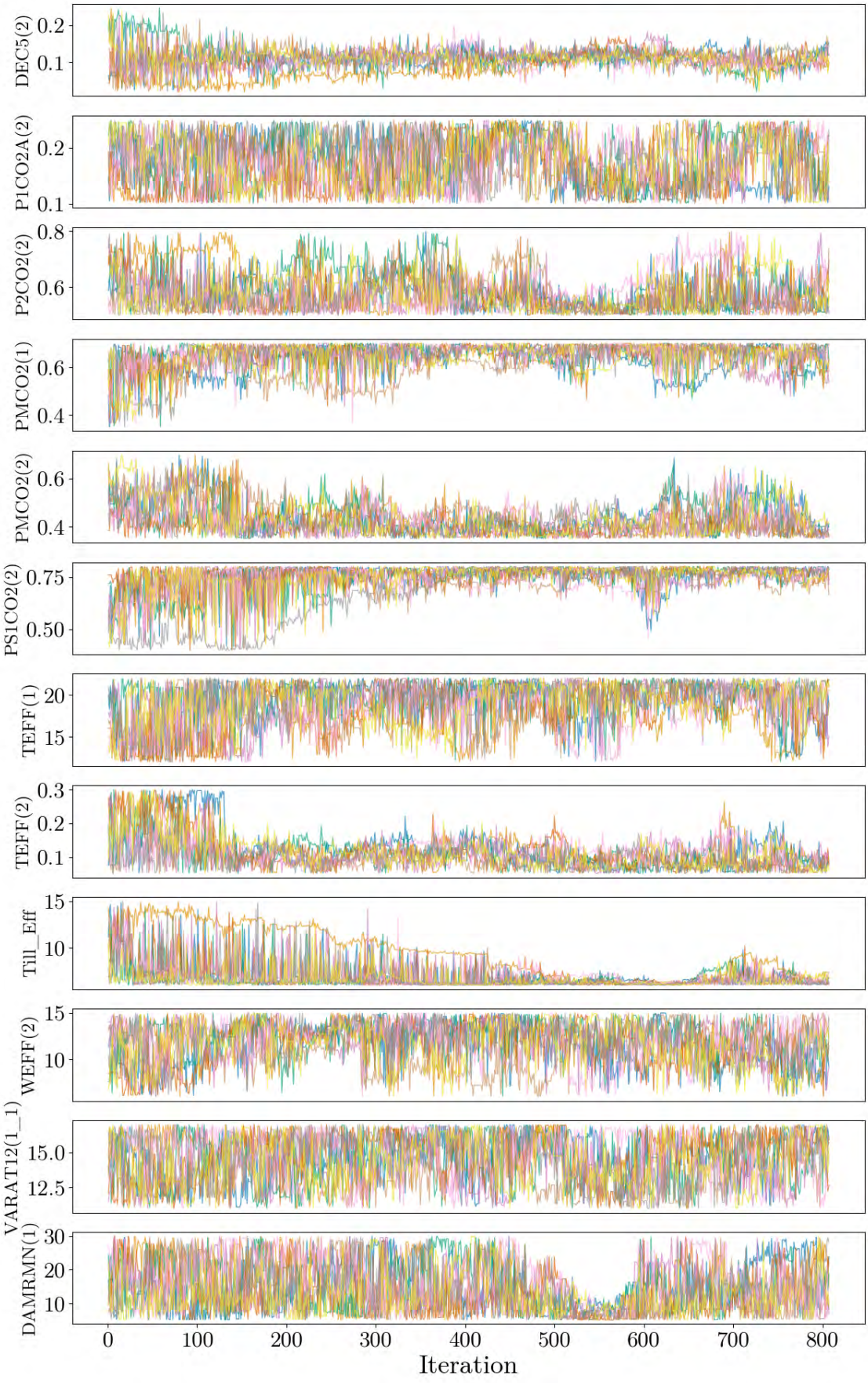

Figure C18: Traceplots for fold 5 (DayCent calibration parameters).

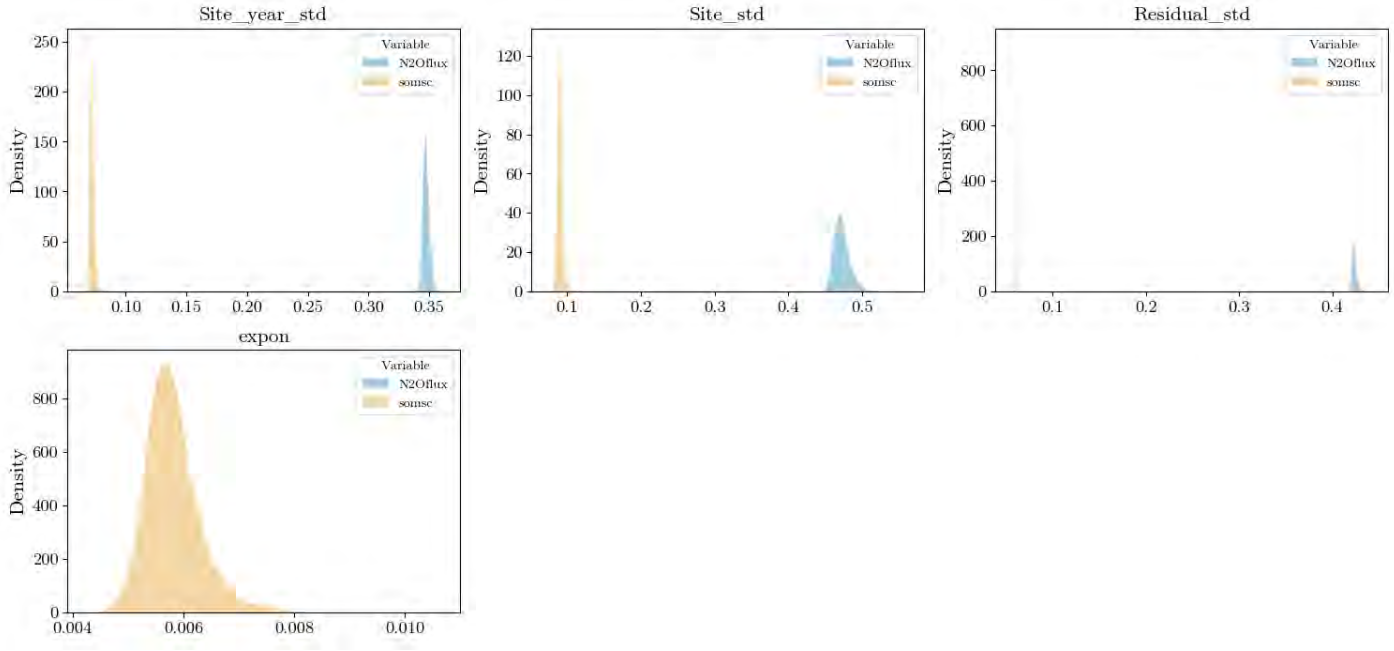

Figure C19: Marginal densities of variance parameters colored by Monte Carlo chain for fold 5.

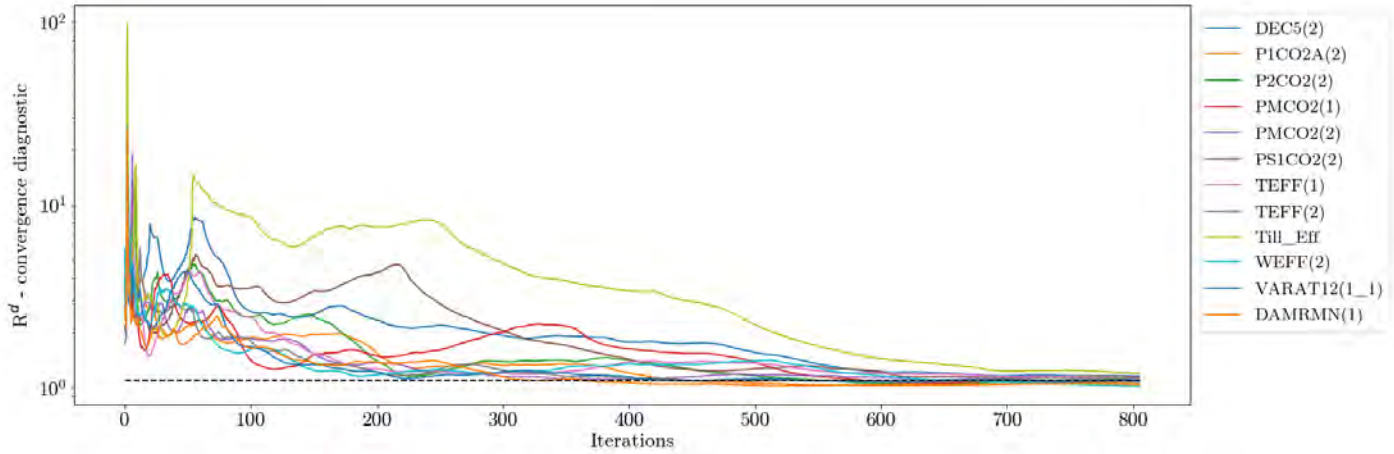

Figure C20: Gelman-Rubin (R) convergence criteria (i.e. shrinkage factors) for the DayCent calibration parameters and fold 5.

Table C9: Summary statistics of marginal posterior distributions of the calibrated DayCent-CR model parameters for fold 5 with 5th percentile, mean, median, 95th percentile and the standard deviation.

Parameter	5th percentile	95th percentile	Mean	Median	Standard deviation
DAMRMN	5.8489	24.4795	14.0914	12.9609	6.0412
DEC_5_2	0.0765	0.1382	0.1066	0.1071	0.0194
P1CO2A_2	0.1095	0.2381	0.1749	0.1806	0.0417
P2CO2_2	0.5055	0.6784	0.576	0.5657	0.0574
PMCO2_1	0.5664	0.6965	0.6461	0.6574	0.044
PMCO2_2	0.3569	0.527	0.4303	0.4213	0.0565
PS1CO2_2	0.6914	0.7971	0.758	0.7666	0.0348
TEFF1	15.904	21.6695	19.3965	19.8128	1.879
TEFF2	0.0576	0.1472	0.0933	0.0884	0.0283
Till_Eff	6.0739	7.6187	6.6147	6.4464	0.5536
VARAT12(1,1)	12.0511	16.7693	14.5698	14.6379	1.4996
WEFF2	8.0481	14.2867	11.2317	11.2663	1.9305

Table C10: Random effects for the calibration dataset.

Parameter	Variable	5th percentile	95th percentile	Mean	Median	Standard deviation
$\sigma_{\text{site-year}}$	N2Oflux	0.343257	0.351935	0.347194	0.346923	0.002687
σ_{site}	N2Oflux	0.457225	0.492323	0.472203	0.470689	0.011282
σ	N2Oflux	0.419633	0.428893	0.423362	0.422867	0.002817
$\sigma_{\text{site-year}}$	SOC	0.069462	0.075156	0.071997	0.071761	0.001917
σ_{site}	SOC	0.084876	0.096371	0.090047	0.089641	0.004227
σ	SOC	0.058933	0.060439	0.059671	0.05967	0.000461
ν	SOC	0.005135	0.006729	0.005812	0.005744	0.000512

C.6 Variograms

For this assessment, we calculated site-level averages for the residuals of differences in change in SOC stock (modeled - measured) and N_2O emission (modeled - measured) as well as for the initial SOC stock and the average N_2O flux. We then generated variograms for each of the previous quantities using approximately 75 equally sized bins.
Both sample and model variograms are shown in Figures C21, C22, C23, and C24. All model variograms use a Gaussian semivariance, and we set the nuggets to the sample semivariance in the bin with the smallest separation distance. We estimated the sample and model semivariance using the `gstat` package (*Gräler, Pebesma, and Heuvelink, 2016*) for R (*R Core Team, 2021*). To compute separation distances, geographic coordinates were projected onto the North American Datum 83 (NAD83) for the continental United States (EPSG: 6350); while two of the 31 site locations in this analysis are located in Canada (Swift Current and Lethbridge), the Canadian sites are near the border with the United States.

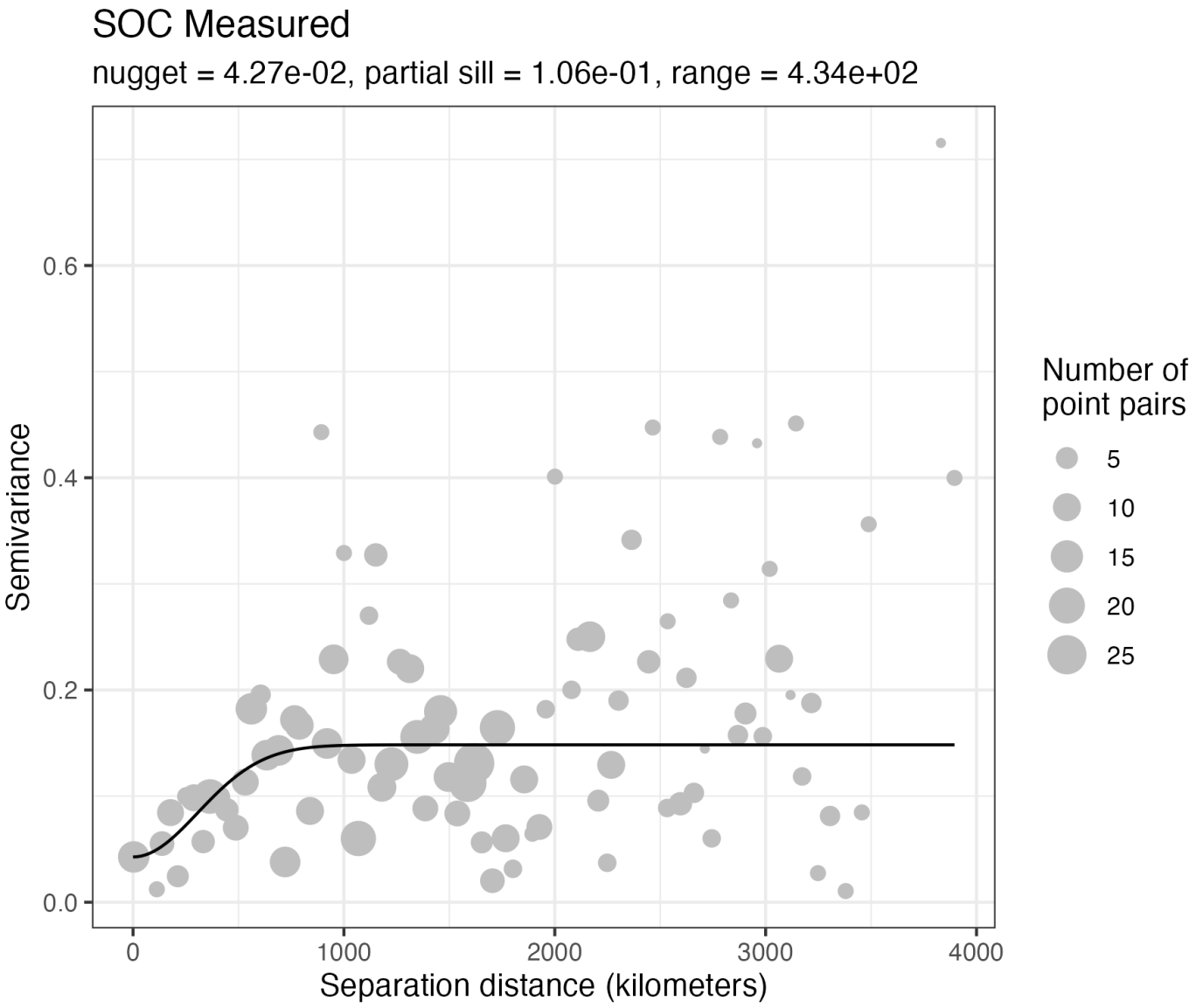

Figure C21: Variogram of the residuals of log(initial SOC stock). The grey dots show the sample semivariance, and the black line shows an estimated gaussian model semivariance.

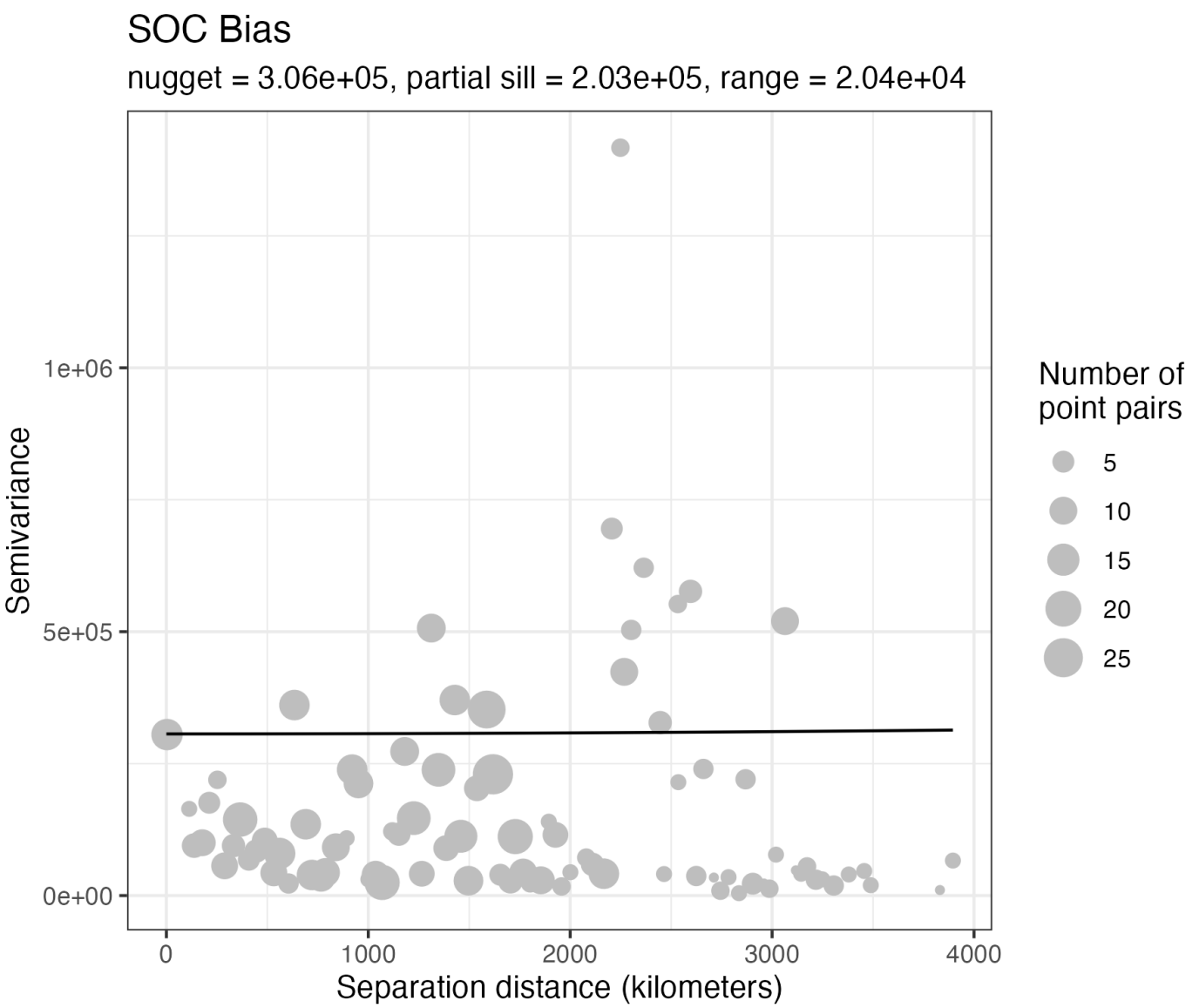

Figure C22: Variogram of the residuals of differences in change in SOC stock (modeled - measured). The grey dots show the sample semivariance, and the black line shows an estimated gaussian model semivariance.

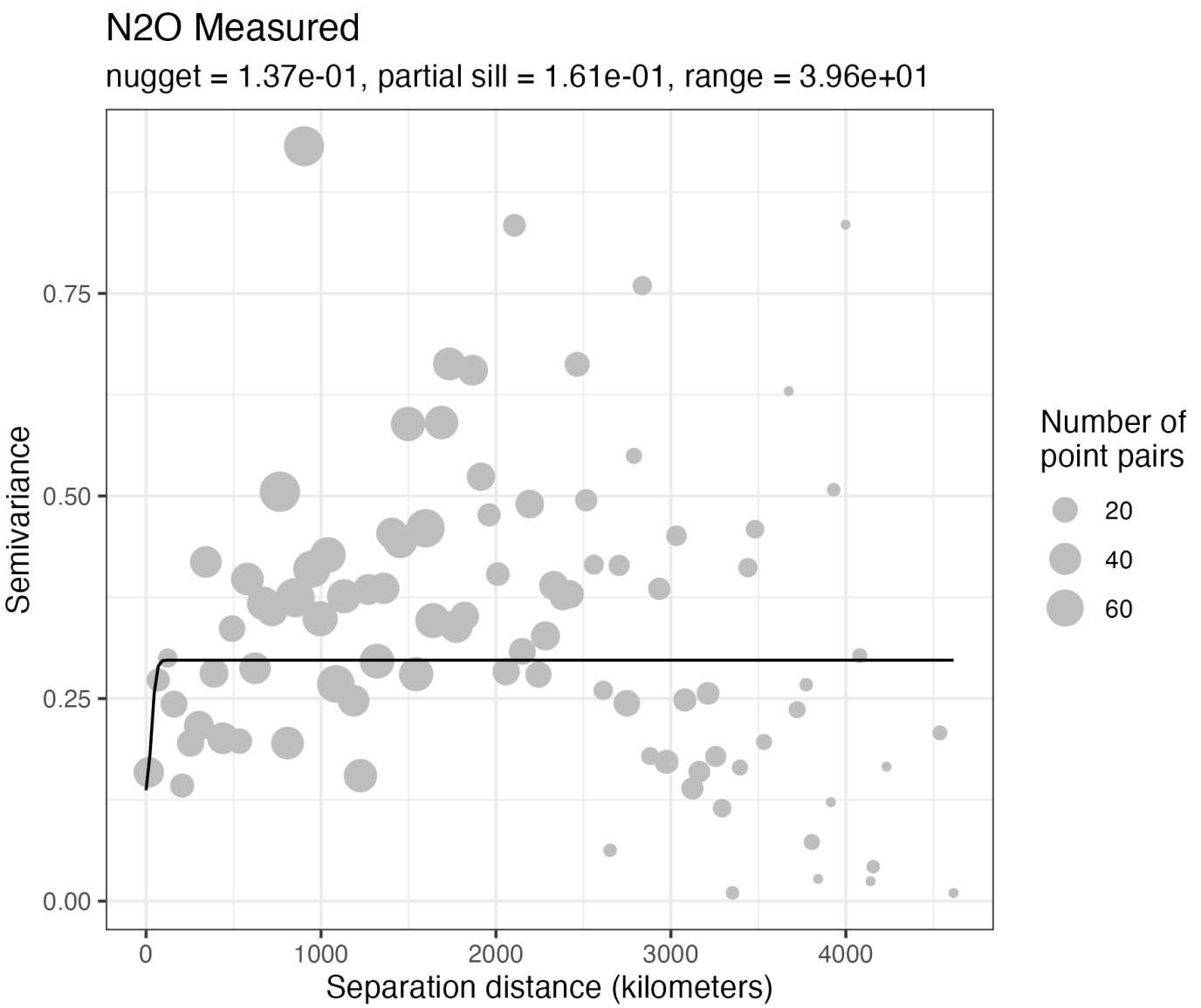

Figure C23: Variogram of the residuals of log(mean N_2O flux + ϵ), where ϵ shifts mean N2O flux > 0. The grey dots show the sample semivariance, and the black line shows an estimated gaussian model semivariance

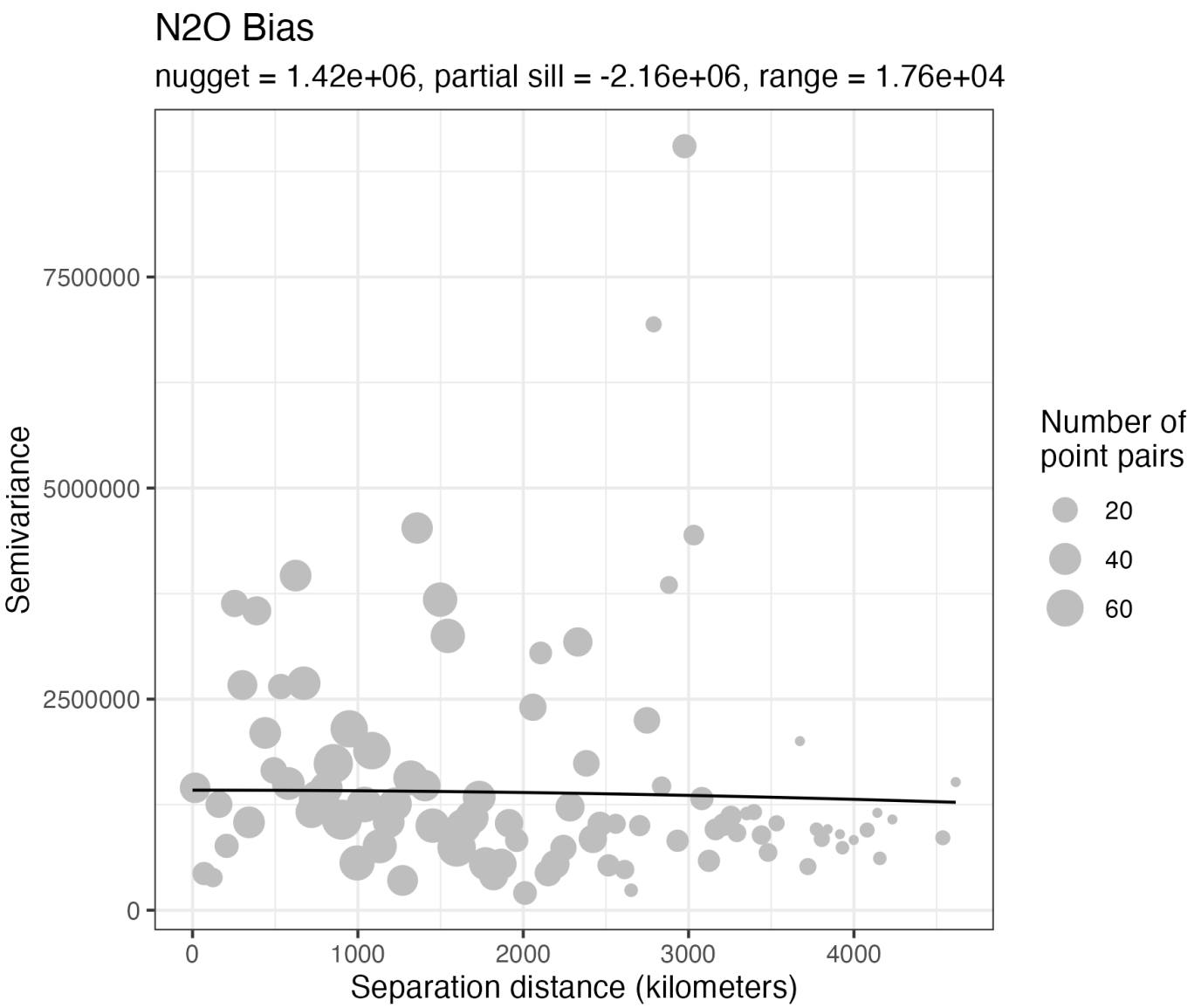

Figure C24: Variogram of the residuals of differences in change in N2O flux (modeled - measured). The grey dots show the sample semivariance, and the black line shows an estimated gaussian model semivariance.

Appendix D: Thinned vs Full Posteriors

As noted in Section 3.1 "Description of model calibration", we thin the post-burn-in chains by seven to keep 176 posterior draws from the final model fit (i.e., the model fit to all of calibration data). We use these 176 posterior draws in crediting. This section compares the distribution of these 176 posterior draws (after thinning) versus the full post-burn-in draws.

Figure 1 shows the marginal densities of the full and thinned post-burn-in chains. Cross-referencing the model description in Section 3.1 "Description of model calibration", in Figure 1 site refers to σ_{site}, siteyr referes to $\sigma_{\text{site-year}}$, and Resi refers to σ_{residual}.

Table 1: Summary statistics of the thinned posterior distributions of the calibrated DayCent-CR model parameters and seven variance parameters with mean, SD, effective sample size (ESS), and $\hat{R}$.

Variable	Full Mean	Thinned Mean	Full SD	Thinned SD	Full ESS	Thinned ESS	Full $\hat{R}$	Thinned $\hat{R}$
DEC5(2)	0.12	0.12	0.023	0.022	36	44	1.2	1.2
P1CO2A(2)	0.16	0.16	0.039	0.04	43	64	1.1	1.1
P2CO2(2)	0.57	0.57	0.056	0.053	56	87	1.1	1.1
PMCO2(1)	0.67	0.67	0.025	0.024	64	84	1.1	1.1
PMCO2(2)	0.4	0.4	0.041	0.04	79	99	1.1	1
PS1CO2(2)	0.77	0.77	0.029	0.03	67	94	1.1	1.1
TEFF(1)	17	17	2.7	2.8	51	87	1.1	1.1
TEFF(2)	0.094	0.095	0.035	0.035	41	77	1.2	1.1
Till_Eff	6.8	6.7	0.72	0.72	36	66	1.2	1.1
WEFF(2)	9.9	9.8	1.9	2	58	83	1.1	1.1
VARAT12(1,1)	15	15	1.5	1.5	43	110	1.1	1.1
DAMRMN(1)	16	17	7.3	7	45	61	1.1	1.1
$\sigma_{\text{site, n2o}}$	0.46	0.46	0.01	0.0091	220	160	1	1
$\sigma_{\text{site-yr, n2o}}$	0.32	0.32	0.0026	0.0025	110	130	1.1	1
$\sigma_{\text{residual, n2o}}$	0.41	0.41	0.0013	0.0011	73	100	1.1	1.1
$\sigma_{\text{site, soc}}$	0.084	0.084	0.0053	0.0051	63	78	1.1	1.1
$\sigma_{\text{site-yr, soc}}$	0.067	0.067	9.8e-4	6.7e-4	130	170	1	1
$\sigma_{\text{residual, soc}}$	0.058	0.058	9.1e-4	8.3e-4	90	80	1.1	1.1
ν_{soc}	0.0072	0.0072	5.9e-4	5.7e-4	100	82	1.1	1.1

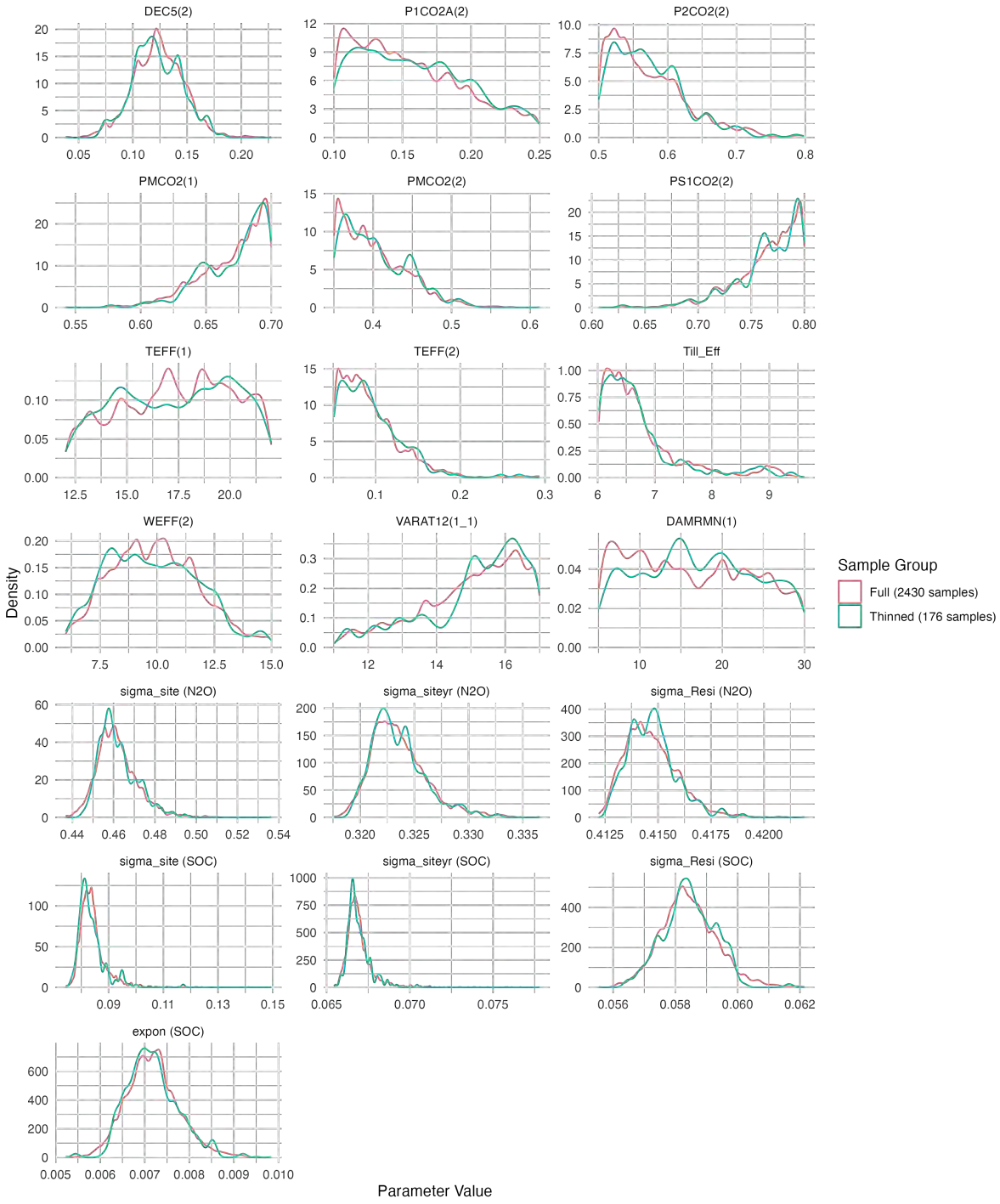

Figure 1: Marginal densities of the full and thinned posteriors from the final model fit.

Appendix E: Confidence interval width and coverage rates as function of time

As a consequence of the exponential residual variance model (see Section 4), the residual variance grows with increasing time since the start of an experiment. As a result, confidence intervals tend to become wider as the time since the start of an experiment increases.

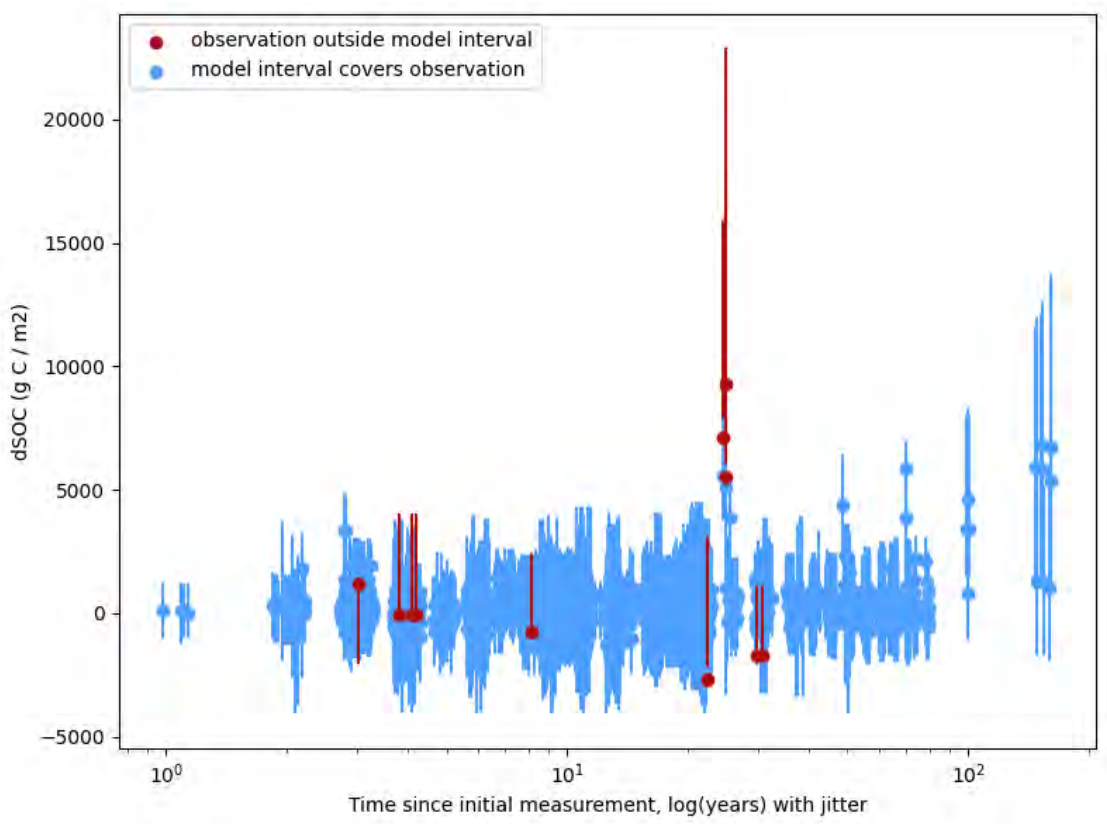

Figure E1: Confidence intervals for change in SOC as a function of number of years since experiment start. Dots show observed changes. Confidence intervals are blue if they contain the observed value and red if they do not contain the observed value. The x-axis is on the log scale, and random noise is added to the x-axis placement (i.e. the x-axis placement is jittered) for visual discernment.

To assess whether the calibrated model remains conservative when predicting SOC over the shorter periods of time used during crediting, Tables E1–E3 show the coverage rates across all folds for each PC x CFG category when filtering to point-pairs with measurements on or before 3, 5, and 10 years since experiment start, respectively. PC x CFG categories with no experimental data in these respective time ranges are not shown. Tables E1–E3 also show 95% Agresti-Coull confidence intervals for the coverage rates (Agresti and Coull, 1998).

As seen in Tables E1–E3, all categories have higher than 90% coverage over 3, 5, and 10 year time periods.

Table E1: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals across all folds of the calibration/validation process filtered to point-pairs with ≤ 3 years between experiment start and measurement.Lower and upper 95% confidence bounds are shown for coverage rates.

PC	CFG	n	n_{in}	n_{out}	coverage(%)	lower	upper
CROP	C3A	59	59	0	100	95	100
CROP	C3AN	59	59	0	100	95	100
CROP	C3PN	12	12	0	100	80	100
CROP	C3S	71	71	0	100	96	100
CROP	C4A	55	55	0	100	95	100
DISTURB	C3A	20	20	0	100	87	100
DISTURB	C3AN	23	23	0	100	88	100
DISTURB	C3S	37	37	0	100	93	100
DISTURB	C4A	47	47	0	100	94	100
NFERT	C3A	19	19	0	100	86	100
NFERT	C3AN	18	18	0	100	86	100
NFERT	C3PN	3	3	0	100	45	100
NFERT	C4A	48	47	1	97.92	91	100
ORG	Annuals	5	5	0	100	60	100
ORG	C3A	2	2	0	100	33	100
ORG	C3AN	2	2	0	100	33	100
ORG	C4A	5	5	0	100	60	100

Table E2: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals across all folds of the calibration/validation process filtered to point-pairs with ≤ 5 years between experiment start and measurement.Lower and upper 95% confidence bounds are shown for coverage rates.

PC	CFG	n	n_{in}	n_{out}	coverage(%)	lower	upper
CROP	C3A	99	98	1	98.99	95	100
CROP	C3AN	93	92	1	98.92	95	100
CROP	C3PN	18	18	0	100	86	100
CROP	C3S	95	95	0	100	97	100
CROP	C4A	63	62	1	98.41	93	100
DISTURB	C3A	26	26	0	100	90	100
DISTURB	C3AN	32	32	0	100	91	100
DISTURB	C3S	39	39	0	100	93	100
DISTURB	C4A	63	63	0	100	96	100
NFERT	C3A	25	25	0	100	89	100
NFERT	C3AN	18	18	0	100	86	100
NFERT	C3PN	3	3	0	100	45	100
NFERT	C4A	49	48	1	97.96	91	100
ORG	Annuals	7	7	0	100	69	100
ORG	C3A	3	3	0	100	45	100
ORG	C3AN	2	2	0	100	33	100
ORG	C4A	6	6	0	100	65	100

Table E3: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals across all folds of the calibration/validation process filtered to point-pairs with ≤ 10 years between experiment start and measurement.Lower and upper 95% confidence bounds are shown for coverage rates.

PC	CFG	n	n_{in}	n_{out}	coverage(%)	lower	upper
CROP	C3A	205	204	1	99.51	98	100
CROP	C3AN	206	205	1	99.51	98	100
CROP	C3PN	25	25	0	100	89	100
CROP	C3S	156	156	0	100	98	100
CROP	C4A	127	126	1	99.21	96	100
DISTURB	C3A	47	47	0	100	94	100
DISTURB	C3AN	42	42	0	100	93	100
DISTURB	C3S	43	43	0	100	94	100
DISTURB	C4A	127	127	0	100	98	100
NFERT	C3A	63	62	1	98.41	93	100
NFERT	C3AN	38	38	0	100	93	100
NFERT	C3PN	3	3	0	100	45	100
NFERT	C4A	95	94	1	98.95	95	100
ORG	Annuals	18	18	0	100	86	100
ORG	C3A	9	9	0	100	74	100
ORG	C3AN	4	4	0	100	54	100
ORG	C4A	9	9	0	100	74	100

Appendix F: Proposal for disambiguating pooled measurement uncertainty (PMU)

F.1 Background

The Model Requirements version 1.1a defines the pooled measurement uncertainty (PMU) by Eq. (1):

$$\text{PMU} = \sqrt{\frac{\sum_{i=1}^{k}(n_i - 1)\sigma_i^2}{\sum_{i=1}^{k}(n_i - 1)}} \tag{1}$$

where i indexes treatment pairs, σ_i^2 is the standard error of the difference in measured emissions between treatments, k is the number of treatment pairs for which standard errors are reported in the literature for both treatments in the pair, and n_i is the number of emissions measurements reported for each treatment. Note that in Eq. (1), σ_i^2 is the standard error, not standard deviation.

However, Eq. (1) is ambiguous when the two treatments being compared do not have the same number of measurements. In particular, let n_{i1} and n_{i2} be the number of emissions measurements taken for the first and second treatment in treatment pair i. If $n_{i1} \neq n_{i2}$, it is not clear what value to use for n_i in Eq. (1). This situation did not arise in the first validation report, but it does occur in the second validation report.

F.2 Proposal

We propose to resolve this ambiguity by following the common practice of computing a pooled variance by weighting individual variances by their respective degrees of freedom. In this case (and assuming the true variances in the two treatments being compared are equal) the degrees of freedom of σ_i^2 is $n_{i1} + n_{i2} - 2$. Therefore the PMU would be calculated using Eq. (2):

$$\text{PMU} = \sqrt{\frac{\sum_{i=1}^{k}(n_{i1} + n_{i2} - 2)\sigma_i^2}{\sum_{i=1}^{k}(n_{i1} + n_{i2} - 2)}} \tag{2}$$

where n_{i1} and n_{i2} are the number of emissions measurements taken in the first and second treatment, respectively, in treatment pair i, and $n_{i1}+n_{i2}-2$ is the degrees of freedom of σ_i^2. Eq. (2) estimates the standard deviation of the difference in mean emissions under the assumptions that: 1) the true mean differences are not the same across treatment pairs, but 2) the true variances are the same.

The impact on PMU estimates using Eq. (2) instead of Eq. (1) is minimal. For example, during model validation of DayCent-CAR 1.0.2 for Indigo Ag U.S. Project No. 1 (CAR1459) the overall PMU of SOC across all practice changes and CFGs is 623.46 using Eq. (2) (see Sections 8.3 and 8.4), and 623.43 using Eq. (1) and randomly setting $n_i = n_{i1}$ or $n_i = n_{i2}$ when $n_{i1} \neq n_{i2}$. The advantage of Eq. (2) over Eq. (1) is that Eq. (2) is not ambiguous in how to calculate the PMU when $n_{i1} \neq n_{i2}$, and it is better grounded in statistical theory and practice (Rosner 2006).