Validation Report DayCent-CR Version 1.0.2

Indigo Ag

Reviewed by:	Michael Dietze, Boston University
First submission:	April 26, 2022
Revised to address reviewer comments:	May 19, 2022 June 9, 2022 June 16, 2022
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

1 Report type

1.1 Report type

Type 1 (Project-specific, for project CAR1459)

1.2 SEP version

Version 1.0, accessed on 23 April 2022

1.3 SEP model requirements version

Requirements and Guidance for Model Calibration, Validation, Uncertainty, and Verification For Soil Enrichment Projects, Version 1.1a, accessed on 22 Mar 2022 (referred to hereafter as the "Model Requirements")

1.4 Model version

DayCent-CR Version 1.0.2.

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 1.0.2:

1. Version 1.0 build 1.0 of the DayCent-CR model executable, corresponding to SVN revision 279 of the DayCent source code repository[1]. This code was originally derived from the branch of DayCent maintained by the National Greenhouse Gas Inventory team and also used for the COMET-Farm system.
2. Version 2.0 of the DayCent-CR model parameters, corresponding to Git commit `85afb5ae2d17802007bf79e264d4100a326aec0b` of the private model parameter repository. These were originally derived from the default parameterizations for the COMET-Farm system and have been modified for carbon crediting, including during the calibration process reported here. In addition to parameter files, this component also includes `R` scripts that are used to perform Monte Carlo simulations using the calibrated parameter set.

During model simulations for project CAR1459, Indigo will submit inputs to the model using the DayCent-CR API, with initial version DCR1.0.2 build 1.0.0.23. This API was not used during calibration

[1] To access materials for academic research purposes, Indigo Ag should be contacted directly.

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 1.0.2 whether it is run directly or accessed through any technically compatible version of the DayCent-CR API.

1.5 Version confirmation materials

The following materials[2] have been provided for use by the reviewer of this report and project verifiers for CAR1459:

- 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 results (`DayCentCR_1.0.2_validation_supporting_files.zip`)
- Appendix A "Documentation of calibrated parameter sets"
- Appendix B "Declaration of Practices"
- Appendix C "Sampler diagnostics"
- Appendix D "Thinned vs full posteriors of final fit"
- Appendix E "Confidence interval width and coverage rates as function of time"
- Appendix F "Variance inflation factor for ORG x All category"
- Appendix G "Proposal for disambiguating pooled measurement uncertainty (PMU)"
- Supporting document `Appendix_D_revision_v4.docx`

All version confirmation 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 files.

1.6 Changes from previous validation report

DayCent-CR version 1.0 has been previously validated and approved for crediting of SOC in Indigo U.S. Project no. 1. For the current validation of DayCent-CR Version 1.0.2, we have made the following changes from version 1.0:

1. The pooled measurement uncertainty (PMU) calculations use Eq. (7) in Appendix G. This change was necessary to address an ambiguity that arose in the second validation report but not in the first validation report. Please see Appendix G for details.

[2] To access materials for academic research purposes, Indigo Ag should be contacted directly.

2. The variance structure fitted during calibration now accounts for the way simulation errors build on each other with time, by allowing the residual variance to scale exponentially with the length of the simulation and thus to attribute more uncertainty to a result that depends on many decades of model time and less uncertainty to a result that has had little time to diverge from the precisely-known initial measurement. For the relatively short simulations used during crediting, this usually results in reduced uncertainty compared to the approach used in the previous report but is still demonstrably conservative (more than 90% of the 90% predictive intervals cover the observed value). See Section 4.1 for further details. In addition, we apply a variance inflation factor to the ORG x All category to ensure conservativness (see Appendix F) and address potential concerns raised in Appendix E.

3. 8 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 three PCs for SOC changes in the newly-added cotton CFG. See Section 5 for details.

4. Three of the validation sites that are new in this report were excluded from the previous report because of insufficient SOC sampling depth, but we have now recovered estimates for these by identifying additional data from other publications. See Section 4.4 for details.

5. Six appendices detailing changes from the Model Requirements (Appendices A–D and H of the previous report) have been removed, because they are no longer needed when using version 1.1a of the Model Requirements.

6. We have removed the restriction of valid model predictions to SOC changes smaller than 5000 g C m^{-2} that was reported in the previous report, as uncertainty coverage in the current validation was adequate across the full range of observed dSOC. We continue to note that 5000 g C m^{-2} is a larger change than we expect to observe within the crediting period of any individual field. See Section 13 for details.

2 Introduction

This report describes the validation of DayCent-CR for use in modeling changes in the emissions source soil carbon for carbon crediting as part of CAR1459, Indigo U.S. Project No. 1.

DayCent-CR is a process-based ecosystem biogeochemical model which 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. The DayCent model (e.g. see Parton et al., 2001; Del Grosso et al., 2006; Del Grosso et al., 2012; Zhang, Suyker, and Paustian, 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, 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 N transformations including N_2, N_2O and NO_x emissions and CH_4 oxidation and emissions from soil. The model is used to estimate net

CO_2, N_2O, and CH_4 emissions from soils in the US national greenhouse house gas inventory submitted by US EPA to the UN Framework Convention on Climate Change. The DayCent model is included within the COMET-Farm platform that implements USDA's entity-scale greenhouse gas inventory methods (Powers et al., 2014) and the model is implemented as part of the Climate Action Reserve's protocol for avoided conversion of grassland (http://www.climateactionreserve.org/how/protocols/grassland/)

The version of the model validated in this report for the CAR Soil Enrichment Protocol project CAR1459 is based on the latest version of the model developed to simulate soil organic matter dynamics to 30 cm soil depth, with additional improvements to several soil and plant processes as documented in Gurung et al. (2020). This version, known as DayCent-CR, is structurally the same as documented in Gurung et al. (2020), with one exception: the procedure used to initialize total soil organic C and N and its distribution across the kinetically-distinct organic matter pools in this version of the model has been adapted to use initial estimates of soil organic carbon based on lab measurements of field sampled soils (see Section 4.2 "Model setup" for details) and soil organic N pools based on the modeled C:N ratios of each SOC pool. This allows the model to operate in compliance with SEP Protocol section 5, using the required direct measurements of soil organic carbon (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 Validation Report.

The DayCent-CR version evaluated in this report uses the DayCent executable compiled from source code with Revision Number 279 in the Subversion system (TortoiseSVN software), used to manage versions of source code and default parameter sets for the simulation. This executable is the same as the one used in Gurung et al. (2020) except for a small change to the file reading system; otherwise it produces exactly the same outputs.

3 Responsible parties

Calibration, validation, and running of DayCent-CR for this project were all performed by Indigo Ag, which is also the project developer of CAR1459. As required in Section 5 of the SEP Model Requirements, Indigo Ag has the requisite expertise to calibrate and validate DayCent-CR for model performance and uncertainty, including the entire team formerly at Soil Metrics, LLC who were approved by CAR for the validation of DayCent-CR version 1.0 on February 12, 2021 (https://soilmetrics.eco/our-team/).

C. Black lead the model validation process, oversaw alignment of available datasets with the SEP Model Requirements, performed data analysis of validation datasets and model outputs, and lead the writing of this report. B. Segal lead implementation of improvements to the model calibration process, performed data analysis of model outputs for calibration and validation, and contributed to the writing and review of this report. Y. Zhang performed modification to the model version to use direct soil measurements to initialize the model, implemented the initial version of Bayesian calibration and cross-validation, and reviewed modifications to the calibration process for the current model version. S. Williams performed literature searches and assembled the validation dataset. M. Easter created the input files required to run the model for each specific site. K. Brown and M. Easter developed APIs for high-throughput processing of model inputs and outputs. K. Paustian supervised the project and provided review of all model development and data collection. R. Gurung developed the Bayesian calibration approach and provided technical guidance on its implementation for this report. M. Motew helped with interpretation of the SEP Model Requirements criteria and provided review of the report. M. DuBuisson and M. Walker oversaw compliance with SEP requirements and provided review of all Validation Report materials. K. McAllister was the project manager and led the development of timelines, the execution of deliverables, and alignment with stakeholders and company priorities. N. Campbell was the project supervisor and provided review to all Validation Report materials.

4 Model Calibration

Follows Model Requirements Section 2 Summary of Requirements (p8)

4.1 Description of model calibration

DayCent-CR Version 1.0.2 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 (see below for details). The joint posterior of DayCent parameters was estimated using the DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt and Ter Braak, 2011; Vrugt, 2016), which is a Markov Chain Monte Carlo (MCMC) algorithm. The DREAM algorithm has been used by Zhang, Arabi, and Paustian (2020) to calibrate DayCent for crop growth/production. To calibrate DayCent-CR for modeling SOC stock and stock change, we used the likelihood function proposed by Gurung et al. (2020) with the exception that we allow for heterogeneous residual variance whereas Gurung et al. (2020) assumed homogeneous residual variance. This function accounts for 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 Eq. (1):

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

where $\boldsymbol{\theta}$ is a vector of parameters that are used by DayCent to predict SOC or that define the variance-covariance matrix Σ, $\hat{\boldsymbol{y}}_{\text{mod}}$ and $\boldsymbol{y}_{\text{obs}}$ are vectors of natural log-transformed SOC values (modeled and observed, respectively), and n is the number of observations. Both $\hat{\boldsymbol{y}}_{\text{mod}}$ and Σ are functions of the parameters $\boldsymbol{\theta}$.

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 $\sigma^2_{\text{resid}} = \sigma^2\exp(2t\nu)$ 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 are included in $\boldsymbol{\theta}$. 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 and Bates, 2000) and an exponential residual variance model that is a function of years since the first measurement. See the supplement to Gurung et al. (2020) for additional discussion of this approach. These models were fit with the R package `nlme` (Pinheiro et al., 2022) using the `lme` function as part of the likelihood evaluation for each MCMC iteration. Thus these variance parameters were estimated via restricted maximum likelihood (REML) applied to the marginal model after plugging in the Monte Carlo draws for the DayCent calibration parameters. This estimation procedure is similar to empirical Bayes (see Casella (1985) and Carlin and Louis (2009, Ch. 5)). However, in empirical Bayes, prediction and inference would be based on a single set of variance parameter estimates, whereas we base prediction and inference on a distribution of variance parameter estimates. As a result, we expect our approach 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 was chosen 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 "Bias evaluation" and Section 10 "Model prediction error"). Note that at time 0 ($t = 0$), the residual variance becomes $\sigma^2_{\text{resid}} = \sigma^2 \exp(2t\nu) = \sigma^2 \exp(0) = \sigma^2$. In other words, σ^2 is the residual variance at time zero. As shown in Appendix A, both σ^2 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 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 R (R Core Team, 2021) using the DREAM package (Guillaume and Andrews, 2012). The DREAM algorithm is described in detail by Vrugt et al. (2009). The calibration was run with 11 MCMC chains, all of which were run until the $\hat{R}$ statistic of Gelman and Rubin (1992) dropped below 1.1 (900-2700 iterations depending on fold), suggesting convergence of the posterior distribution of model parameters. The first 50% of each chain was discarded as the "burn-in" period and the remaining 50% of each chain was used to summarize the posterior of the parameters $\boldsymbol{\theta}$[3]. Traceplots and Gelman-Rubin $\hat{R}$ 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 $k = 5$. 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. If a given experiment was assigned to a fold, all the individual observations associated with that experiment were then assigned to that fold (see Section 4.4 "Justification for splitting of experimental data" for details).

[3] The post-burn-in simulations were thinned further to reduce computational burden during crediting, as described in the k-fold fold validation steps.

The fold configuration used in this report is provided in Table A2 and a map of study sites in Figure 1.

2. Second, for each fold (fold $= 1, 2, \ldots, 5$) one group was reserved for validation and the remaining four groups were used for model calibration, giving approximately an 80%-20% split between calibration and validation datasets, respectively.

3. Third, for each fold, 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% of each chain was discarded as the "burn-in" period. This resulted in 450–1350 post-burn-in iterations (depending on fold) for each of 11 chains, for a total of 4,950–14,850 posterior draws.

4. Fourth, out-of-sample predictions were made in each fold 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 distribution of SOC differences between the experimental treatments at the second time point, similar to the methods described in Gurung et al. (2020).

5. Fifth, model performance was quantified by computing model bias, RMSE, and 90% 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[4]. Whereas we need accurate estimates of tail probabilities to assess validation criteria (i.e. coverage rates of 90% intervals), for crediting we only need accurate estimates of variance, so require fewer posterior draws; 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 and Hinkley (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 11 "Model validation outputs for use in SEP uncertainty calculations" for a description of how the saved posteriors are used during crediting, and Appendix D "Thinned vs full posteriors of final fit" for a comparison between thinned and full posteriors.

The prior distributions of parameters adjusted during the calibration process (Table A1) and summary statistics of marginal posterior distributions of model parameters (Table A3) for the final step using the full dataset are provided. For the full parameter set and auxiliary files needed to reproduce the validation, please see the supplement `DayCentCR_1.0.2_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 of the available training data. This approach complies with Section 2.3.1.2

[4]From each of the 11 chains, we kept n_{thin} evenly spaced iterations from the post-burn-in period, where $11 \times n_{\text{thin}} = 176$.

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 we claim that 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 k separate parameter sets that will all 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, we completed a comparison between the parameter distributions obtained from cross-validation and from fitting the full dataset (Figure A1), as well as between the distributions of model outputs (Figure 66 and Section 12) to ensure differences between the validated and final parameterizations, particularly for the most sensitive parameters, are not such as would materially change model results.

4.2 Model setup

For calibration and validation, we ran DayCent-CR for all treatments and sites (See site-level summaries in tables of Section 8 "Documentation of validation and calibration datasets, per CFG-PC-ES combination", and full dataset in data file `validation_data_datapoints.csv` in `DayCentCR_1.0.2_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 6 "Description of data requirements". Where site-specific data were not available from the experimental publication, we used soil data (texture and pH, which were then used to estimate other missing soil parameters) from the gSSURGO database (Soil Survey Staff, 2022), management information estimated from typical agronomic practice in the region (see Section 6.2 "Management information" and Section 6.3 "Procedures for missing data" for details), and climate data (minimum and maximum daily temperature, precipitation) from the PRISM database (`http://prism.oregonstate.edu`) for the experiments located in the United States. The nearest weather station was used for sites in the United Kingdom (Barré et al., 2010) and Canada (Environment and Climate Change Canada: `https://climate.weather.gc.ca/historical_data/search_historic_data_e.html`. For sites in Brazil temperature and precipitation were obtained from the SWAT global weather data (`swat.tamu.edu/data/cfsr`), and for Australia temperature was obtained from SWAT and precipitation was obtained from the nearest weather station (Australian Government; Bureau of Meteorology; `http://www.bom.gov.au/climate/data/`).

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, we conducted equilibrium simulations of native grassland (5000 to 7000 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 US National Greenhouse Gas 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 the measured value, simulations of the experimental period were used to perform the calibration and validation process (see Section 4.1 "Description of model calibration").

14 of the 41 experimental sites that generated 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 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 an SEP project, and calibration and validation are constrained to the time periods for which SOC observations are available. We note that 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 we believe that 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 an SEP project, using site latitude and longitude, soil C measurements, and soil physical and chemical properties (described in Section 6 "Description of data requirements"). 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 SEP projects for the initial period simulated to reach a model steady-state (consistent with the US 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 SEP 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 SEP requirements and will be described in the CAR1459 Monitoring Plan.

4.3 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 US GHG inventory simulations (U.S. EPA, 2020), and not all model parameters have an impact on SOC dynamics. Therefore we selected 28 parameters to consider in the calibration exercise (Table A1). These consisted of 27 parameters directly related to SOC processes and to DayCent-CR's soil organic matter decomposition sub-routine, plus one parameter related to soil water ("FWLOSS(2)", which scales potential evapotranspiration) that was chosen by scientist Yao Zhang (a Responsible Party to this report) based on his previous work developing DayCent water modeling processes. These parameters were selected because they control the decay rate of the SOC pools and C transfer efficiency between pools and directly affect the magnitude of SOC stocks and SOC stock differences. Other parameters associated with other processes, such as plant production, influence modeled SOC but in an indirect manner mediated by the selected parameters. Consequently, they were left as constants and assigned the default values used in COMET-Farm and the US GHG Inventory.

All 28 parameters were assigned independent uniform prior distributions defined by the lower and upper bounds shown in Table A1. The initial list of parameters and their prior ranges were taken from the values reported in Gurung et al. (2020) and are based on theoretical understanding, previous studies by the team of scientists at Colorado State University where the model was developed, initial testing of model algorithms, and direct review and input from William J. Parton who created the Century model and was the primary investigator on many of the previous studies. This initial list was then updated by Y. Zhang 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.

A variance based Global Sensitivity Analysis (GSA) was performed on the calibration dataset used for the validation of DayCent-CR version 1.0, using the method of Sobol (1993) to identify the model parameters most strongly controlling SOC response. The GSA quantified the relative importance of the parameters that have a substantial influence on model output, and allowed us to identify which group of parameters was most important. The Sobol method implements a Monte Carlo simulation to propagate parameter uncertainty from the priors to uncertainty in model outputs. Similar to analysis of variance, the method partitions the total variance of the model output into first-order and higher-order interaction terms to estimate the proportion of variance explained by each parameter. The method is model independent and has been previously used with a closely related version of the DayCent model in Gurung et al. (2020). The total sensitivity indices for each of the 28 parameters are plotted in Figure C17. From this analysis, we identified 10 parameters that each contributed more than 0.5% of variance (see Table A1). This very inclusive cutoff was chosen to reduce the dependence of the GSA on the calibration dataset.

Bayesian calibration was performed on the 10 most influential parameters shown in Table A1, 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) (Table A3), as described in Section 4.1.

The prior ranges for all DayCent parameters included in the Bayesian calibration, along with descriptions of each parameter, are provided in Table A1. The summary statistics of the marginal posterior distribution are provided in Table A3 and Figure A1. Sampler diagnostics, marginal posterior plots, and parameter correlation plots are available in Appendix C. Only the parameters included in the GSA are shown in Appendix A; for the full parameter set and auxiliary files needed to reproduce the validation, please see supporting file `DayCentCR_1.0.2_validation_supporting_files.zip`.

During the calibration process, instead of estimating the posterior distribution of model parameters for each LRR separately, we treated the model parameters as population-level variables and conducted a single calibration for all LRRs. Because we use the joint posterior from this single calibraiton 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.

4.4 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 carbon 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, we assigned experimental sites 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/PC combination, per Model Requirements, Section 2), the data from these experiments may be correlated in space (climate and soil factors, conditions during model spinup) but are likely uncorrelated in management. Therefore, these experiments were considered as separate "sites" and were separately 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, we also created spatial variograms of log(initial SOC stock), measured and modeled SOC change, and model residuals after calibration (see Figures C1–C4 in Appendix C.1). While there apears to be spatial correlation in log(initial SOC stock), the range of spatial correlation was estimated to be only 217 kilometers, and there does not appear to be spatial correlation in the modeled or measured differences in SOC stock change or in residuals.

The data from 3 sites (brookings_REAP, dalhart, tribune) were late additions to the dataset that were excluded from calibration because of time constraints but were included in the validation by randomizing each site into one fold of the dataset when making out-of-sample predictions.

5 Project domain

Follows Model Requirements, Sections 3.1 and 3.2, and Summary of Section 3.2 (p10)

5.1 Practice categories

The project intends to credit 14 practices (Appendix B) falling into four Practice Categories (PCs):

- “Inorganic N fertilizer application” (NFERT)
- “Organic amendments application” (ORG)
- “Soil disturbance and/or residue management” (DISTURB)
- “Cropping practices” (CROP)

5.2 Crop functional groups

The project includes crops spanning four crop functional groups (CFGs):

- Annual, C4, herbaceous, non-N-fixing, non-flooded crops (“corn”)
- Annual, C3, herbaceous, non-N-fixing, non-flooded crops (“wheat”)
- Annual, C3, herbaceous, N-fixing, non-flooded crops (“soy”)[5].
- Annual, C3, shrubby, non-N-fixing, non-flooded crops (“cotton”)

5.3 Land resource regions

The project encompasses 16 LRRs (Table 1) and 8 IPCC climate zones (Table 2).

[5]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

Table 1: Land Resource Regions (LRRs) occurring in the project area.

LRR	Name
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

Table 2: Climate zones defined by IPCC (2019) appearing in the project

ipcc_climate_zone	ipcc_climate_zone_abbrev
warm temperate dry	WTD
cool temperate dry	CTD
warm temperate moist	WTM
cool temperate moist	CTM
boreal moist	BM
boreal dry	BD
tropical moist	TrM
tropical dry	TrD

5.4 Soils

The project includes all 12 soil textures in the USDA soil texture classification, listed in Table 3 along with the clay contents at the midpoint of each texture class definition.

Table 3: Names, abbreviations, and midpoint clay contents for USDA soil texture classes occurring in the project area.

Abbreviation	Texture class	% 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

5.5 Emission sources

The model was validated for changes in soil organic carbon. Emissions of CH_4 and N_2O are not included in this report.

5.6 Domain covered by this validation

The domain validated in this report includes a total of 12 combinations of CFG, PC, and emissions source (ES), as summarized in Table 4). Additionally, following Model Requirements section 3.3.1 paragraph 5 (allowing multiple project CFGs to be aggregated when validating the ORG PC), we present a combined dataset for the ORG PC from all annual crops combined, with an additional variance inflation factor to ensure conservatism (Appendix F). Data for ORG x corn x SOC and ORG x soy x SOC are provided for context only, as we consider them validated by the combined ORG x All x SOC dataset. Further, we follow Model Requirements section 3.3.1 paragraph 7 (allowing cropping systems that use irrigation as a background practice to not require validation of the WATER PC). This provision is valid because at least one study in the validation dataset uses irrigation as a management practice. The range of precipitation regimes included in the validation dataset (173-1627 mm yr^{-1}), covering at least 3 LRRs, are considered an adequate proxy for testing the effects of artificial rainfall.

Table 4: Combinations of CFG and PC that are validated for SOC in this project

PC	corn	cotton	soy	wheat
CROP	+	+	+	+
DISTURB	+	+	+	+
NFERT	+	-	+	+
ORG	Via ORG x All	Via ORG x All	Via ORG x All	+

Table 5: 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 the data declaration table for each PC x CFG combination in Section 8 "Documentation of validation and calibration datasets, per CFG-PC-ES combination" for counts of stacked and unstacked observations.

PC	CFG	n sites	n observations	LRRs	climate zones	countries	soils	clay range
CROP	corn	17	210	C, H, L, M, P, S	CTD, CTM, TrM, WTD, WTM	Brazil, USA	Cl, Lo, SaLo, SiCl, SiClLo, SiLo	40
CROP	cotton	6	162	C, P	TrM, WTD, WTM	Australia, Brazil, USA	Cl, ClLo, SaLo	54
CROP	soy	20	295	C, H, L, M, P, S	CTD, CTM, TrM, WTD, WTM	Australia, Brazil, Canada, USA	Cl, ClLo, Lo, SaLo, SiCl, SiClLo, SiLo	54

CROP	wheat	23	326	C, H, L, M, P, S	CTD, CTM, WTD, WTM	Australia, Canada, USA	Cl, ClLo, Lo, SaClLo, SaLo, SiCl, SiClLo, SiLo	54
DISTURB	corn	13	225	K, L, M, N, P	CTD, CTM, TrM, WTD, WTM	Brazil, USA	Cl, Lo, SaLo, SiClLo, SiLo	40
DISTURB	cotton	4	49	C, P	TrM, WTD, WTM	Australia, Brazil, USA	Cl, ClLo, SaLo	43
DISTURB	soy	9	66	C, L, M, N, P	CTD, CTM, TrM, WTD, WTM	Brazil, USA	Cl, ClLo, Lo, SaLo, SiClLo, SiLo	40
DISTURB	wheat	11	87	B, C, F, G, H, L, M, P	CTD, CTM, WTD, WTM	USA	ClLo, Lo, SaLo, SiClLo, SiLo	30
NFERT	corn	15	166	C, E, H, K, L, M, N, P, S	CTD, CTM, WTD, WTM	USA	ClLo, Lo, LoSa, SaLo, SiClLo, SiLo	25
NFERT	soy	7	77	C, L, M, P, S	CTM, WTD, WTM	USA	Lo, SaLo, SiClLo, SiLo	25
NFERT	wheat	14	173	B, C, F, H, L, M, P, S	CTD, CTM, WTD, WTM	Canada, England, USA	Lo, SaLo, SiClLo, SiLo	25

ORG	All	10	58	B, C, E, L, M, S	CTD, CTM, WTD, WTM	Canada, England, USA	ClLo, Lo, LoSa, SiClLo, SiLo	29
ORG	corn	6	15	C, E, L, M, S	CTD, CTM, WTD, WTM	USA	ClLo, Lo, LoSa, SiClLo, SiLo	25
ORG	soy	3	7	C, M, S	WTD, WTM	USA	Lo, SiClLo, SiLo	18
ORG	wheat	8	53	B, C, M, S	CTD, CTM, WTD, WTM	Canada, England, USA	ClLo, Lo, SiClLo, SiLo	22

6 Description of data requirements

Follows Model Requirements, Section 3.3 Summary of Requirements (p14)

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

6.1 Site-specific model drivers

- Daily weather data for the site and time period to be simulated: precipitation, maximum and minimum temperature, and optionally solar radiation, relative humidity, and windspeed. When the optional weather inputs are not provided, the model estimates them 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.
- Initial SOC stock in the 0–30 cm soil layer
- Depth to bedrock
- Site latitude

6.2 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 dates, methods, and types (e.g. grain, hay %offtake, fruit, etc.)
- Residue management (e.g. burning, straw/stover removal)
- Nitrogen fertilization dates, types, amounts, and application methods
- Herbicide dates and types
- Irrigation dates, types, amounts
- Organic matter addition dates, types (e.g. manure, green manure, compost, straw amendments, N fraction, C:N ratio, mass of the dry fraction)

6.3 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. 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 simulation periods much before the experiment, common regional practices can be derived from available sources on crops grown, tillage and fertilizer inputs (NASS, ERS-ARMS, CTIC). Where more soils detail is needed than provided in the material on hand, information is pulled from USDA Web Soil Survey for the soil series mentioned in the publications.

7 Description of validation data collection process

Follows Model Requirements, Section 3.3 Requirement 1 (p12)

All studies used for model validation were identified from a database of long-term SOC experiments that is contributed to and maintained by DayCent model developers from multiple research teams. This database tracks experiments found in peer-reviewed literature that report effects of management on soil organic carbon. 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 US National GHG Inventory (U.S. EPA, 2020), in which the DayCent model simulates US agriculture GHG emissions for reporting to the UNFCC.

These experiments are considered to have sufficient management detail and reliable soils information to support model testing and development activities, i.e. all parameters listed above in Section 6 "Description of data requirements" were reported, or could be inferred according to the procedures reported above in Section 6.3 "Procedures for missing data". 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 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 database is believed to contain effectively all publicly reported long-term soil research sites where the effect of agronomic practices on soil carbon have been experimentally evaluated for at least three years, measured at two or more timepoints, and reported in sufficient detail to allow parameterization of DayCent-CR models that match the experimental conditions. Much effort by the DayCent model development team has gone into assembling all relevant publication and databases associated with each experiment modeled. This includes all 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 2021 were considered for inclusion.

For this validation and calibration, data were evaluated from 152 sites reporting SOC changes in cropland. Sites were excluded only when they failed one or more of the following criteria:

- Sufficient information was provided to model the site accurately, as described above in Section 6 "Description of data requirements" or missing data could be inferred according to procedures reported above in Section 6.3 "Procedures for missing data".
- SOC was measured to a depth of 30 cm, or to depths allowing a reasonable approximation to 30 cm (not less than 23 cm) by interpolation across the depths that were reported. See supplementary data sheet `validation_data_datapoints.csv` in `DayCentCR_1.0.2_validation_supporting_files.zip` for details of the transformations applied to each measurement. Most of the excluded sites were excluded at this stage 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 SOC measurement was not taken at the onset of the experiment, only the data from timepoints after the first SOC measurement were used.

- When a study was conducted outside the United States, the IPCC climate zone of the site could be determined and was included in the Project's domain (Table 2).

After this evaluation process, 41 sites (Figure 1) were identified as usable for the calibration and validation process, collectively containing 291 treatments and 668 measurements that could be combined into 1,018 pairs of observed practice-change effects. 1 of these observation pairs were from PC x CFG combinations not validated in this report (1 H2O x wheat), and these pairs were included in the calibration runs (therefore allowing the final parameter set to be informed by these observations) but are not used for model validation in this report.

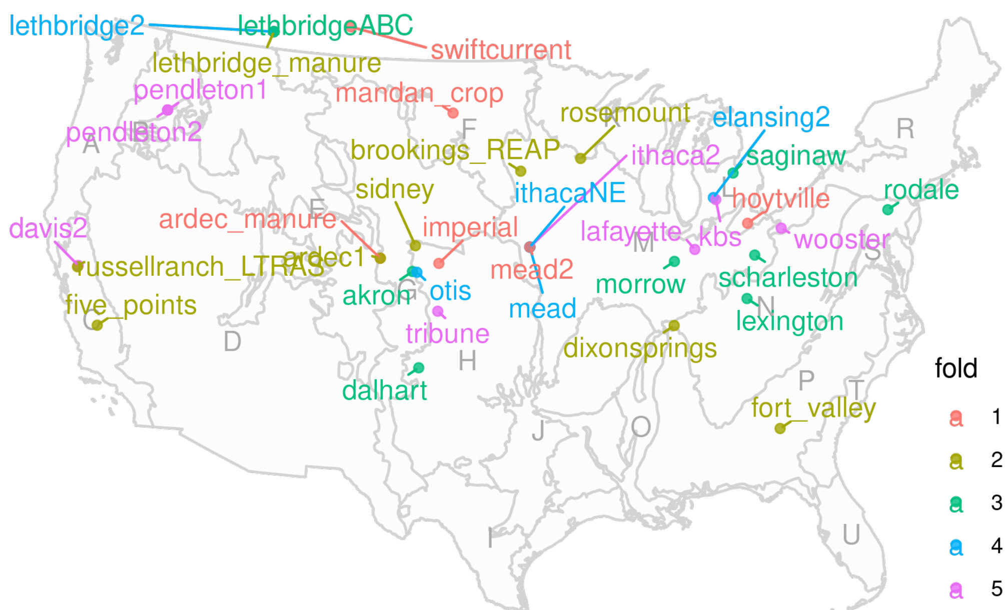

Figure 1: Locations of experimental sites used for calibration and validation of DayCent-CR. Land Resource Regions are shown in grey. Colors indicate which fold of the cross-validation held this site out for validation. Sites outside of North America not shown: broadbalk (Rothamsted, England; fold 4); goias (Goias, Brazil; fold 1); narrabri_field6 (Narrabri, New South Wales; fold 4); narrabri_fieldD1 (Narrabri, New South Wales; fold 3); narrabri_fieldC1 (Narrabri, New South Wales; fold 1) .

Where information from multiple publications was combined for a single validation point, all publications used are included in the citation list for that site (Tables 6–20, Section 8). 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.

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 publication reported a variance, standard deviation, or standard error for that treatment. Because the database was originally compiled for validation of individual treatments rather than of the differences between them, uncertainties were not extracted that were reported for differences between treatments rather than for the individuale treatments. In particular, this means that the PMUs reported here contain no observations from papers whose uncertainties were reported solely in forms such as least significant differences, HSD tests, or MSE values from ANOVA results.

8 Documentation of validation and calibration datasets, per CFG-PC-ES combination

Follows Model Requirements Section 3.3 Summary of Requirements (p14)

Throughout this section, we use the abbreviations for IPCC climate zones shown in Table 2, for soil textural classes shown in Table 3, and for Land Resource Regions shown in Table 1.

8.1 CROP x corn x SOC

This category's validation is usable in all project LRRs and soil textures because:

- The selected studies span 6 LRRs (C, H, L, M, P, S), 4 of which (H, L, M, S) are in the declared project domain, as well as one site outside the US that is within the declared project climate zones (WTM).
- 6 soil textures are included, all of which are in the declared project area: Cl, Lo, SaLo, SiCl, SiClLo, SiLo.
- Clay content spans 40 percentage points, from 10% to 50%.
- At least one study isolates effects, i.e. only 18 of the 210 pairs of observations compare stacks of PC changes.

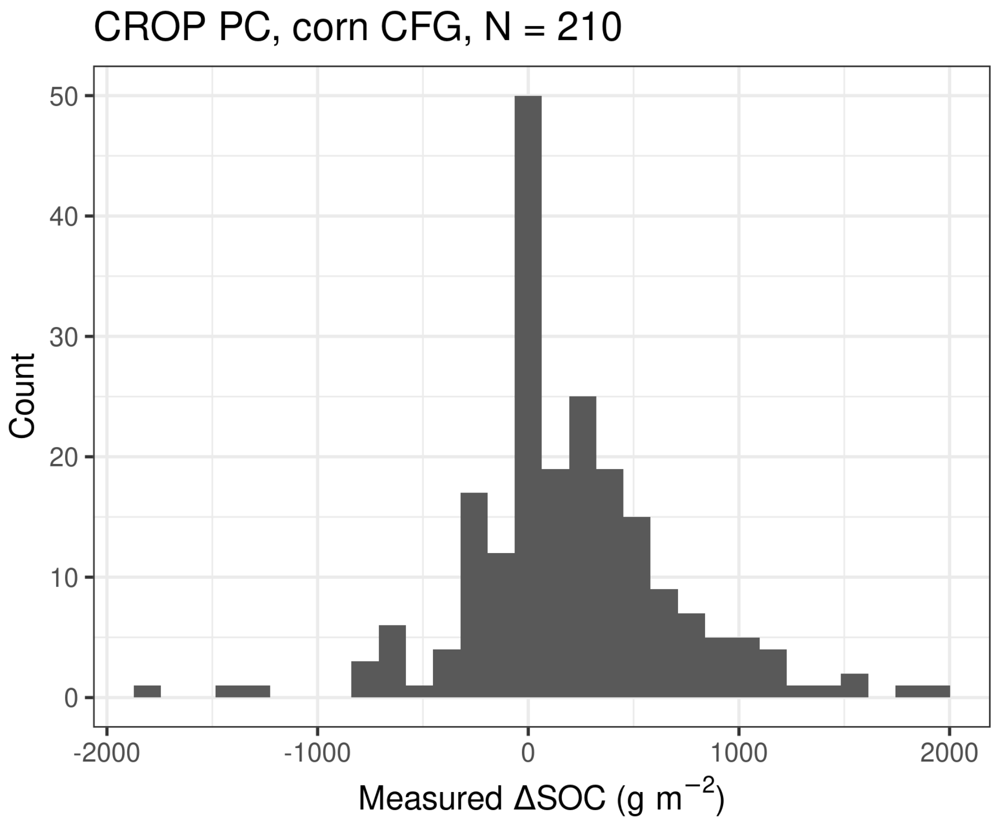

Figure 2: Histogram of changes in SOC observed by the studies used for model validation in response to changed cropping practices involving crops from the corn-type CFG.

Table 6: Descriptive dataset attributes for studies used in validation of CROP x corn.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
brookings_REAP	Wegner et al. (2018) and Osborne and Lehman (2018)	Brookings, SD	2008	2012	M	CTD	SiClLo	35	dry combustion	6
dalhart	Halvorson et al. (2009)	Dalhart, TX	1999	2006	H	WTD	SaLo	18	dry combustion	1
davis2	Clark et al. (1998)	Davis, CA	1988	1996	C	WTD	Lo	17	Walkley-Black method	5 (4 stack PCs)
fort_valley	Sainju, Whitehead, and Singh (2005)	Fort Valley, GA	1999	2002	P	WTM	SaLo	10	dry combustion	54
goias	Ferreira et al. (2019)	Goias, Brazil	2005	2014		TrM	Cl	50	dry combustion	3 (1 stack PCs)
hoytville	Collins et al. (1999)	Hoytville, OH	1963	1993	L	CTM	SiClLo	40	dry combustion	3
imperial	Denef et al. (2008)	Imperial, NE	1970	2012	H	CTD	Lo	24	dry combustion	3 (3 stack PCs)
kbs	Station (2021)	Hickory Corners KBS, MI	1993	2001	L	CTM	Lo	19	dry combustion	1 (1 stack PCs)
mead	Elliott et al. (1994)	Mead, NE	1975	1992	M	WTD	SiClLo	35	dry combustion	1
mead2	Varvel (2006)	Mead, NE	1982	1992, 1998, 2002	M	WTD	SiClLo	31	dry combustion	54

morrow	Khan et al. (2007)	Champaign-Urbana, IL	1955	2005	M	WTM	SiLo	25	dichromate oxidation technique of Mebius (1960)	4
otis	Denef et al. (2008)	Otis, CO	1966	2012	H	CTD	Lo	26	dry combustion	3 (3 stack PCs)
rodale	Elliott et al. (1994) and Pimentel et al. (2005)	Kutztown, PA	1981	1992, 2002	S	WTM	SiLo	30	not reported	6 (6 stack PCs)
russellranch_LTRAS	Kong et al. (2005)	Winter, CA	1993	1997, 2003, 2012	C	WTD	SiLo	18	dry combustion	6
saginaw	Christenson (1997)	Saginaw, MI	1972	1981, 1991	L	CTM	SiCl	47	dry combustion	30
tribune	Halvorson and Schlegel (2012)	Tribune, KS	2001	2010	H	WTD	SiLo	26	dry combustion	6
wooster	Collins et al. (1999) and Dick, Edwards, and McCoy (1997)	Wooster, OH	1962	1971, 1980, 1992	M	CTM	SiLo	15	dry combustion, Walkley-Black method	24

8.2 CROP x cotton x SOC

This category's validation is usable in all project climate zones and soil textures because:

- The observations within the US span only 2 LRRs (C, P), only one of which (P) is in the declared project domain, but the studies also include sites outside the US that are within the declared project climate zones (TrM). Collectively across US and international sites, the validation data are taken from four distinct agricultural regions (LRR C, LRR P, Brazil, Australia) across three climate zones (TrM, WTD, WTM), all of which are in the declared project domain. Following Model Requirements section 3.3, requirement 1 ("Datasets may be used from studies outside of the US. However, the associated IPCC climate zone where these datasets were collected should correspond to the declared IPCC climate zones of the project."), we interpret three project climate zones as equivalent to three project LRRs for purposes of meeting the bioclimatic distribution requirements using data from outside the US.

- 3 soil textures are included, all of which are in the declared project area: Cl, ClLo, SaLo.
- Clay content spans 54 percentage points, from 10% to 64%.
- At least one study isolates effects, i.e. only 1 of the 162 pairs of observations compare stacks of PC changes.

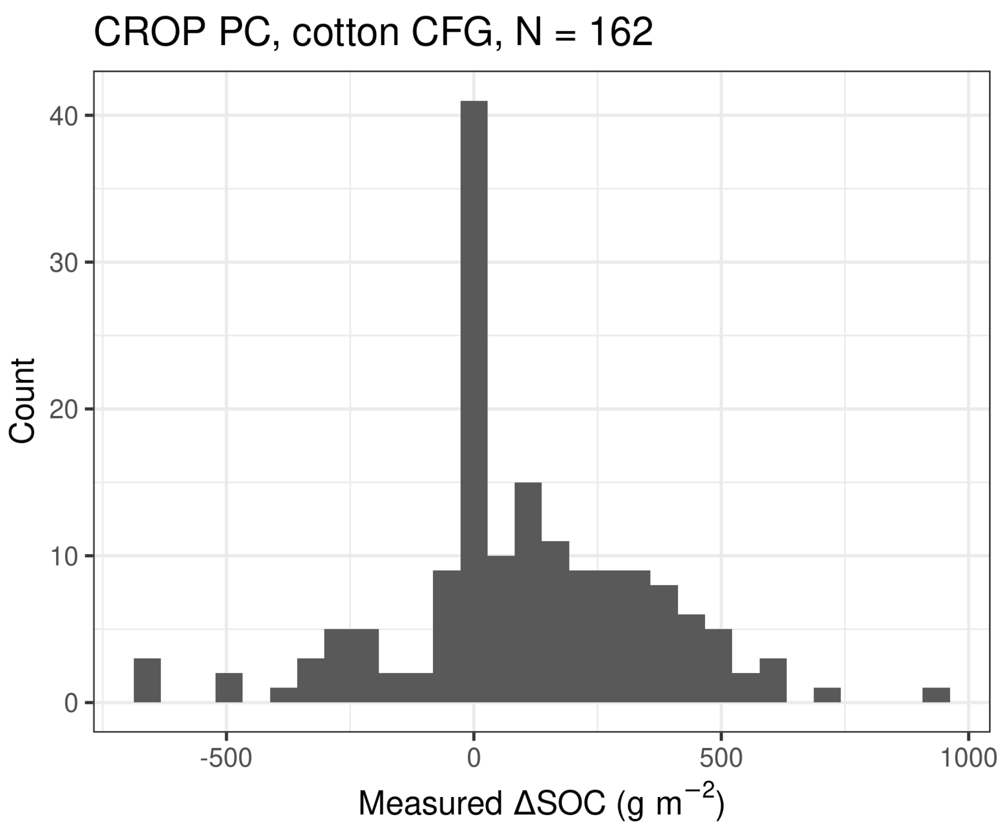

Figure 3: Histogram of changes in SOC observed by the studies used for model validation in response to changed cropping practices involving crops from the cotton-type CFG.

Table 7: Descriptive dataset attributes for studies used in validation of CROP x cotton.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
five_points	Mitchell et al. (2015), Mitchell et al. (2017), and Veenstra et al. (2006)	Five Points, CA	1999	2004, 2007, 2013	C	WTD	ClLo	39	dry combustion	6
fort_valley	Sainju, Whitehead, and Singh (2005)	Fort Valley, GA	1999	2002	P	WTM	SaLo	10	dry combustion	54
goias	Ferreira et al. (2019)	Goias, Brazil	2005	2014		TrM	Cl	50	dry combustion	4 (1 stack PCs)
narrabri_field6	Rochester (2011)	Narrabri, New South Wales	1995	2000, 2002, 2004, 2006, 2008		WTD	Cl	56	wet oxidation	50
narrabri_fieldC1	Senapati et al. (2014)	Narrabri, New South Wales	1985	1998, 2004, 2006, 2008, 2011, 2012		WTD	Cl	53	dry combustion	6

narrabri_fieldD1	Hulugalle et al. (2013)	Narrabri, New South Wales	2002	2005, 2006, 2007, 2008, 2009, 2010, 2011	WTD	Cl	64	wet oxidation	42

8.3 CROP x soy x SOC

This category's validation is usable in all project LRRs and soil textures because:

- The selected studies span 6 LRRs (C, H, L, M, P, S), 5 of which (H, L, M, P, S) are in the declared project domain, as well as sites outside the US that are within the declared project climate zones (CTD, TrM, WTD).
- 7 soil textures are included, all of which are in the declared project area: Cl, ClLo, Lo, SaLo, SiCl, SiClLo, SiLo.
- Clay content spans 54 percentage points, from 10% to 64%.
- At least one study isolates effects, i.e. only 18 of the 295 pairs of observations compare stacks of PC changes.

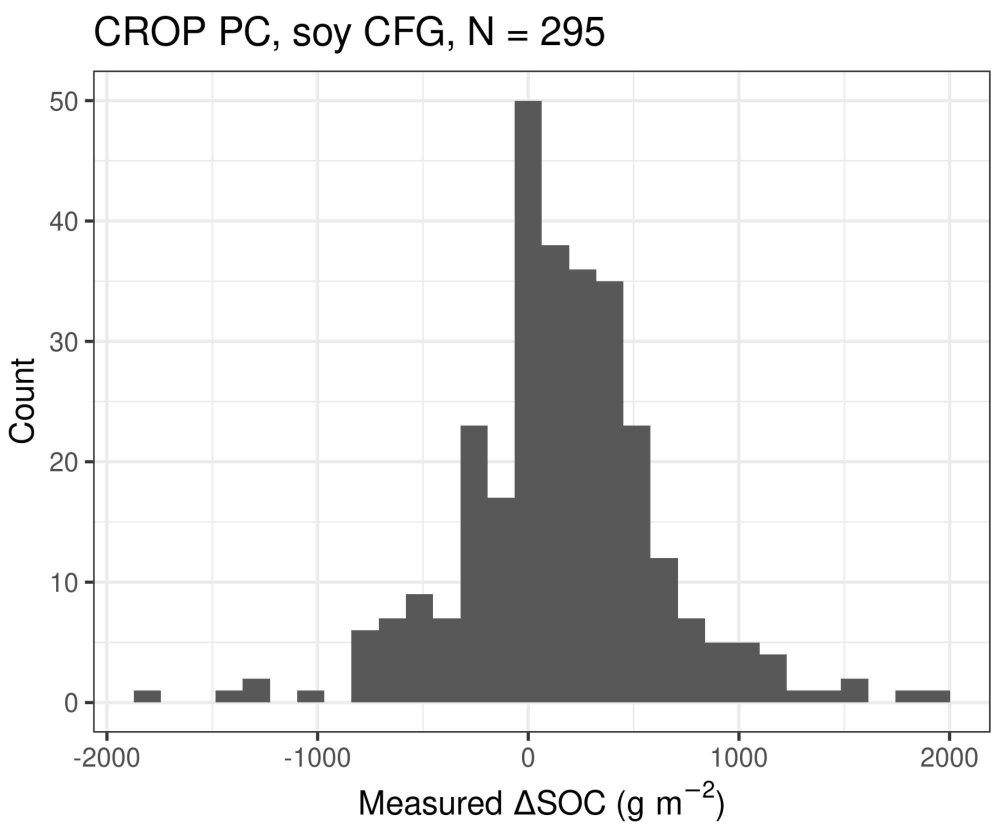

Figure 4: Histogram of changes in SOC observed by the studies used for model validation in response to changed cropping practices involving crops from the soy-type CFG.

Table 8: Descriptive dataset attributes for studies used in validation of CROP x soy.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
brookings_REAP	Wegner et al. (2018) and Osborne and Lehman (2018)	Brookings, SD	2008	2012	M	CTD	SiClLo	35	dry combustion	6
davis2	Clark et al. (1998)	Davis, CA	1988	1996	C	WTD	Lo	17	Walkley-Black method	5 (4 stack PCs)
five_points	Mitchell et al. (2015), Mitchell et al. (2017), and Veenstra et al. (2006)	Five Points, CA	1999	2004, 2007, 2013	C	WTD	ClLo	39	dry combustion	6
fort_valley	Sainju, Whitehead, and Singh (2005)	Fort Valley, GA	1999	2002	P	WTM	SaLo	10	dry combustion	45
goias	Ferreira et al. (2019)	Goias, Brazil	2005	2014		TrM	Cl	50	dry combustion	4 (1 stack PCs)
hoytville	Collins et al. (1999)	Hoytville, OH	1963	1993	L	CTM	SiClLo	40	dry combustion	3
imperial	Denef et al. (2008)	Imperial, NE	1970	2012	H	CTD	Lo	24	dry combustion	3 (3 stack PCs)

kbs	Station (2021)	Hickory Corners KBS, MI	1993	2001	L	CTM	Lo	19	dry combustion	1 (1 stack PCs)
mead	Elliott et al. (1994)	Mead, NE	1975	1992	M	WTD	SiClLo	35	dry combustion	1
mead2	Varvel (2006)	Mead, NE	1982	1992, 1998, 2002	M	WTD	SiClLo	31	dry combustion	45
morrow	Khan et al. (2007)	Champaign-Urbana, IL	1955	2005	M	WTM	SiLo	25	dichromate oxidation technique of Mebius (1960)	4
narrabri_field6	Rochester (2011)	Narrabri, New South Wales	1995	2000, 2002, 2004, 2006, 2008		WTD	Cl	56	wet oxidation	45
narrabri_fieldD1	Hulugalle et al. (2013)	Narrabri, New South Wales	2002	2005, 2006, 2007, 2008, 2009, 2010, 2011		WTD	Cl	64	wet oxidation	35
otis	Denef et al. (2008)	Otis, CO	1966	2012	H	CTD	Lo	26	dry combustion	3 (3 stack PCs)
rodale	Elliott et al. (1994) and Pimentel et al. (2005)	Kutztown, PA	1981	1992, 2002	S	WTM	SiLo	30	not reported	6 (6 stack PCs)

russellranch_LTRAS	Kong et al. (2005)	Winter, CA	1993	1997, 2003, 2012	C	WTD	SiLo	18	dry combustion	8
saginaw	Christenson (1997)	Saginaw, MI	1972	1981, 1991	L	CTM	SiCl	47	dry combustion	28
swiftcurrent	Campbell and Zentner (1997) and Campbell et al. (2007)	Swift Current, SK	1966	1981, 1984, 1990, 1993, 1996		CTD	SiLo	20	dry combustion	20
tribune	Halvorson and Schlegel (2012)	Tribune, KS	2001	2010	H	WTD	SiLo	26	dry combustion	3
wooster	Collins et al. (1999) and Dick, Edwards, and McCoy (1997)	Wooster, OH	1962	1971, 1980, 1992	M	CTM	SiLo	15	dry combustion, Walkley-Black method	24

8.4 CROP x wheat x SOC

This category's validation is usable in all project LRRs and soil textures because:

- The selected studies span 6 LRRs (C, H, L, M, P, S), 5 of which (H, L, M, P, S) are in the declared project domain, as well as sites outside the US that are within the declared project climate zones (CTD, WTD).
- 8 soil textures are included, all of which are in the declared project area: Cl, ClLo, Lo, SaClLo, SaLo, SiCl, SiClLo, SiLo.

- Clay content spans 54 percentage points, from 10% to 64%.
- At least one study isolates effects, i.e. only 17 of the 326 pairs of observations compare stacks of PC changes.

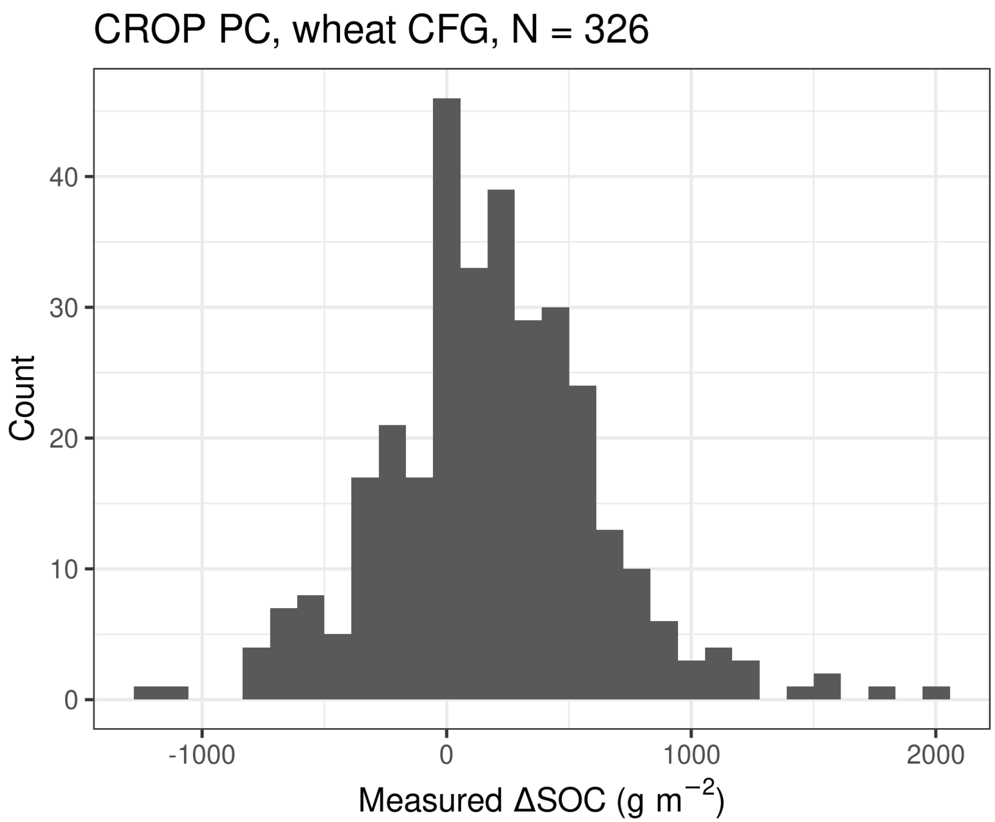

Figure 5: Histogram of changes in SOC observed by the studies used for model validation in response to changed cropping practices involving crops from the wheat-type CFG.

Table 9: Descriptive dataset attributes for studies used in validation of CROP x wheat.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
brookings_REAP	Wegner et al. (2018) and Osborne and Lehman (2018)	Brookings, SD	2008	2012	M	CTD	SiClLo	35	dry combustion	6
dalhart	Halvorson et al. (2009)	Dalhart, TX	1999	2006	H	WTD	SaLo	18	dry combustion	1
davis2	Clark et al. (1998)	Davis, CA	1988	1996	C	WTD	Lo	17	Walkley-Black method	5 (4 stack PCs)
five_points	Mitchell et al. (2015), Mitchell et al. (2017), and Veenstra et al. (2006)	Five Points, CA	1999	2004, 2007, 2013	C	WTD	ClLo	39	dry combustion	6
fort_valley	Sainju, Whitehead, and Singh (2005)	Fort Valley, GA	1999	2002	P	WTM	SaLo	10	dry combustion	45
hoytville	Collins et al. (1999)	Hoytville, OH	1963	1993	L	CTM	SiClLo	40	dry combustion	1
imperial	Denef et al. (2008)	Imperial, NE	1970	2012	H	CTD	Lo	24	dry combustion	3 (3 stack PCs)

kbs	Station (2021)	Hickory Corners KBS, MI	1993	2001	L	CTM	Lo	19	dry combustion	1 (1 stack PCs)
lethbridge2	Janzen et al. (1997)	Lethbridge, AB	1951	1967, 1974, 1985, 1992		CTD	Lo	25	dry combustion	24
lethbridgeABC	Monreal and Janzen (1993)	Lethbridge, AB	1910	1922, 1940, 1953, 1967, 1990		CTD	SaClLo	31	dry combustion (total C) and hot digestion with HCI (inorganic C) to determine organic C indirectly	10
mead	Elliott et al. (1994)	Mead, NE	1975	1992	M	WTD	SiClLo	35	dry combustion	1
mead2	Varvel (2006)	Mead, NE	1982	1992, 1998, 2002	M	WTD	SiClLo	31	dry combustion	18
morrow	Khan et al. (2007)	Champaign-Urbana, IL	1955	2005	M	WTM	SiLo	25	dichromate oxidation technique of Mebius (1960)	2
narrabri_field6	Rochester (2011)	Narrabri, New South Wales	1995	2000, 2002, 2004, 2006, 2008		WTD	Cl	56	wet oxidation	35

narrabri_fieldC1	Senapati et al. (2014)	Narrabri, New South Wales	1985	1998, 2004, 2006, 2008, 2011, 2012		WTD	Cl	53	dry combustion	6
narrabri_fieldD1	Hulugalle et al. (2013)	Narrabri, New South Wales	2002	2005, 2006, 2007, 2008, 2009, 2010, 2011		WTD	Cl	64	wet oxidation	35
otis	Denef et al. (2008)	Otis, CO	1966	2012	H	CTD	Lo	26	dry combustion	3 (3 stack PCs)
rodale	Elliott et al. (1994) and Pimentel et al. (2005)	Kutztown, PA	1981	1992, 2002	S	WTM	SiLo	30	not reported	6 (6 stack PCs)
russellranch_LTRAS	Kong et al. (2005)	Winter, CA	1993	1997, 2003, 2012	C	WTD	SiLo	18	dry combustion	11
saginaw	Christenson (1997)	Saginaw, MI	1972	1981, 1991	L	CTM	SiCl	47	dry combustion	30
swiftcurrent	Campbell and Zentner (1997) and Campbell et al. (2007)	Swift Current, SK	1966	1981, 1984, 1990, 1993, 1996		CTD	SiLo	20	dry combustion	55

tribune	Halvorson and Schlegel (2012)	Tribune, KS	2001	2010	H	WTD	SiLo	26	dry combustion	6
wooster	Collins et al. (1999) and Dick, Edwards, and McCoy (1997)	Wooster, OH	1962	1971, 1980, 1992	M	CTM	SiLo	15	dry combustion, Walkley-Black method	16

8.5 DISTURB x corn x SOC

This category's validation is usable in all project LRRs and soil textures because:

- The selected studies span 5 LRRs (K, L, M, N, P), all of which are in the declared project domain, as well as one site outside the US that is within the declared project climate zones (TrM).
- 5 soil textures are included, all of which are in the declared project area: Cl, Lo, SaLo, SiClLo, SiLo.
- Clay content spans 40 percentage points, from 10% to 50%.
- At least one study isolates effects, i.e. only 1 of the 225 pairs of observations compare stacks of PC changes.

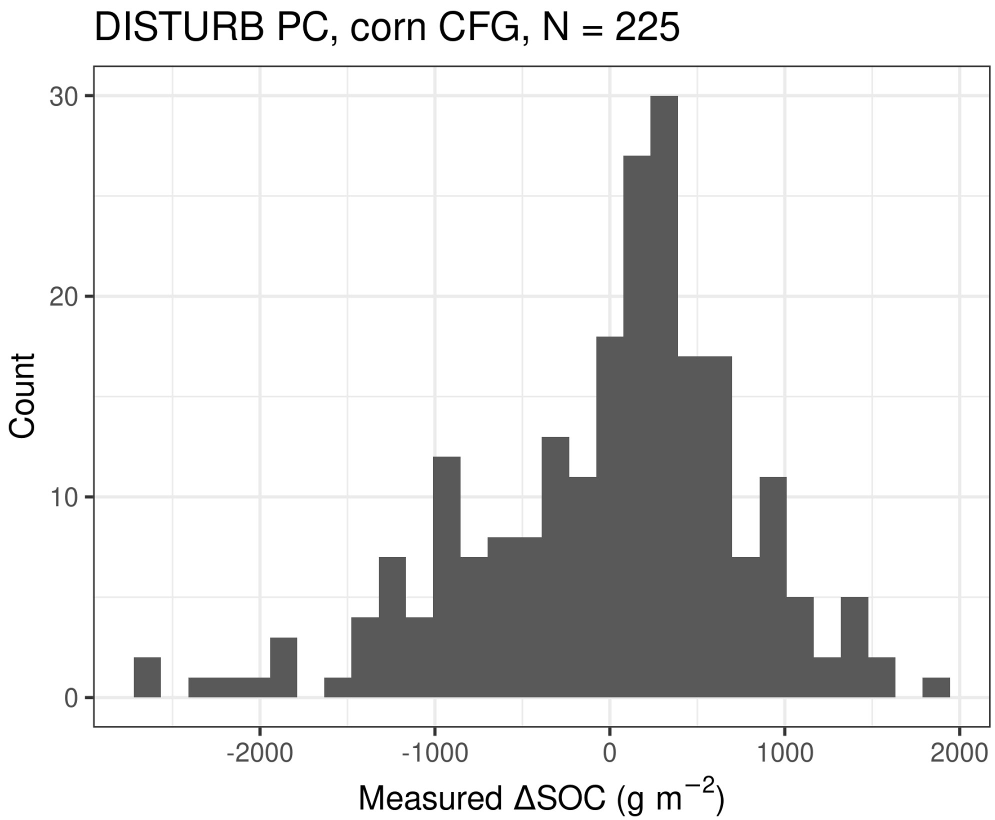

Figure 6: Histogram of changes in SOC observed by the studies used for model validation in response to changed tillage practices involving crops from the corn-type CFG.

Table 10: Descriptive dataset attributes for studies used in validation of DISTURB x corn.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
brookings_REAP	Wegner et al. (2018) and Osborne and Lehman (2018)	Brookings, SD	2008	2012	M	CTD	SiClLo	35	dry combustion	8
dixonsprings	Olson, Ebelhar, and Lang (2010)	Dixon Springs, IL	1989	1992, 2000, 2003, 2007, 2009	N	WTM	SiLo	19	Walkley-Black method	15
fort_valley	Sainju, Whitehead, and Singh (2005)	Fort Valley, GA	1999	2002	P	WTM	SaLo	10	dry combustion	36
goias	Ferreira et al. (2019)	Goias, Brazil	2005	2014		TrM	Cl	50	dry combustion	1 (1 stack PCs)
hoytville	Collins et al. (1999), Jarecki and Lal (2005), and Mestelan (2008)	Hoytville, OH	1963	1993, 2003, 2005	L	CTM	SiClLo	40	dry combustion	5

ithaca2	Jin and Varvel (2018b) and Jin et al. (2015)	Ithaca, NE	1998	2001, 2007, 2011	M	WTD	SiClLo	32	dry combustion	9
ithacaNE	Jin and Varvel (2018a) and Schmer et al. (2014)	Ithaca, NE	2001	2010, 2014	M	WTD	SiLo	26	dry combustion	18
kbs	Station (2021)	Hickory Corners KBS, MI	1993	2001	L	CTM	Lo	19	dry combustion	1
lafayette	Elliott et al. (1994)	Lafayette, IN	1975	1992	M	WTM	SiClLo	36	dry combustion	3
lexington	Blevins et al. (1983) and Ismail, Blevins, and Frye (1994)	Lexington, KY	1970	1980, 1989	N	WTM	SiLo	23	dry combustion, sulfuric acid-permanganate method of Allison (1965)	8
rosemount	Clapp et al. (2000) and Dolan et al. (2006)	Rosemount, MN	1980	1982, 1984, 1986, 1989, 1991, 1993, 2002	K	CTD	SiLo	24	dry combustion	93

scharleston	Collins et al. (1999) and Jarecki and Lal (2005)	South Charleston, OH	1962	1992, 2003	M	WTM	SiLo	20	dry combustion	6
wooster	Collins et al. (1999), Dick, Edwards, and McCoy (1997), and Mestelan (2008)	Wooster, OH	1962	1971, 1980, 1992, 2005	M	CTM	SiLo	15	dry combustion, Walkley-Black method	22

8.6 DISTURB x cotton x SOC

This category's validation is usable in all project climate zones and soil textures because:

- The observations within the US span only 2 LRRs (C, P), only one of which (P) is in the declared project domain, but the studies also include sites outside the US that are within the declared project climate zones (TrM, WTD). Collectively across US and international sites, the validation data are taken from four distinct agricultural regions (LRR C, LRR P, Brazil, Australia) across three climate zones (TrM, WTD, WTM), all of which are in the declared project domain. Following Model Requirements section 3.3, requirement 1 ("Datasets may be used from studies outside of the US. However, the associated IPCC climate zone where these datasets were collected should correspond to the declared IPCC climate zones of the project."), we interpret three project climate zones as equivalent to three project LRRs for purposes of meeting the bioclimatic distribution requirements using data from outside the US.

- 3 soil textures are included, all of which are in the declared project area: Cl, ClLo, SaLo.
- Clay content spans 43 percentage points, from 10% to 53%.
- At least one study isolates effects, i.e. only 1 of the 49 pairs of observations compare stacks of PC changes.

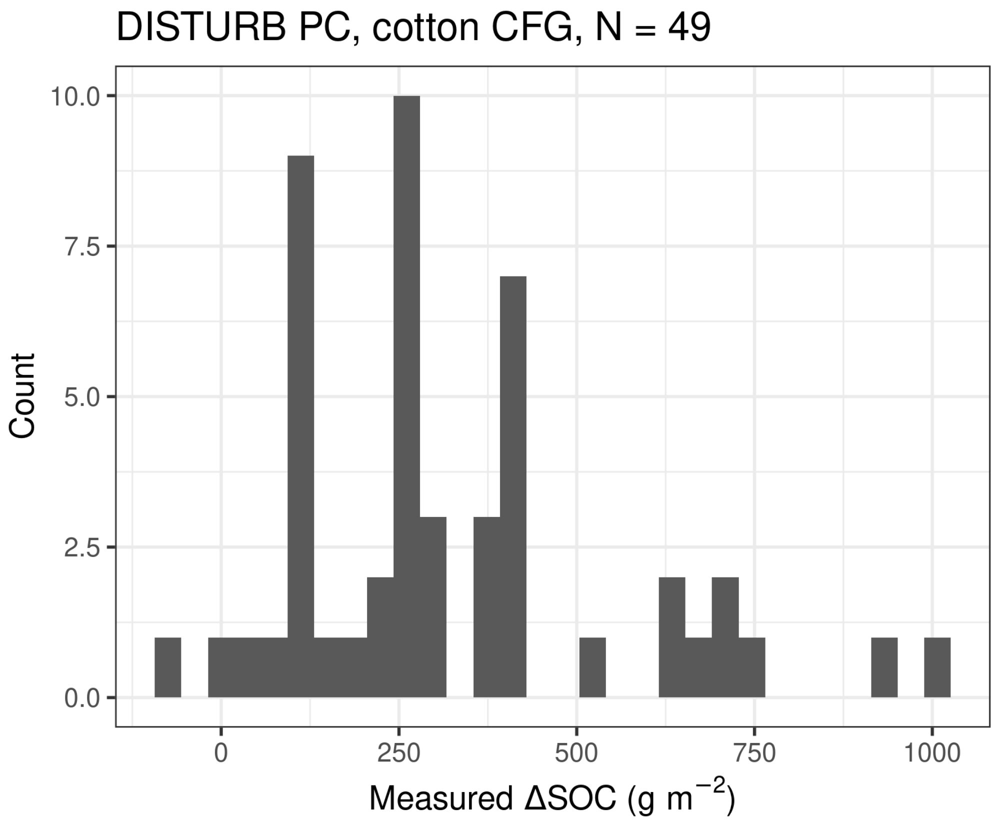

Figure 7: Histogram of changes in SOC observed by the studies used for model validation in response to changed tillage practices involving crops from the cotton-type CFG.

Table 11: Descriptive dataset attributes for studies used in validation of DISTURB x cotton.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
five_points	Mitchell et al. (2015), Mitchell et al. (2017), and Veenstra et al. (2006)	Five Points, CA	1999	2004, 2007, 2013	C	WTD	ClLo	39	dry combustion	6
fort_valley	Sainju, Whitehead, and Singh (2005)	Fort Valley, GA	1999	2002	P	WTM	SaLo	10	dry combustion	36
goias	Ferreira et al. (2019)	Goias, Brazil	2005	2014		TrM	Cl	50	dry combustion	1 (1 stack PCs)
narrabri_fieldC1	Senapati et al. (2014)	Narrabri, New South Wales	1985	1998, 2004, 2006, 2008, 2011, 2012		WTD	Cl	53	dry combustion	6

8.7 DISTURB x soy x SOC

This category's validation is usable in all project LRRs and soil textures because:

- The selected studies span 5 LRRs (C, L, M, N, P), 4 of which (L, M, N, P) are in the declared project domain, as well as one site outside the US that is within the declared project climate zones (TrM).

- 6 soil textures are included, all of which are in the declared project area: Cl, ClLo, Lo, SaLo, SiClLo, SiLo.
- Clay content spans 40 percentage points, from 10% to 50%.
- At least one study isolates effects, i.e. only 1 of the 66 pairs of observations compare stacks of PC changes.

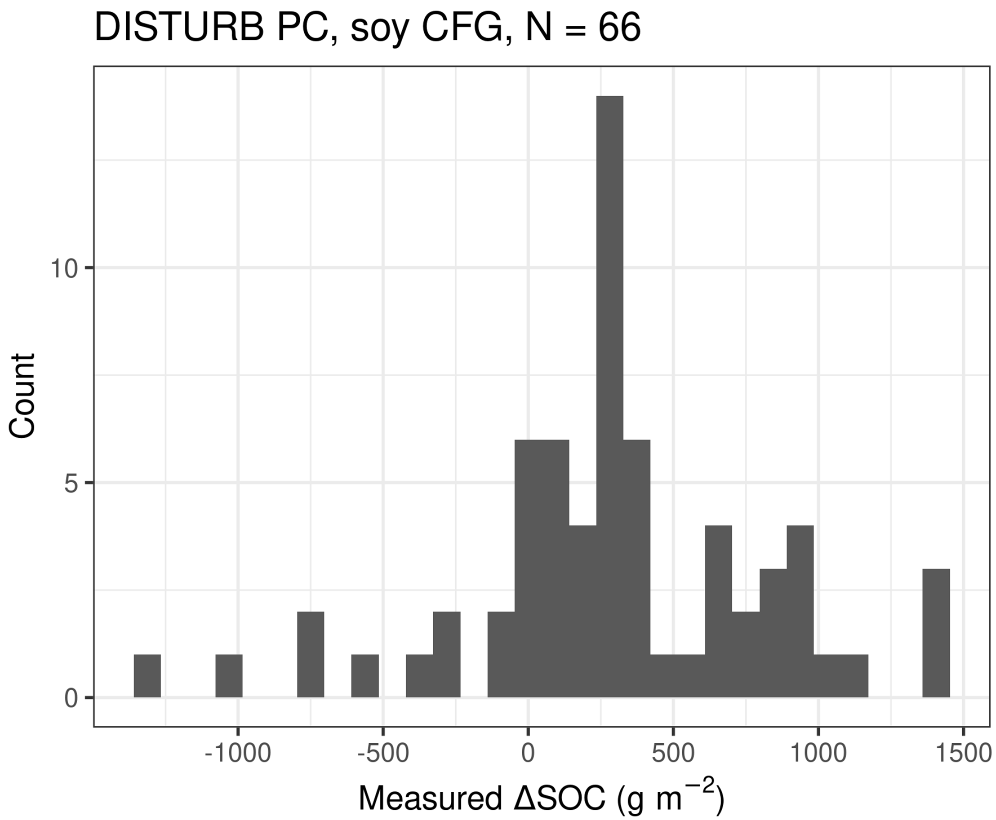

Figure 8: Histogram of changes in SOC observed by the studies used for model validation in response to changed tillage practices involving crops from the soy-type CFG.

Table 12: Descriptive dataset attributes for studies used in validation of DISTURB x soy.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
brookings_REAP	Wegner et al. (2018) and Osborne and Lehman (2018)	Brookings, SD	2008	2012	M	CTD	SiClLo	35	dry combustion	8
dixonsprings	Olson, Ebelhar, and Lang (2010)	Dixon Springs, IL	1989	1992, 2000, 2003, 2007, 2009	N	WTM	SiLo	19	Walkley-Black method	15
five_points	Mitchell et al. (2015), Mitchell et al. (2017), and Veenstra et al. (2006)	Five Points, CA	1999	2004, 2007, 2013	C	WTD	ClLo	39	dry combustion	3
fort_valley	Sainju, Whitehead, and Singh (2005)	Fort Valley, GA	1999	2002	P	WTM	SaLo	10	dry combustion	18
goias	Ferreira et al. (2019)	Goias, Brazil	2005	2014		TrM	Cl	50	dry combustion	1 (1 stack PCs)

hoytville	Collins et al. (1999) and Jarecki and Lal (2005)	Hoytville, OH	1963	1993, 2003	L	CTM	SiClLo	40	dry combustion	3
kbs	Station (2021)	Hickory Corners KBS, MI	1993	2001	L	CTM	Lo	19	dry combustion	1
lafayette	Elliott et al. (1994)	Lafayette, IN	1975	1992	M	WTM	SiClLo	36	dry combustion	3
wooster	Collins et al. (1999) and Dick, Edwards, and McCoy (1997)	Wooster, OH	1962	1971, 1980, 1992	M	CTM	SiLo	15	dry combustion, Walkley-Black method	14

8.8 DISTURB x wheat x SOC

This category's validation is usable in all project LRRs and soil textures because:

- The selected studies span 8 LRRs (B, C, F, G, H, L, M, P), 6 of which (F, G, H, L, M, P) are in the declared project domain.
- 5 soil textures are included, all of which are in the declared project area: ClLo, Lo, SaLo, SiClLo, SiLo.
- Clay content spans 30 percentage points, from 10% to 40%.
- At least one study isolates effects, i.e. 0 of the 87 pairs of observations compare stacks of PC changes.

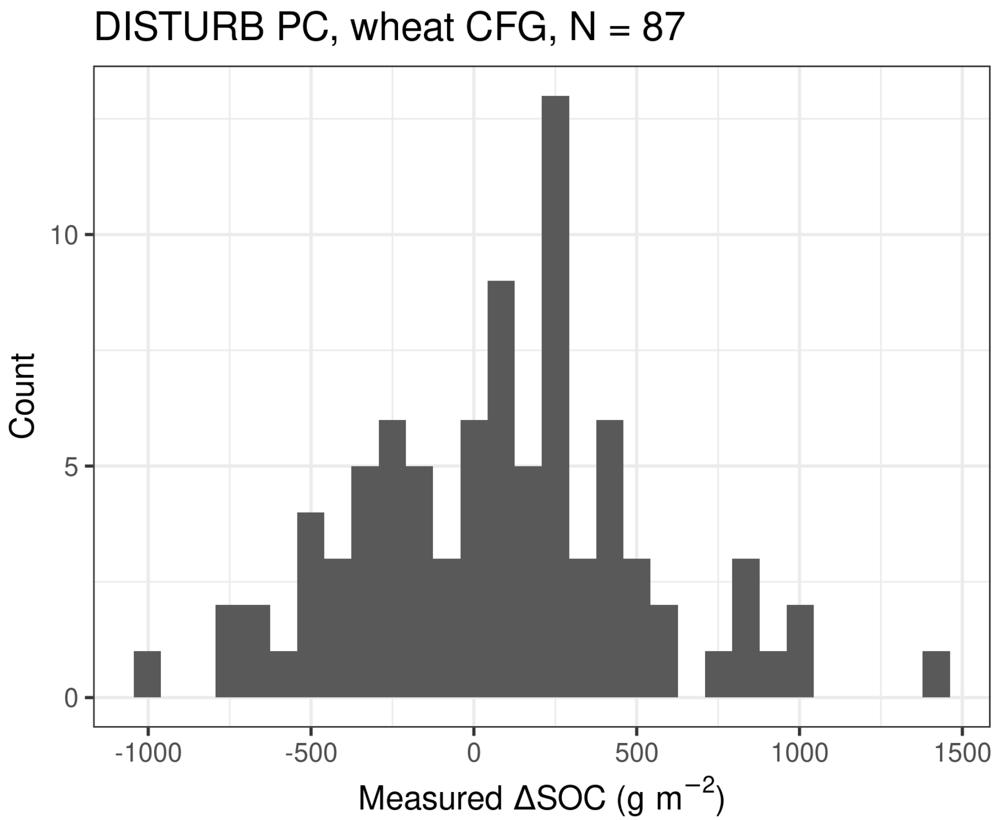

Figure 9: Histogram of changes in SOC observed by the studies used for model validation in response to changed tillage practices involving crops from the wheat-type CFG.

Table 13: Descriptive dataset attributes for studies used in validation of DISTURB x wheat.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
akron	Mikha, Vigil, and Benjamin (2013)	Akron, CO	1992	2006	G	CTD	SiLo	25	dry combustion	1
brookings_REAP	Wegner et al. (2018) and Osborne and Lehman (2018)	Brookings, SD	2008	2012	M	CTD	SiClLo	35	dry combustion	4
five_points	Mitchell et al. (2015), Mitchell et al. (2017), and Veenstra et al. (2006)	Five Points, CA	1999	2004, 2007, 2013	C	WTD	ClLo	39	dry combustion	6
fort_valley	Sainju, Whitehead, and Singh (2005)	Fort Valley, GA	1999	2002	P	WTM	SaLo	10	dry combustion	18
hoytville	Collins et al. (1999)	Hoytville, OH	1963	1993	L	CTM	SiClLo	40	dry combustion	1
kbs	Station (2021)	Hickory Corners KBS, MI	1993	2001	L	CTM	Lo	19	dry combustion	1

mandan_crop	Halvorson, Wienhold, and Black (2002)	Mandan, ND	1984	1990	F	CTD	SiLo	20	dry combustion	12
pendleton1	Bista et al. (2016), Ghimire, Machado, and Rhinhart (2015), and Rasmussen and Smiley (1997)	Pendleton, OR	1931	1941, 1951, 1976, 1986, 1995, 2005, 2010	B	WTD	SiLo	22	dry combustion, Walkley-Black	16
pendleton2	Ghimire, Machado, and Bista (2017)	Pendleton, OR	1940	1995	B	WTD	SiLo	24	dry combustion	18
sidney	Elliott et al. (1994)	Sidney, NE	1970	1993	H	CTD	Lo	25	dry combustion	3
wooster	Collins et al. (1999) and Dick, Edwards, and McCoy (1997)	Wooster, OH	1962	1971, 1980, 1992	M	CTM	SiLo	15	dry combustion, Walkley-Black method	7

8.9 NFERT x corn x SOC

This category's validation is usable in all project LRRs and soil textures because:

- The selected studies span 9 LRRs (C, E, H, K, L, M, N, P, S), 8 of which (E, H, K, L, M, N, P, S) are in the declared project domain.
- 6 soil textures are included, all of which are in the declared project area: ClLo, Lo, LoSa, SaLo, SiClLo, SiLo.
- Clay content spans 25 percentage points, from 10% to 35%.
- At least one study isolates effects, i.e. only 19 of the 166 pairs of observations compare stacks of PC changes.

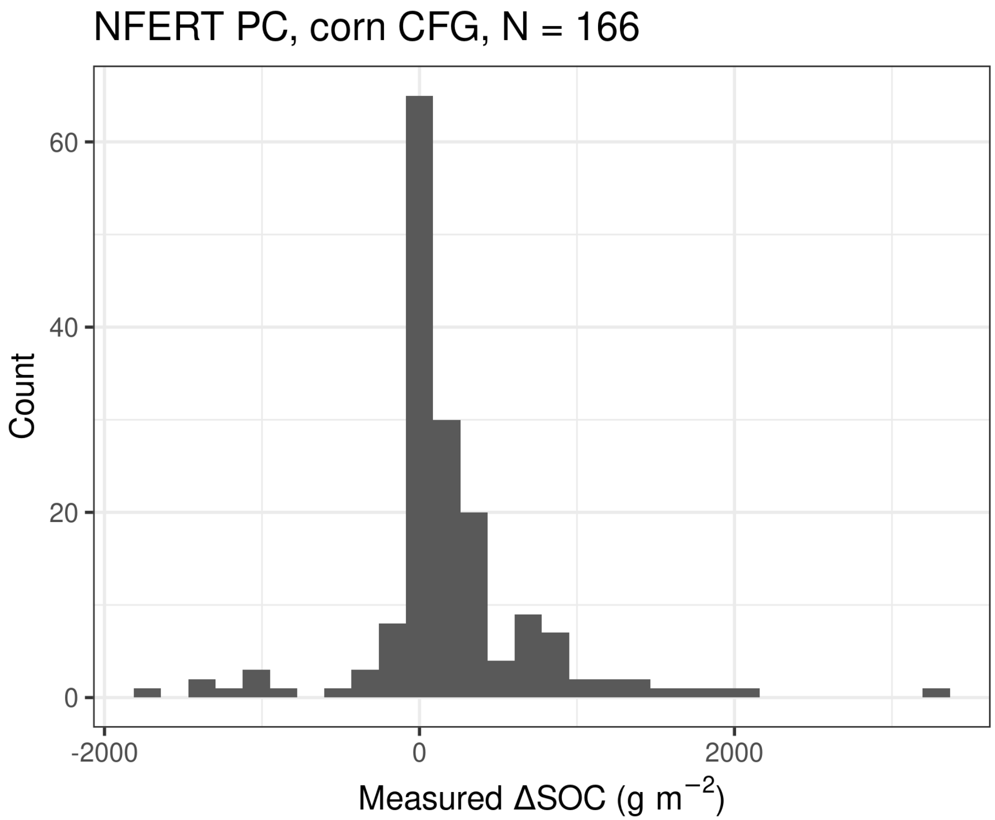

Figure 10: Histogram of changes in SOC observed by the studies used for model validation in response to changed nitrogen practices involving crops from the corn-type CFG.

Table 14: Descriptive dataset attributes for studies used in validation of NFERT x corn.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
ardec_manure	Halvorson, Stewart, and Del Grosso (2016)	Fort Collins, CO	2011	2014	E	CTD	ClLo	34	dry combustion	2 (1 stack PCs)
ardec1	Halvorson and Jantalia (2011)	Fort Collins, CO	1999	2009	E	CTD	ClLo	34	dry combustion	4
dalhart	Halvorson et al. (2009)	Dalhart, TX	1999	2002	H	WTD	SaLo	18	dry combustion	1
davis2	Clark et al. (1998)	Davis, CA	1988	1996	C	WTD	Lo	17	Walkley-Black method	4 (4 stack PCs)
elansing2	Vitosh, Davis, and Knezek (1973)	East Lansing, MI	1963	1971	L	CTM	LoSa	10	high frequency induction furnace	3 (3 stack PCs)
fort_valley	Sainju, Whitehead, and Singh (2005)	Fort Valley, GA	1999	2002	P	WTM	SaLo	10	dry combustion	36
ithaca2	Jin and Varvel (2018b) and Jin et al. (2015)	Ithaca, NE	1998	2001, 2007, 2011	M	WTD	SiClLo	32	dry combustion	18
kbs	Station (2021)	Hickory Corners KBS, MI	1993	2001	L	CTM	Lo	19	dry combustion	1 (1 stack PCs)

lexington	Blevins et al. (1983) and Ismail, Blevins, and Frye (1994)	Lexington, KY	1970	1980, 1989	N	WTM	SiLo	23	dry combustion, sulfuric acid-permanganate method of Allison (1965)	24
mead	Elliott et al. (1994)	Mead, NE	1975	1992	M	WTD	SiClLo	35	dry combustion	1 (1 stack PCs)
mead2	Varvel (2006)	Mead, NE	1982	1992, 1998, 2002	M	WTD	SiClLo	31	dry combustion	54
morrow	Khan et al. (2007)	Champaign-Urbana, IL	1955	2005	M	WTM	SiLo	25	dichromate oxidation technique of Mebius (1960)	3
rodale	Elliott et al. (1994) and Pimentel et al. (2005)	Kutztown, PA	1981	1992, 2002	S	WTM	SiLo	30	not reported	6 (6 stack PCs)
rosemount	Dolan et al. (2006)	Rosemount, MN	1980	2002	K	CTD	SiLo	24	dry combustion	6
russellranch_LTRAS	Kong et al. (2005)	Winter, CA	1993	1997, 2003, 2012	C	WTD	SiLo	18	dry combustion	3 (3 stack PCs)

8.10 NFERT x soy x SOC

This category's validation is usable in all project LRRs and soil textures because:

- The selected studies span 5 LRRs (C, L, M, P, S), 4 of which (L, M, P, S) are in the declared project domain.

- 4 soil textures are included, all of which are in the declared project area: Lo, SaLo, SiClLo, SiLo.
- Clay content spans 25 percentage points, from 10% to 35%.
- At least one study isolates effects, i.e. only 12 of the 77 pairs of observations compare stacks of PC changes.

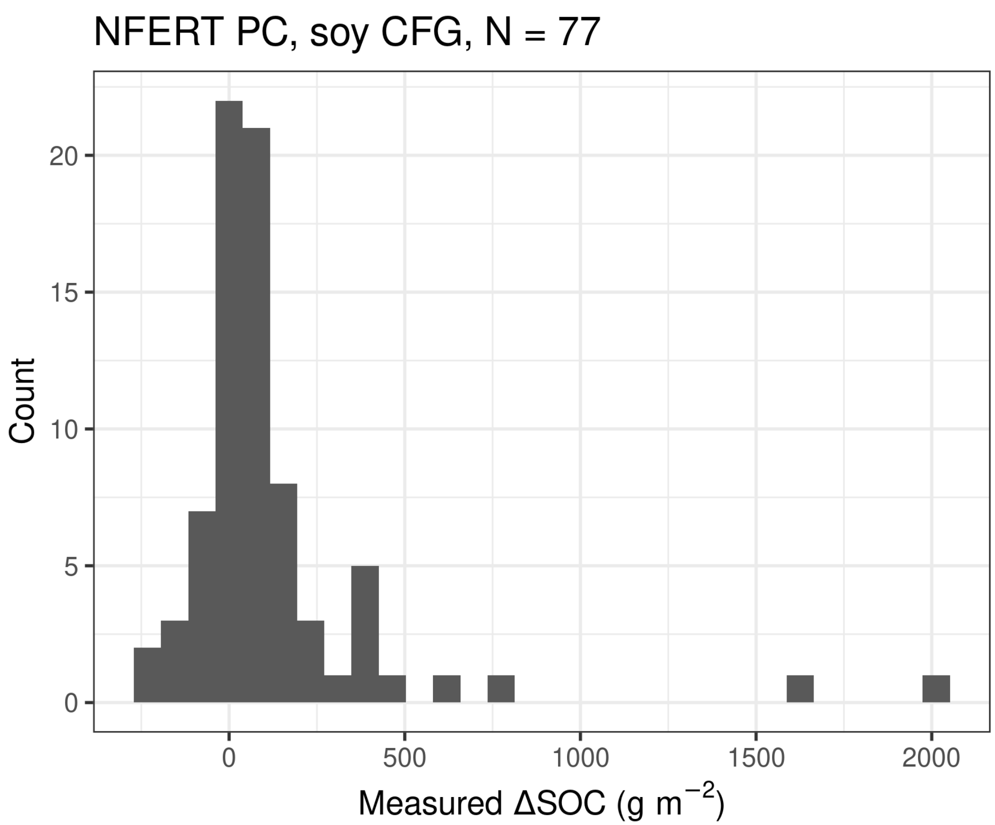

Figure 11: Histogram of changes in SOC observed by the studies used for model validation in response to changed nitrogen practices involving crops from the soy-type CFG.

Table 15: Descriptive dataset attributes for studies used in validation of NFERT x soy.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
davis2	Clark et al. (1998)	Davis, CA	1988	1996	C	WTD	Lo	17	Walkley-Black method	4 (4 stack PCs)
fort_valley	Sainju, Whitehead, and Singh (2005)	Fort Valley, GA	1999	2002	P	WTM	SaLo	10	dry combustion	18
kbs	Station (2021)	Hickory Corners KBS, MI	1993	2001	L	CTM	Lo	19	dry combustion	1 (1 stack PCs)
mead	Elliott et al. (1994)	Mead, NE	1975	1992	M	WTD	SiClLo	35	dry combustion	1 (1 stack PCs)
mead2	Varvel (2006)	Mead, NE	1982	1992, 1998, 2002	M	WTD	SiClLo	31	dry combustion	45
morrow	Khan et al. (2007)	Champaign-Urbana, IL	1955	2005	M	WTM	SiLo	25	dichromate oxidation technique of Mebius (1960)	2
rodale	Elliott et al. (1994) and Pimentel et al. (2005)	Kutztown, PA	1981	1992, 2002	S	WTM	SiLo	30	not reported	6 (6 stack PCs)

8.11 NFERT x wheat x SOC

This category's validation is usable in all project LRRs and soil textures because:

- The selected studies span 8 LRRs (B, C, F, H, L, M, P, S), 6 of which (F, H, L, M, P, S) are in the declared project domain, as well as sites outside the US that are within the declared project climate zones (CTM, CTD).
- 4 soil textures are included, all of which are in the declared project area: Lo, SaLo, SiClLo, SiLo.
- Clay content spans 25 percentage points, from 10% to 35%.
- At least one study isolates effects, i.e. only 21 of the 173 pairs of observations compare stacks of PC changes.

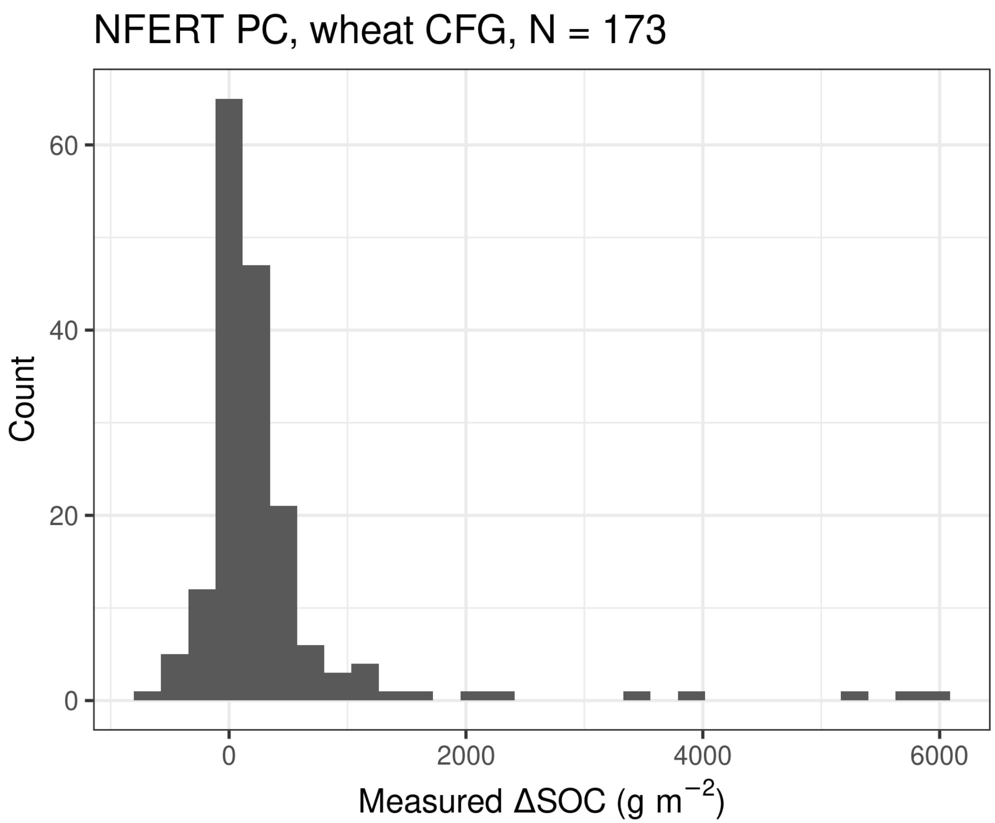

Figure 12: Histogram of changes in SOC observed by the studies used for model validation in response to changed nitrogen practices involving crops from the wheat-type CFG.

Table 16: Descriptive dataset attributes for studies used in validation of NFERT x wheat.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
broadbalk	Research (2014)	Rothamsted, England	1844	1893, 1914, 1944, 1992, 1997, 2005		CTM	SiClLo	25	dry combustion	12 (6 stack PCs)
dalhart	Halvorson et al. (2009)	Dalhart, TX	1999	2002	H	WTD	SaLo	18	dry combustion	1
davis2	Clark et al. (1998)	Davis, CA	1988	1996	C	WTD	Lo	17	Walkley-Black method	4 (4 stack PCs)
fort_valley	Sainju, Whitehead, and Singh (2005)	Fort Valley, GA	1999	2002	P	WTM	SaLo	10	dry combustion	18
kbs	Station (2021)	Hickory Corners KBS, MI	1993	2001	L	CTM	Lo	19	dry combustion	1 (1 stack PCs)
mandan_crop	Halvorson, Wienhold, and Black (2002)	Mandan, ND	1984	1990	F	CTD	SiLo	20	dry combustion	18
mead	Elliott et al. (1994)	Mead, NE	1975	1992	M	WTD	SiClLo	35	dry combustion	1 (1 stack PCs)
mead2	Varvel (2006)	Mead, NE	1982	1992, 1998, 2002	M	WTD	SiClLo	31	dry combustion	18

morrow	Khan et al. (2007)	Champaign-Urbana, IL	1955	2005	M	WTM	SiLo	25	dichromate oxidation technique of Mebius (1960)	1
pendleton1	Bista et al. (2016), Ghimire, Machado, and Rhinhart (2015), and Rasmussen and Smiley (1997)	Pendleton, OR	1931	1941, 1951, 1976, 1986, 1995, 2005, 2010	B	WTD	SiLo	22	dry combustion, Walkley-Black	24
pendleton2	Ghimire, Machado, and Bista (2017)	Pendleton, OR	1940	1995	B	WTD	SiLo	24	dry combustion	45
rodale	Elliott et al. (1994) and Pimentel et al. (2005)	Kutztown, PA	1981	1992, 2002	S	WTM	SiLo	30	not reported	6 (6 stack PCs)
russellranch_LTRAS	Kong et al. (2005)	Winter, CA	1993	1997, 2003, 2012	C	WTD	SiLo	18	dry combustion	9 (3 stack PCs)
swiftcurrent	Campbell and Zentner (1997) and Campbell et al. (2007)	Swift Current, SK	1966	1981, 1984, 1990, 1993, 1996		CTD	SiLo	20	dry combustion	15

8.12 ORG x All x SOC

Follows Model Requirements section 3.3.1, paragraph 5

This category's validation is usable in all project LRRs and soil textures because, when considering observations of the ORG PC across all annual CFGS:

- The selected studies span 6 LRRs (B, C, E, L, M, S), 4 of which (E, L, M, S) are in the declared project domain, as well as sites outside the US that are within the declared project climate zones (CTM, CTD).
- 5 soil textures are included, all of which are in the declared project area: ClLo, Lo, LoSa, SiClLo, SiLo.
- Clay content spans 29 percentage points, from 10% to 39%.
- At least one study isolates effects, i.e. only 20 of the 58 pairs of observations compare stacks of PC changes.

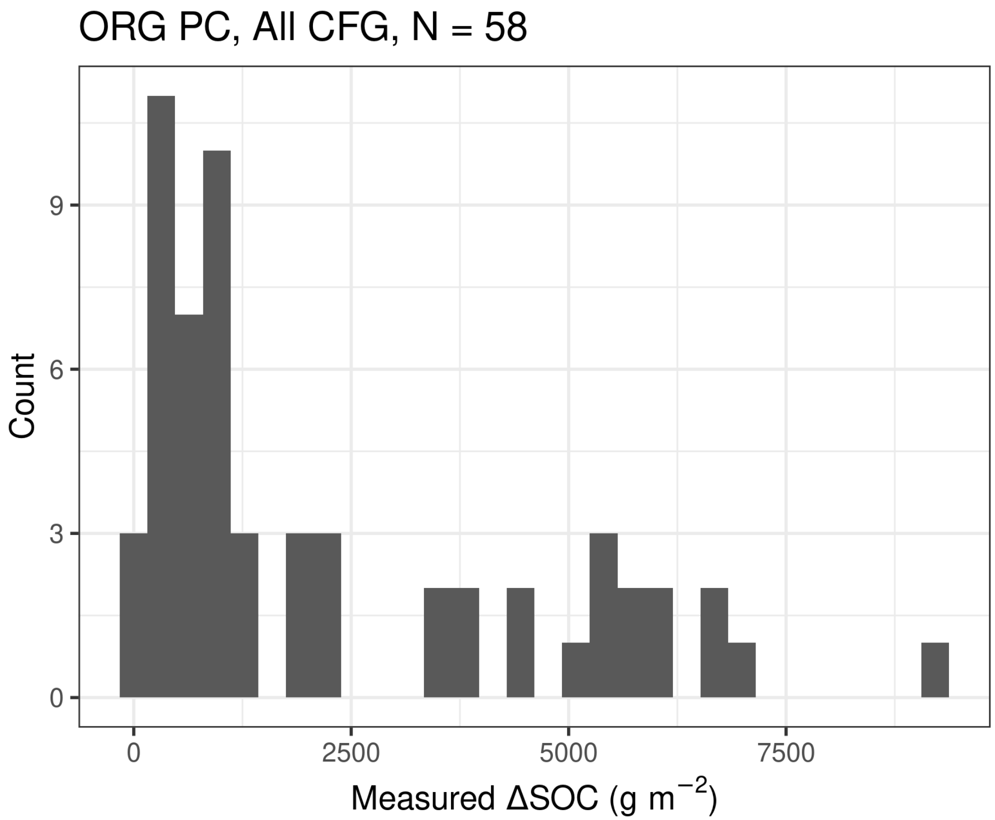

Figure 13: Histogram of changes in SOC observed by the studies used for model validation in response to changed organic amendment practices involving crops from the All-type CFG.

Table 17: Descriptive dataset attributes for studies used in validation of ORG x All.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
ardec_manure	Halvorson, Stewart, and Del Grosso (2016)	Fort Collins, CO	2011	2014	E	CTD	ClLo	34	dry combustion	2 (1 stack PCs)
broadbalk	Research (2014)	Rothamsted, England	1844	1893, 1914, 1944, 1992, 1997, 2005		CTM	SiClLo	25	dry combustion	12 (6 stack PCs)
davis2	Clark et al. (1998)	Davis, CA	1988	1996	C	WTD	Lo	17	Walkley-Black method	2 (2 stack PCs)
elansing2	Vitosh, Davis, and Knezek (1973)	East Lansing, MI	1963	1971	L	CTM	LoSa	10	high frequency induction furnace	3 (3 stack PCs)
lethbridge_manure	Hao et al. (2003)	Lethbridge, AB	1973	1998		CTD	ClLo	39	dry combustion	6
lethbridge2	Janzen et al. (1997)	Lethbridge, AB	1951	1967, 1974, 1985, 1992		CTD	Lo	25	dry combustion	4
mead	Elliott et al. (1994)	Mead, NE	1975	1992	M	WTD	SiClLo	35	dry combustion	1 (1 stack PCs)

pendleton1	Bista et al. (2016), Ghimire, Machado, and Rhinhart (2015), and Rasmussen and Smiley (1997)	Pendleton, OR	1931	1941, 1951, 1976, 1986, 1995, 2005, 2010	B	WTD	SiLo	22	dry combustion, Walkley-Black	21
rodale	Elliott et al. (1994) and Pimentel et al. (2005)	Kutztown, PA	1981	1992, 2002	S	WTM	SiLo	30	not reported	4 (4 stack PCs)
russellranch_LTRAS	Kong et al. (2005)	Winter, CA	1993	1997, 2003, 2012	C	WTD	SiLo	18	dry combustion	3 (3 stack PCs)

8.13 ORG x corn x SOC

Follows Model Requirements section 3.3.1, paragraph 5

This category is presented for context only to support the use of its data as part of 8.12 "ORG x All x SOC", but the validation would be usable in all project LRRs and soil textures because:

- The selected studies span 5 LRRs (C, E, L, M, S), 4 of which (E, L, M, S) are in the declared project domain.
- 5 soil textures are included, all of which are in the declared project area: ClLo, Lo, LoSa, SiClLo, SiLo.

- Clay content spans 25 percentage points, from 10% to 35%.
- At least one study isolates effects, i.e. only 14 of the 15 pairs of observations compare stacks of PC changes.

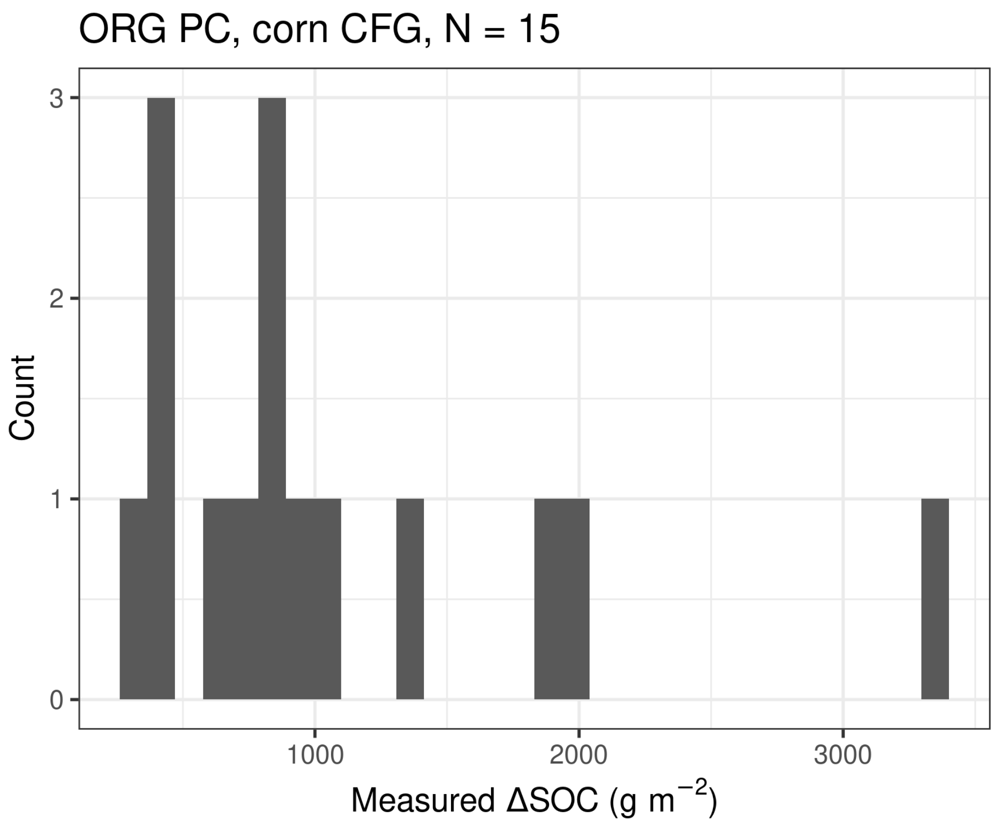

Figure 14: Histogram of changes in SOC observed by the studies used for model validation in response to changed organic amendment practices involving crops from the corn-type CFG.

Table 18: Descriptive dataset attributes for studies used in validation of ORG x corn.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
ardec_manure	Halvorson, Stewart, and Del Grosso (2016)	Fort Collins, CO	2011	2014	E	CTD	ClLo	34	dry combustion	2 (1 stack PCs)
davis2	Clark et al. (1998)	Davis, CA	1988	1996	C	WTD	Lo	17	Walkley-Black method	2 (2 stack PCs)
elansing2	Vitosh, Davis, and Knezek (1973)	East Lansing, MI	1963	1971	L	CTM	LoSa	10	high frequency induction furnace	3 (3 stack PCs)
mead	Elliott et al. (1994)	Mead, NE	1975	1992	M	WTD	SiClLo	35	dry combustion	1 (1 stack PCs)
rodale	Elliott et al. (1994) and Pimentel et al. (2005)	Kutztown, PA	1981	1992, 2002	S	WTM	SiLo	30	not reported	4 (4 stack PCs)
russellranch_LTRAS	Kong et al. (2005)	Winter, CA	1993	1997, 2003, 2012	C	WTD	SiLo	18	dry combustion	3 (3 stack PCs)

8.14 ORG x soy x SOC

Follows Model Requirements section 3.3.1, paragraph 5

This category is presented for context only to support the use of its data as part of 8.12 “ORG x All x SOC”, and has insufficient data to validate for the entire project:

- The selected studies span 3 LRRs (C, M, S), only 2 of which (M, S) are in the declared project domain. Validating this category for all the geographies in project CAR1459 would require data from at least 3 project LRRS (Model Requirements section 3.3 Requirement 2), so we present the data only for context to support the validation of 8.12 “ORG x All x SOC”
- 3 soil textures are included, all of which are in the declared project area: Lo, SiClLo, SiLo.
- Clay content spans 18 percentage points, from 17% to 35%.
- No study isolates effects, i.e. all 7 of the 7 pairs of observations compare stacks of PC changes. Therefore the Model Requirement that a validation dataset not be made up *exclusively* of stacked-practice observations is not met.

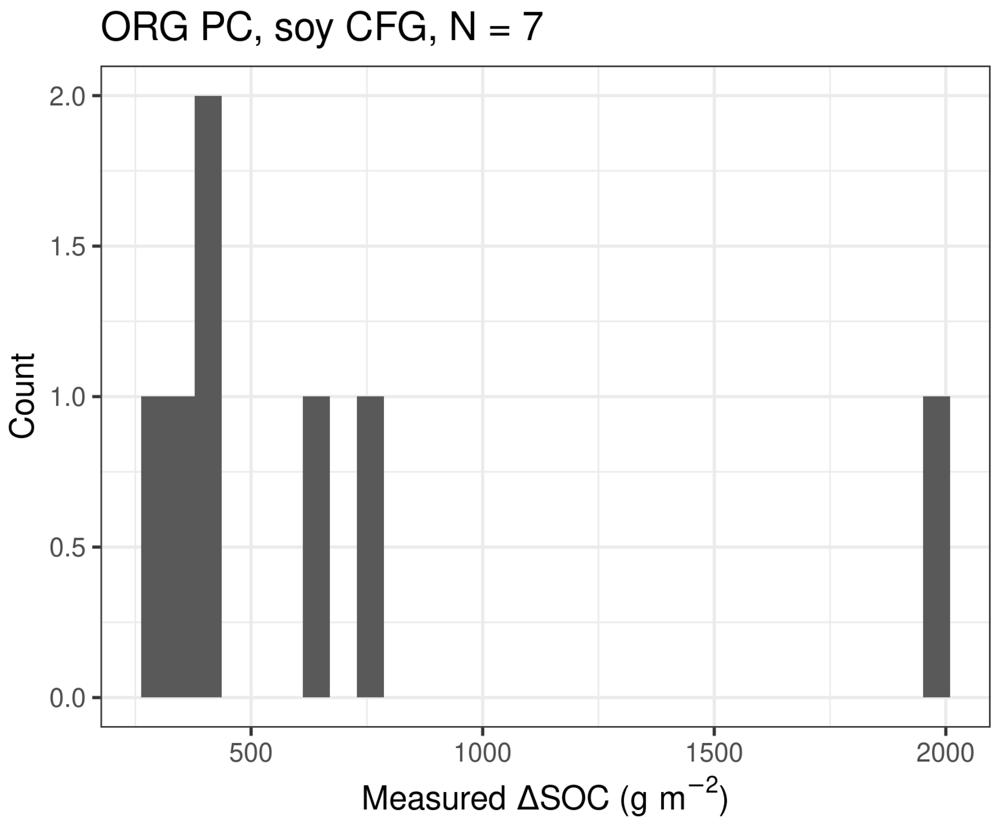

Figure 15: Histogram of changes in SOC observed by the studies used for model validation in response to changed organic amendment practices involving crops from the soy-type CFG.

Table 19: Descriptive dataset attributes for studies used in validation of ORG x soy.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
davis2	Clark et al. (1998)	Davis, CA	1988	1996	C	WTD	Lo	17	Walkley-Black method	2 (2 stack PCs)
mead	Elliott et al. (1994)	Mead, NE	1975	1992	M	WTD	SiClLo	35	dry combustion	1 (1 stack PCs)
rodale	Elliott et al. (1994) and Pimentel et al. (2005)	Kutztown, PA	1981	1992, 2002	S	WTM	SiLo	30	not reported	4 (4 stack PCs)

8.15 ORG x wheat x SOC

This category's validation is usable in all project LRRs and soil textures because:

- The selected studies span 4 LRRs (B, C, M, S), only two of which (M, S) are in the declared project domain, but the studies also include sites outside the US that are within the declared project climate zones (CTM, CTD). Collectively across US and international sites, the validation data are taken from six distinct agricultural regions (LRR B, LRR C, LRR M, LRR S, Canada, England) across four climate zones (CTD, CTM, WTD, WTM), all of which are in the declared project domain. Following Model Requirements section 3.3, requirement 1 ("Datasets may be used from studies outside of the US. However, the associated IPCC climate zone where these datasets were collected should correspond to the declared IPCC climate zones of the project."), we interpret four project climate zones as equivalent to four project LRRs for purposes of meeting the bioclimatic distribution requirements using data from outside the US.
- 4 soil textures are included, all of which are in the declared project area: ClLo, Lo, SiClLo, SiLo.
- Clay content spans 22 percentage points, from 17% to 39%.

- At least one study isolates effects, i.e. only 16 of the 53 pairs of observations compare stacks of PC changes.

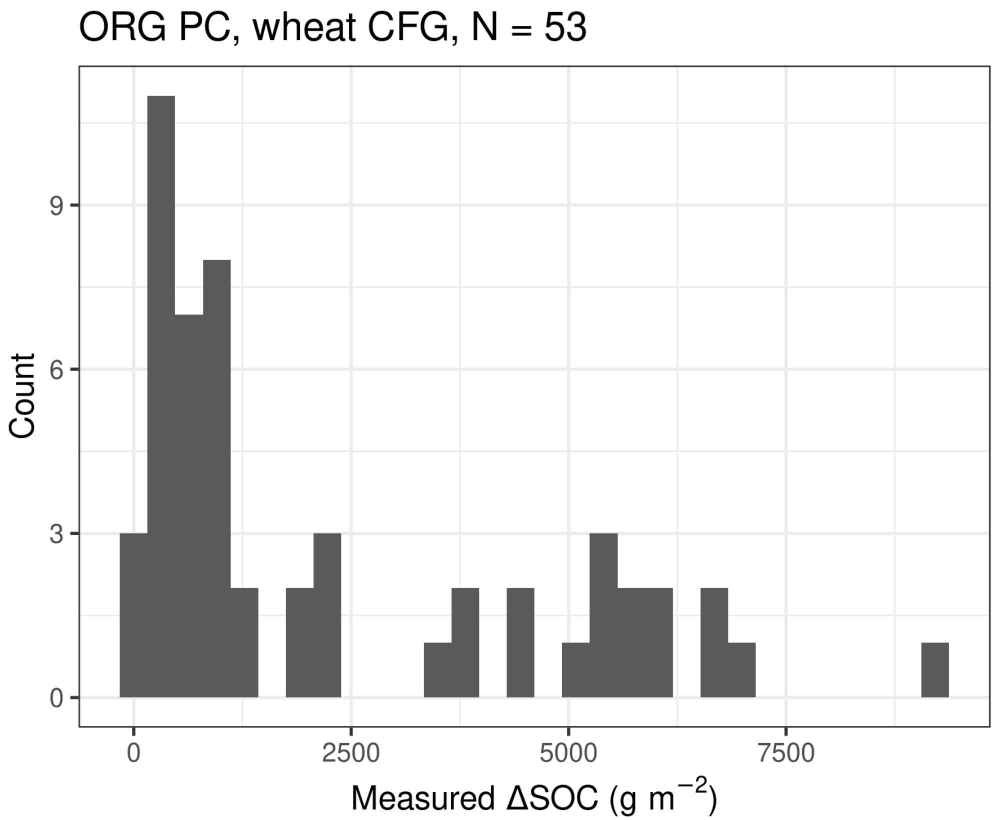

Figure 16: Histogram of changes in SOC observed by the studies used for model validation in response to changed organic amendment practices involving crops from the wheat-type CFG.

Table 20: Descriptive dataset attributes for studies used in validation of ORG x wheat.

Study name	Citation(s)	Location	Year initiated	Year(s) measured	LRR	IPCC climate zone	Soil texture	Clay content (%)	SOC measurement method(s)	N observations
broadbalk	Research (2014)	Rothamsted, England	1844	1893, 1914, 1944, 1992, 1997, 2005		CTM	SiClLo	25	dry combustion	12 (6 stack PCs)
davis2	Clark et al. (1998)	Davis, CA	1988	1996	C	WTD	Lo	17	Walkley-Black method	2 (2 stack PCs)
lethbridge_manure	Hao et al. (2003)	Lethbridge, AB	1973	1998		CTD	ClLo	39	dry combustion	6
lethbridge2	Janzen et al. (1997)	Lethbridge, AB	1951	1967, 1974, 1985, 1992		CTD	Lo	25	dry combustion	4
mead	Elliott et al. (1994)	Mead, NE	1975	1992	M	WTD	SiClLo	35	dry combustion	1 (1 stack PCs)
pendleton1	Bista et al. (2016), Ghimire, Machado, and Rhinhart (2015), and Rasmussen and Smiley (1997)	Pendleton, OR	1931	1941, 1951, 1976, 1986, 1995, 2005, 2010	B	WTD	SiLo	22	dry combustion, Walkley-Black	21

rodale	Elliott et al. (1994) and Pimentel et al. (2005)	Kutztown, PA	1981	1992, 2002	S	WTM	SiLo	30	not reported	4 (4 stack PCs)
russellranch_LTRAS	Kong et al. (2005)	Winter, CA	1993	1997, 2003, 2012	C	WTD	SiLo	18	dry combustion	3 (3 stack PCs)

9 Bias evaluation

Follows Model Requirements Section 3.4 Summary of Requirements (p18)

9.1 Calculating Bias

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 Eq. (2):

$$\text{bias} = \frac{1}{n}\sum_{i=1}^{n} \text{modeled}_i - \text{observed}_i \tag{2}$$

where observed_i is the observed difference in SOC at the second time point for the i^{th} experimental treatment pair (e.g. $\text{SOC}_{\text{no-till}} - \text{SOC}_{\text{till}}$), modeled_i is the modeled difference in SOC at the second time point for the i^{th} experimental treatment pair, 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 crop categories, only the observations matching the category of interest were included in the calculation.

Note that while Eq. (2) calculates bias at the second time point, it is identical to calculating bias in emission reductions between the first and second time point. This is because measured and observed SOC at the first time point are always identical (the SEP requires that modeled SOC be constrained to equal observed SOC at the first time point), so these values cancel out when subtracting observed_i from modeled_i.

Bias for each category was then computed as the 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. Per Section 1.6 "Changes from previous validation report" and Appendix G, the PMU was calculated using Eq. (3):

$$\text{PMU} = \sqrt{\frac{\sum_{i=1}^{k} \sigma_i^2 (n_{i1} + n_{i2} - 2)}{\sum_{i=1}^{k} (n_{i1} + n_{i2} - 2)}} \tag{3}$$

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, n_{i1} and n_{i2} are the number of replicates included in first and second study in the i^{th} treatment pair, and $n_{i1} + n_{i2} - 2$ is the degrees of freedom of σ_i^2.

9.2 Example PMU calculation

The CROP $\times$ Corn validation dataset contained $k = 15$ observations of practice changes (Table 21) that reported uncertainties for both observed treatments and therefore allow computing the standard error of the difference between treatments as $\sqrt{\sigma_1^2 + \sigma_2^2}$.

Table 21: Pairs of observations from the CROP x corn validation dataset for which estimates of measurement uncertainty were available, showing calculation of standard error of difference to be used for calculating PMU of SOC change.

site	n trt1	n trt2	se trt1	se trt2	se diff	df se diff
hoytville	3	3	670	419	790.2	4
hoytville	3	3	419	37	420.6	4
hoytville	3	3	670	509	841.4	4
kbs	30	30	276	208	345.6	58
mead	4	4	455	288	538.5	6
wooster	3	3	171	215	274.7	4
wooster	3	3	154	80	173.5	4
wooster	3	3	127	215	249.7	4
wooster	3	3	179	80	196.1	4
wooster	3	3	127	171	213	4
wooster	3	3	179	154	236.1	4

Note that in this report's validation dataset, uncertainty was expressed as standard error in all studies that reported it. If any sites had reported standard deviation, the standard error of the difference for those pairs of observation would have been $\sqrt{\mathrm{sd}_1^2/n_1 + \mathrm{sd}_2^2/n_2}$. The needed summations over the product of degrees of freedom (df se diff) and standard error (se diff) to compute PMU are shown in Table 22.

Table 22: Computing pooled measurement uncertainty for CROP x corn from the standard errors of differences shown in 21.

site	n trt1	n trt2	se trt1	se trt2	se diff	df se diff	se^2	$\mathrm{se}^2 \cdot df$
hoytville	3	3	670	419	790.2	4	624416	2.498e+06
hoytville	3	3	419	37	420.6	4	176904	7.076e+05
hoytville	3	3	670	509	841.4	4	707954	2.832e+06
kbs	30	30	276	208	345.6	58	119439	6.927e+06
mead	4	4	455	288	538.5	6	289982	1.74e+06
wooster	3	3	171	215	274.7	4	75460	3.018e+05
wooster	3	3	154	80	173.5	4	30102	1.204e+05
wooster	3	3	127	215	249.7	4	62350	2.494e+05
wooster	3	3	179	80	196.1	4	38455	1.538e+05
wooster	3	3	127	171	213	4	45369	1.815e+05
wooster	3	3	179	154	236.1	4	55743	2.23e+05
sum	NA	NA	NA	NA	NA	100	NA	1.593e+07
$\sqrt{\sum \mathrm{se}^2 \cdot df / \sum df}$ = **PMU**	NA	NA	NA	NA	NA	NA	NA	399.2

9.3 PMU coverage by category

Table 23: Pooled measurement uncertainty of difference in SOC between treatments (g m−2 across entire observation interval), computed for each CFG x PC. 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. % obs: percentage of the observation pairs in the full dataset (Table 5) with uncertainty available. % sites: percentage of the sites in the full dataset (Table 5) 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. Citations: a: (Collins et al., 1999); b: (Elliott et al., 1994); c: (Ghimire, Machado, and Rhinhart, 2015); d: (Jin and Varvel, 2018b; Jin et al., 2015); e: (Station, 2021); f: (Senapati et al., 2014)

PC	CFG	n obs	n stacked	n sites	PMU	citations	% obs	% sites
CROP	corn	11	1	4	399.2	a,b,e	5	24
CROP	cotton	6	0	1	240.2	f	4	17
CROP	soy	11	1	4	399.2	a,b,e	4	20
CROP	wheat	13	1	5	342.4	a,b,e,f	4	22
DISTURB	corn	22	0	6	656.8	a,b,d,e	10	46
DISTURB	cotton	6	0	1	209.3	f	12	25
DISTURB	soy	8	0	4	467.5	a,b,e	12	44
DISTURB	wheat	6	0	4	378.6	a,c,e	7	36
NFERT	All	1	1	1	529.6	b	5	14
NFERT	corn	20	2	3	868.6	b,d,e	12	20
NFERT	soy	2	2	2	366.8	b,e	3	29
NFERT	wheat	8	2	3	306.1	b,c,e	5	21
ORG	All	4	1	2	372.8	b,c	7	20
ORG	corn	1	1	1	529.6	b	7	17
ORG	soy	1	1	1	529.6	b	14	33
ORG	wheat	4	1	2	372.8	b,c	8	25
All PCs	All CFGs	76	2	9	622.2	a,b,c,d,e,f	7	22

9.4 Bias across all categories

Pooled measurement uncertainty for the entire dataset: 622.19 g C m^{-2}

Table 24: Model bias per study (g C/m^2) across all PCs and CFGs. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Fold	Study	n treatment pairs	Bias
1	ardec_manure	3	335.9
1	swiftcurrent	70	166.4

1	mead2	117	60.17
1	hoytville	8	60.02
1	narrabri_fieldC1	12	36.96
1	mandan_crop	30	-36.44
1	imperial	3	-119.4
1	goias	4	-140.3
2	lethbridge_manure	6	1537
2	brookings_REAP	14	252.3
2	rosemount	99	224.3
2	ardec1	4	192.8
2	russellranch_LTRAS	20	31.06
2	fort_valley	126	-28.39
2	sidney	3	-57.64
2	five_points	12	-244.7
2	dixonsprings	15	-277
3	dalhart	2	207
3	narrabri_fieldD1	42	170.2
3	lexington	32	46.39
3	scharleston	6	-101.5
3	saginaw	30	-103.2
3	rodale	6	-109.5
3	lethbridgeABC	10	-315.4
3	akron	1	-545.9
3	morrow	7	-624.2
4	mead	2	223.1
4	lethbridge2	28	97.09
4	ithacaNE	18	3.69
4	otis	3	-25.16
4	narrabri_field6	50	-67.63
4	broadbalk	18	-485.4
4	elansing2	3	-785.2
5	davis2	5	315.2
5	tribune	6	270.1
5	pendleton2	63	134
5	lafayette	3	31.88
5	pendleton1	61	-50.47
5	wooster	46	-88.42
5	ithaca2	27	-97.89
5	kbs	2	-250.4
1	All studies	247	45.4
2	All studies	299	181.1
3	All studies	136	-152.9
4	All studies	122	-148.5
5	All studies	213	32.99

NA	Across all folds and studies	1017	-3.87

Mean bias across all studies and PC x CFG combinations: -3.87 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.5 CROP x corn x SOC

Pooled measurement uncertainty (PMU) = 399.18 g C m^{-2}

Table 25: Model bias per study (g C/m^2) for CROP x corn. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
hoytville	3	475.3
dalhart	1	454.8
brookings_REAP	6	340.5
davis2	5	315.2
tribune	6	270.1
russellranch_LTRAS	6	45.09
fort_valley	54	10.42
otis	3	-25.16
wooster	24	-48.29
mead	1	-76.01
saginaw	30	-103.2
rodale	6	-109.5
imperial	3	-119.4
mead2	54	-150.5
goias	3	-188.4
morrow	4	-320.5
kbs	1	-620.9
Across all studies	210	8.79

Mean bias across all studies = 8.79 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.6 CROP x cotton x SOC

Pooled measurement uncertainty (PMU) = 240.25 g C m^{-2}

Table 26: Model bias per study (g C/m^2) CROP x cotton. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
narrabri_fieldD1	42	170.2
fort_valley	54	10.42
narrabri_fieldC1	6	-37.74
narrabri_field6	50	-67.63
goias	4	-140.3
five_points	6	-228.5
Across all studies	162	-48.93

Mean bias across all studies = -48.93 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.7 CROP x soy x SOC

Pooled measurement uncertainty (PMU) = 399.18 g C m^{-2}

Table 27: Model bias per study (g C/m^2) for CROP x soy. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
hoytville	3	475.3
brookings_REAP	6	340.5
tribune	3	334.2
davis2	5	315.2
narrabri_fieldD1	35	110
fort_valley	45	11.57
otis	3	-25.16
russellranch_LTRAS	8	-37.06
narrabri_field6	45	-45.15
wooster	24	-48.29
swiftcurrent	20	-60.94
mead	1	-76.01
saginaw	28	-95.06
rodale	6	-109.5
imperial	3	-119.4

goias	4	-140.3
mead2	45	-171.6
five_points	6	-228.5
morrow	4	-320.5
kbs	1	-620.9
Across all studies	295	-25.59

Mean bias across all studies = -25.59 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.8 CROP x wheat x SOC

Pooled measurement uncertainty (PMU) = 342.4 g C m^{-2}

Table 28: Model bias per study (g C/m^2) for CROP x wheat. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
hoytville	1	544
dalhart	1	454.8
brookings_REAP	6	340.5
davis2	5	315.2
tribune	6	270.1
swiftcurrent	55	133.8
lethbridge2	24	56.5
narrabri_fieldD1	35	50.58
fort_valley	45	13.88
russellranch_LTRAS	11	-12.08
otis	3	-25.16
narrabri_fieldC1	6	-37.74
narrabri_field6	35	-55.34
mead	1	-76.01
saginaw	30	-103.2
rodale	6	-109.5
imperial	3	-119.4
mead2	18	-225
five_points	6	-228.5
wooster	16	-275.1
lethbridgeABC	10	-315.4
kbs	1	-620.9

morrow	2	-782.4
Across all studies	326	-35.06

Mean bias across all studies = -35.06 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.9 DISTURB x corn x SOC

Pooled measurement uncertainty (PMU) = 656.8 g C m^{-2}

Table 29: Model bias per study (g C/m^2) for DISTURB x corn. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
rosemount	93	209.8
brookings_REAP	8	186.1
lexington	8	129.9
kbs	1	120.2
lafayette	3	31.88
ithacaNE	18	3.69
fort_valley	36	-94.9
scharleston	6	-101.5
wooster	22	-132.2
ithaca2	9	-162.7
hoytville	5	-189.1
dixonsprings	15	-277
goias	1	-477.6
Across all studies	225	-57.95

Mean bias across all studies = -57.95 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.10 DISTURB x cotton x SOC

Pooled measurement uncertainty (PMU) = 209.31 g C m^{-2}

Table 30: Model bias per study (g C/m^2) for DISTURB x cotton. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
narrabri_fieldC1	6	111.7
fort_valley	36	-94.9
five_points	6	-260.9
goias	1	-477.6
Across all studies	49	-180.4

Mean bias across all studies = -180.45 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.11 DISTURB x soy x SOC

Pooled measurement uncertainty (PMU) = 467.52 g C m^{-2}

Table 31: Model bias per study (g C/m^2) for DISTURB x soy. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
brookings_REAP	8	186.1
kbs	1	120.2
lafayette	3	31.88
fort_valley	18	-82.35
wooster	14	-198.3
five_points	3	-241.4
dixonsprings	15	-277
hoytville	3	-313.1
goias	1	-477.6
Across all studies	66	-139.1

Mean bias across all studies = -139.06 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.12 DISTURB x wheat x SOC

Pooled measurement uncertainty (PMU) = 378.63 g C m^{-2}

Table 32: Model bias per study (g C/m^2) for DISTURB x wheat. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
brookings_REAP	4	225.5
pendleton2	18	139.7
kbs	1	120.2
pendleton1	16	86.55
mandan_crop	12	85.43
wooster	7	-44.63
sidney	3	-57.64
fort_valley	18	-82.04
five_points	6	-260.9
akron	1	-545.9
hoytville	1	-701.7
Across all studies	87	-94.13

Mean bias across all studies = -94.13 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.13 NFERT x corn x SOC

Pooled measurement uncertainty (PMU) = 868.65 g C m^{-2}

Table 33: Model bias per study (g C/m^2) for NFERT x corn. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
mead	1	522.2
rosemount	6	447.9
davis2	4	385.1
mead2	54	266.5
ardec_manure	2	251.9
ardec1	4	192.8

lexington	24	18.54
fort_valley	36	-20.09
dalhart	1	-40.76
ithaca2	18	-65.5
russellranch_LTRAS	3	-84.2
rodale	6	-109.5
kbs	1	-620.9
elansing2	3	-785.2
morrow	3	-1029
Across all studies	166	-44.69

Mean bias across all studies = -44.69 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.14 NFERT x soy x SOC

Pooled measurement uncertainty (PMU) = 366.79 g C m^{-2}

Table 34: Model bias per study (g C/m^2) for NFERT x soy. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
mead	1	522.2
davis2	4	385.1
mead2	45	166.7
fort_valley	18	-31.3
rodale	6	-109.5
kbs	1	-620.9
morrow	2	-732.7
Across all studies	77	-60.05

Mean bias across all studies = -60.05 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.15 NFERT x wheat x SOC

Pooled measurement uncertainty (PMU) = 306.06 g C m^{-2}

Table 35: Model bias per study (g C/m^2) for NFERT x wheat. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
mead	1	522.2
davis2	4	385.1
swiftcurrent	15	285.6
mead2	18	223
pendleton2	45	131.6
russellranch_LTRAS	9	83.79
pendleton1	24	-9.44
fort_valley	18	-34.63
dalhart	1	-40.76
rodale	6	-109.5
mandan_crop	18	-117.7
broadbalk	12	-268.6
morrow	1	-416
kbs	1	-620.9
Across all studies	173	0.98

Mean bias across all studies = 0.98 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.16 ORG x All x SOC

Follows Model Requirements section 3.3.1, paragraph 5

Pooled measurement uncertainty (PMU) = 372.78 g C m^{-2}

Table 36: Model bias per study (g C/m^2) for ORG x all annual CFGs. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
lethbridge_manure	6	1537
davis2	2	633.9
ardec_manure	2	540.5
mead	1	522.2
lethbridge2	4	340.6

russellranch_LTRAS	3	-84.2
pendleton1	21	-201.8
rodale	4	-446.4
broadbalk	12	-630.4
elansing2	3	-785.2
Across all studies	58	142.6

Mean bias across all studies = 142.64 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.17 ORG x corn x SOC

This category is presented for context only to support the use of its data as part of Section 9.16 "ORG x All x SOC".

Pooled measurement uncertainty (PMU) = 529.59 g C m^{-2}

Table 37: Model bias per study (g C/m^2) for ORG x corn. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
davis2	2	633.9
ardec_manure	2	540.5
mead	1	522.2
russellranch_LTRAS	3	-84.2
rodale	4	-446.4
elansing2	3	-785.2
Across all studies	15	63.45

Mean bias across all studies = 63.45 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.18 ORG x soy x SOC

This category is presented for context only to support the use of its data as part of Section 9.16 "ORG x All x SOC".

Pooled measurement uncertainty (PMU) = 529.59 g C m^{-2}

Table 38: Model bias per study (g C/m^2) for ORG x soy. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
davis2	2	633.9
mead	1	522.2
rodale	4	-446.4
Across all studies	7	236.6

Mean bias across all studies = 236.55 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

9.19 ORG x wheat x SOC

Pooled measurement uncertainty (PMU) = 372.78 g C m^{-2}

Table 39: Model bias per study (g C/m^2) for ORG x wheat. Note that each study is validated in exactly one of the 5 k-folds so only one bias value can be calculated. See Table A2 for study fold assignments.

Study	n treatment pairs	Bias
lethbridge_manure	6	1537
davis2	2	633.9
mead	1	522.2
lethbridge2	4	340.6
russellranch_LTRAS	3	-84.2
pendleton1	21	-201.8
rodale	4	-446.4
broadbalk	12	-630.4
Across all studies	53	208.9

Mean bias across all studies = 208.88 g C m^{-2}

Is the absolute bias smaller than the PMU? Yes

10 Model prediction error

Follows Model Requirements, Section 3.5 Summary of Requirements (p20)

10.1 Description of calculation method

Model uncertainty bounds on the difference in SOC change 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 4 "Model Calibration"). 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% 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% posterior prediction interval.

Similar to the bias calculation described in Section 9.1 "Calculating Bias", for each treatment pair the posterior prediction intervals are formed for the difference in SOC at the second time point. Coverage rates are then calculated as the proportion of these posterior prediction intervals that contain the observed difference in SOC at the second time point. As in Section 9.1 "Calculating Bias", this is equivalent to calculating coverage rates for 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 we use weakly informative independent priors (as recommended in Model Requirements section 3.5) that have a uniform distribution defined by their lower and upper bounds (see Table A1 and Section 4.3 "Documentation of model parameter sets" 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.

10.2 Model prediction error across all categories

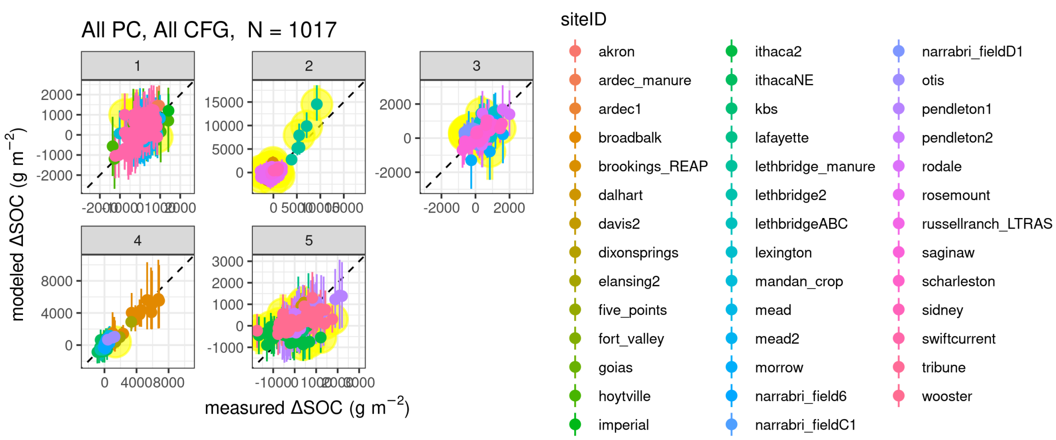

Figure 17: Model predictions versus measurements of SOC change in all practice changes and crop types, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 40 for coverage rates).

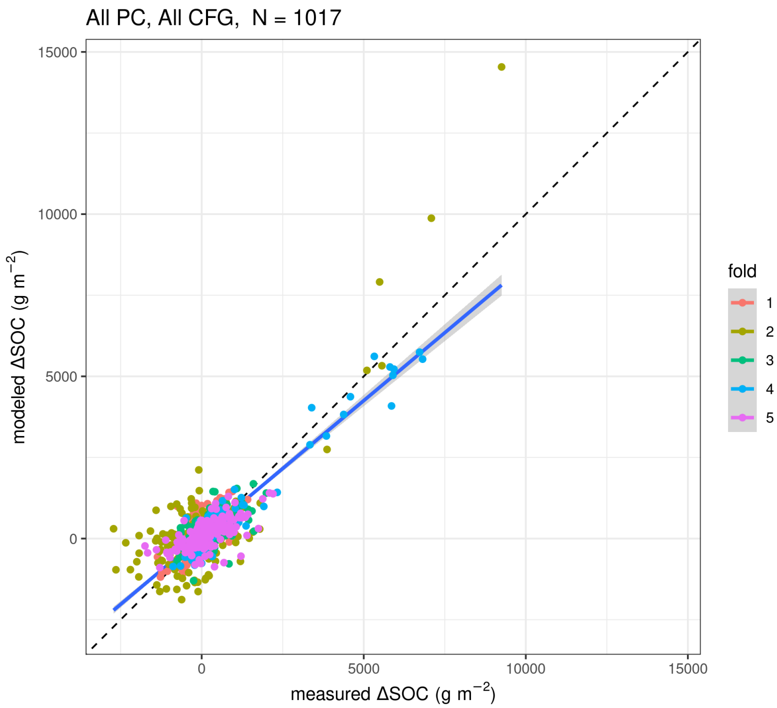

Figure 18: Scatterplot of model predictions versus measurements of SOC change in all practice changes and crop types. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

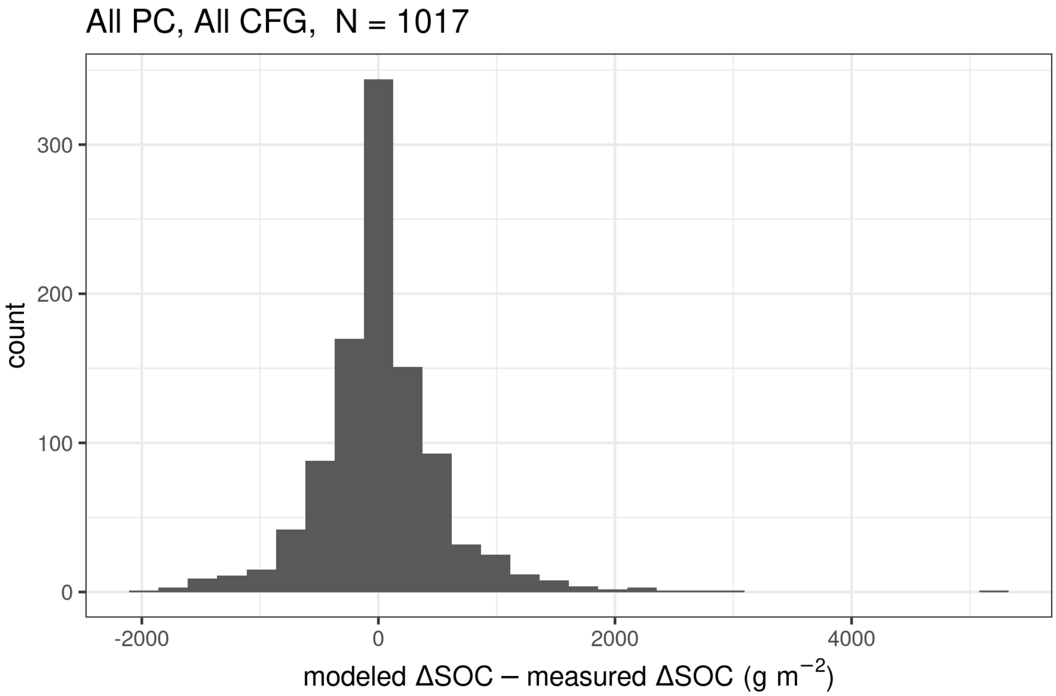

Figure 19: Histogram of model residuals (predicted - observed) for change in SOC in all studies used for model validation across all practices and crop types.

Table 40: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for each fold of the calibration/validation process across all practices and crop types.

fold	n	n in	n out	% coverage
1	247	242	5	98
2	299	277	22	93
3	136	117	19	86
4	122	121	1	99
5	213	195	18	92
All folds	1017	952	65	94

We calculate the MSE and RMSE for each MCMC iteration, and report the mean ± the standard deviation for each:

Mean squared error: 719393 ± 71026 (g C m^{-2})^2; RMSE: 847 ± 41 g C m^{-2}

Do 90% prediction intervals cover observed data 90% of the time? Yes

10.3 CROP x corn x SOC

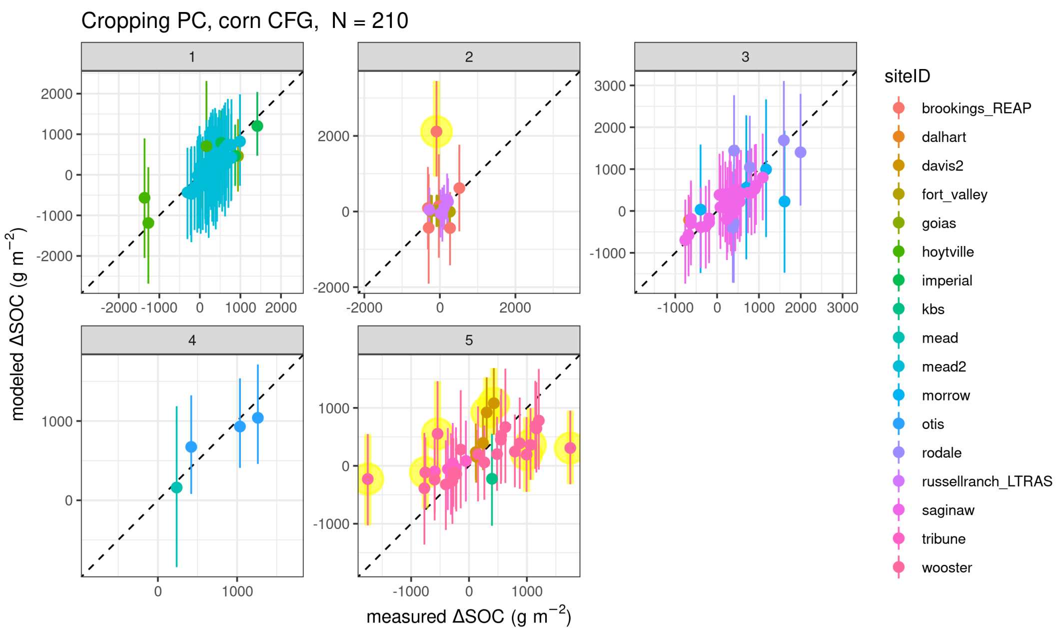

Figure 20: Model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the corn-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 41 for coverage rates).

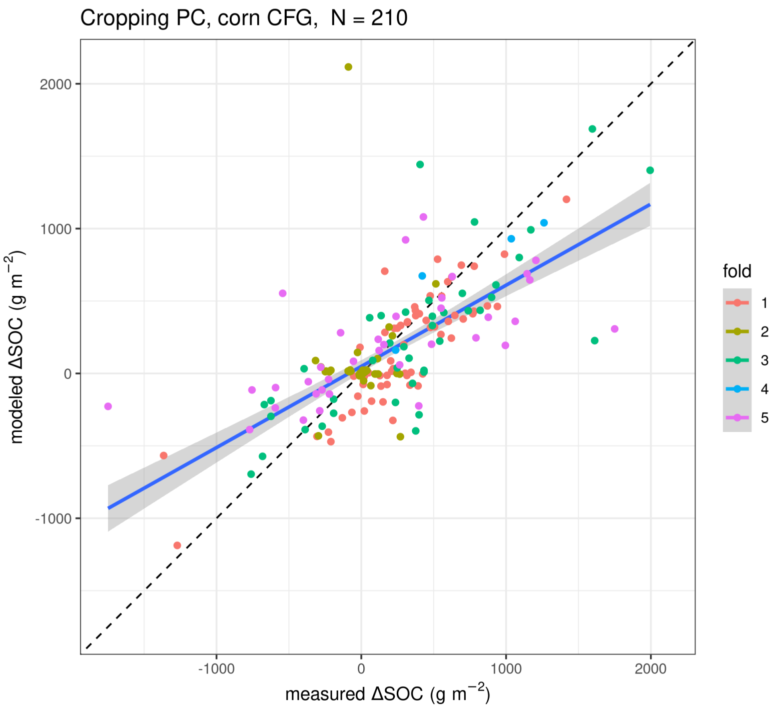

Figure 21: Scatterplot of model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the corn-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

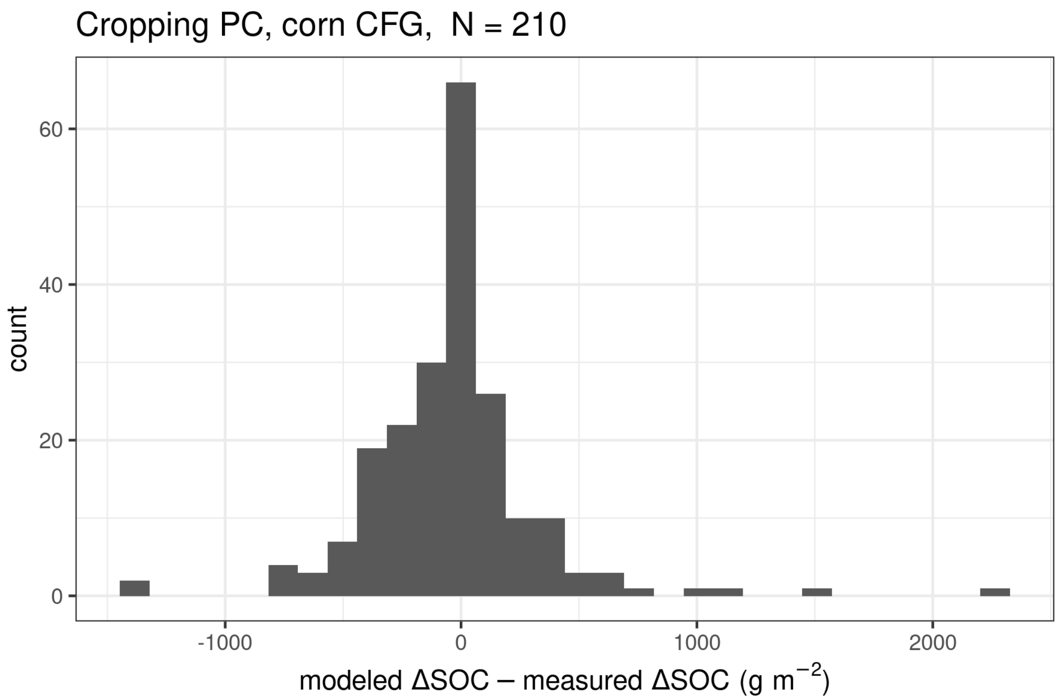

Figure 22: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, cropping practices involving crops from the corn-type CFG.

Table 41: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed cropping practices involving corn-type CFG.

fold	n	n in	n out	% coverage
1	63	63	0	100
2	66	65	1	98
3	41	41	0	100
4	4	4	0	100
5	36	28	8	78
All folds	210	201	9	96

Mean squared error: 438595 ± 74591; RMSE: 660 ± 56

Do 90% prediction intervals cover observed data 90% of the time? Yes

10.4 CROP x cotton x SOC

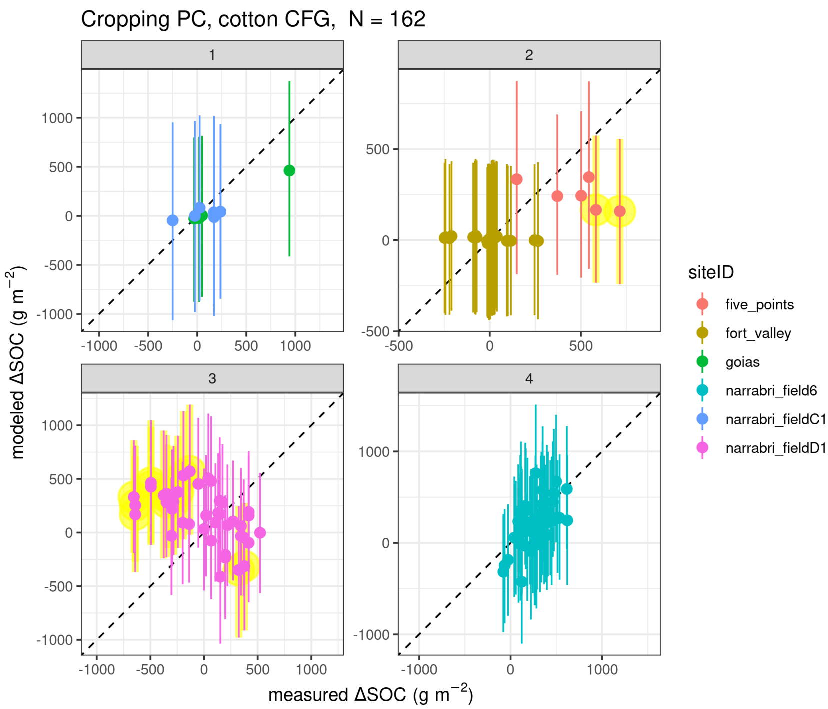

Figure 23: Model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the cotton-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 42 for coverage rates).

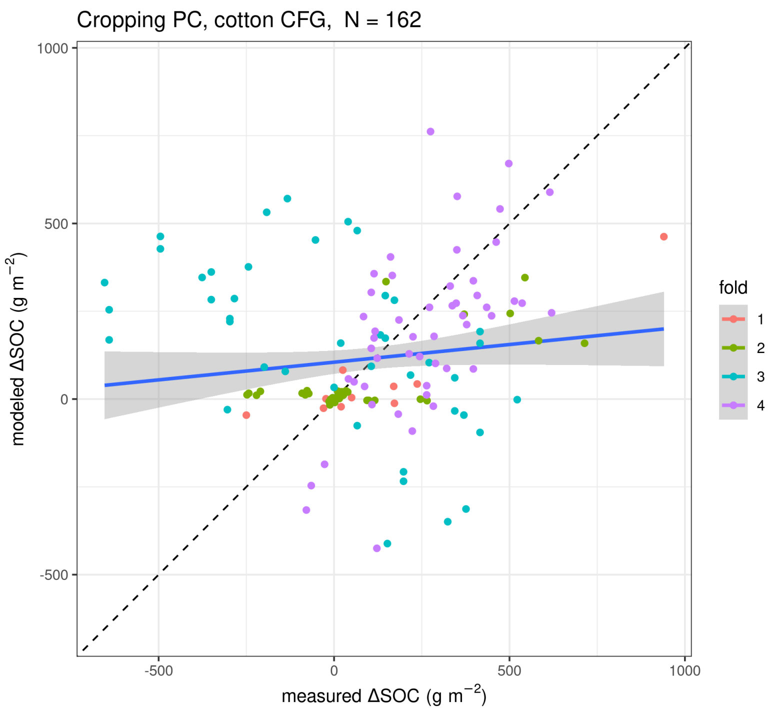

Figure 24: Scatterplot of model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the cotton-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

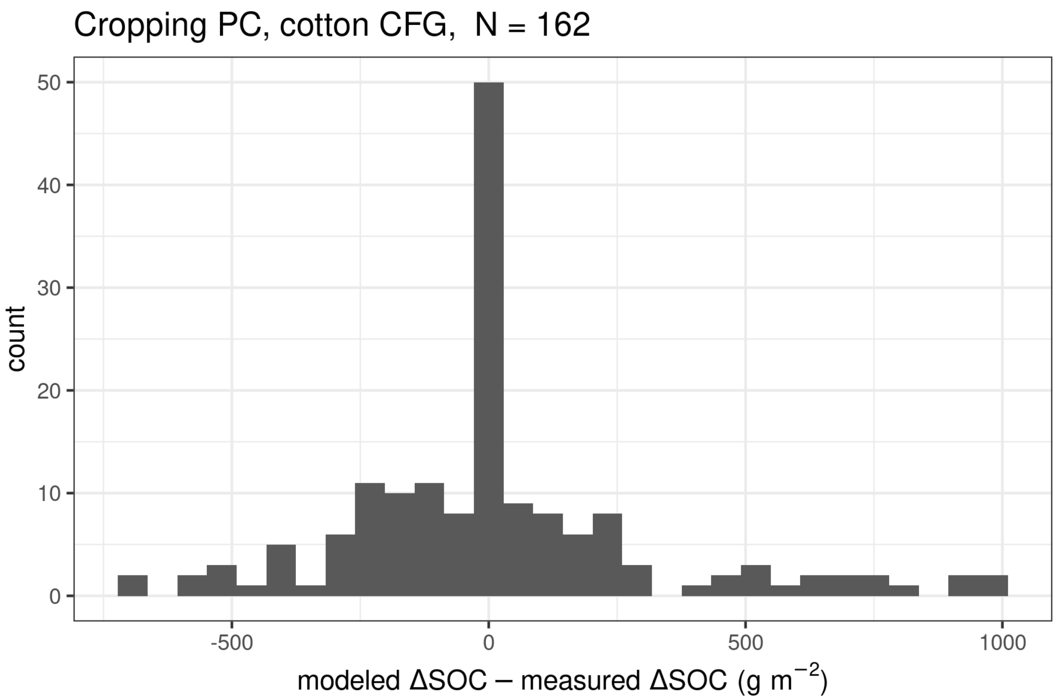

Figure 25: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, cropping practices involving crops from the cotton-type CFG.

Table 42: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed cropping practices involving cotton-type CFG.

fold	n	n in	n out	% coverage
1	10	10	0	100
2	60	58	2	97
3	42	28	14	67
4	50	50	0	100
All folds	162	146	16	90

Mean squared error: 215426 ± 37481; RMSE: 462 ± 40

Do 90% prediction intervals cover observed data 90% of the time? Yes

10.5 CROP x soy x SOC

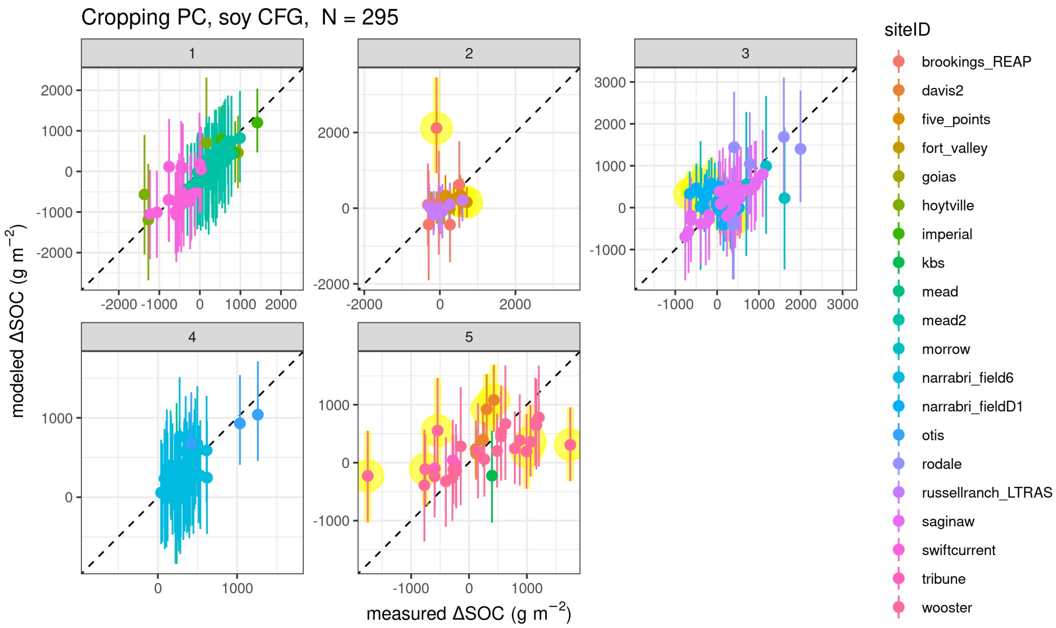

Figure 26: Model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the soy-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 43 for coverage rates).

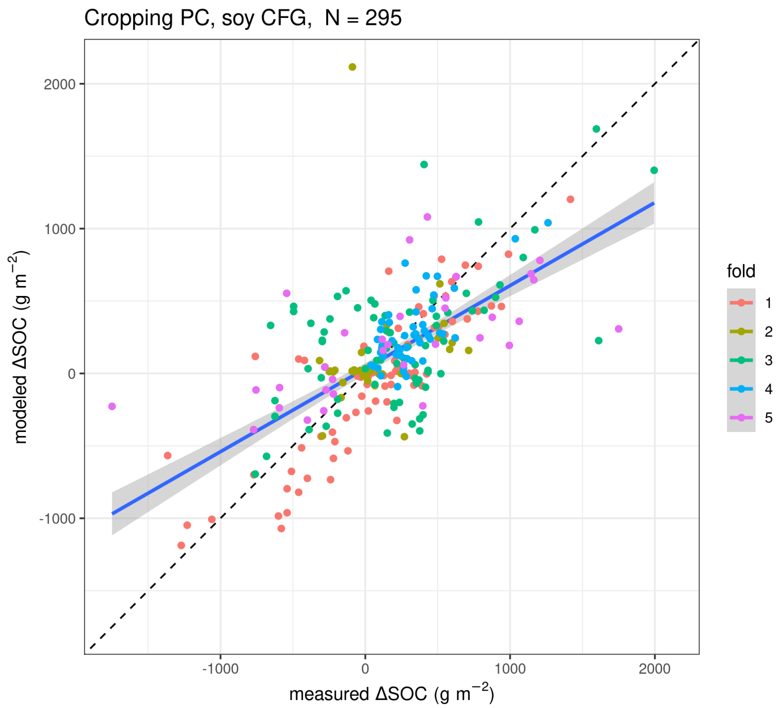

Figure 27: Scatterplot of model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the soy-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

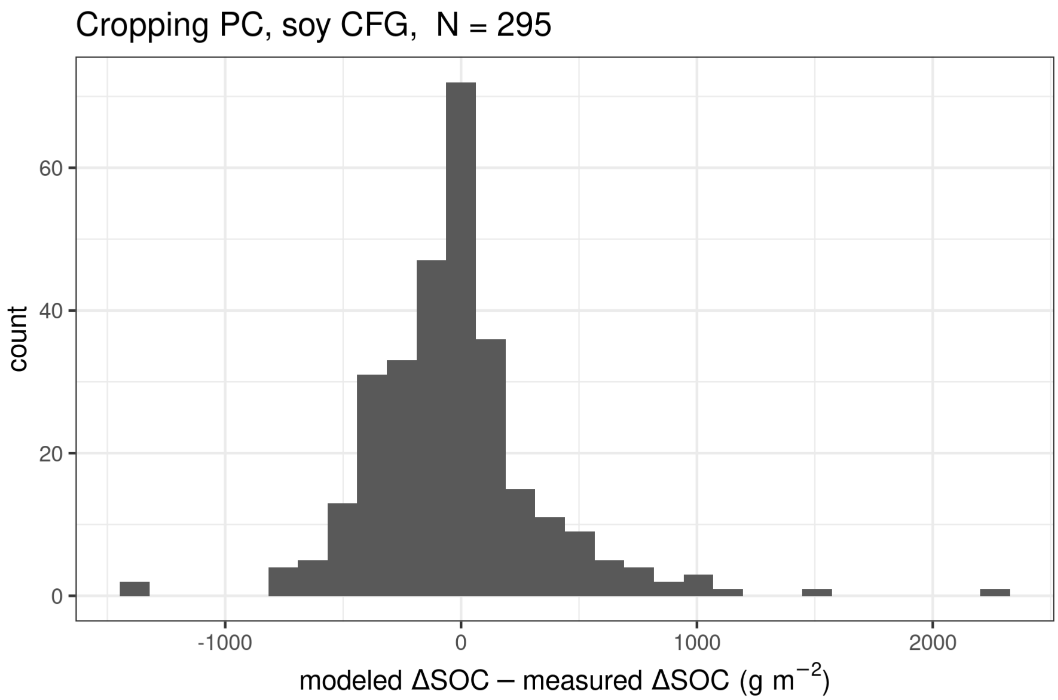

Figure 28: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, cropping practices involving crops from the soy-type CFG.

Table 43: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed cropping practices involving soy-type CFG.

fold	n	n in	n out	% coverage
1	75	75	0	100
2	65	62	3	95
3	73	63	10	86
4	49	49	0	100
5	33	25	8	76
All folds	295	274	21	93

Mean squared error: 407857 ± 55371; RMSE: 637 ± 43

Do 90% prediction intervals cover observed data 90% of the time? Yes

10.6 CROP x wheat x SOC

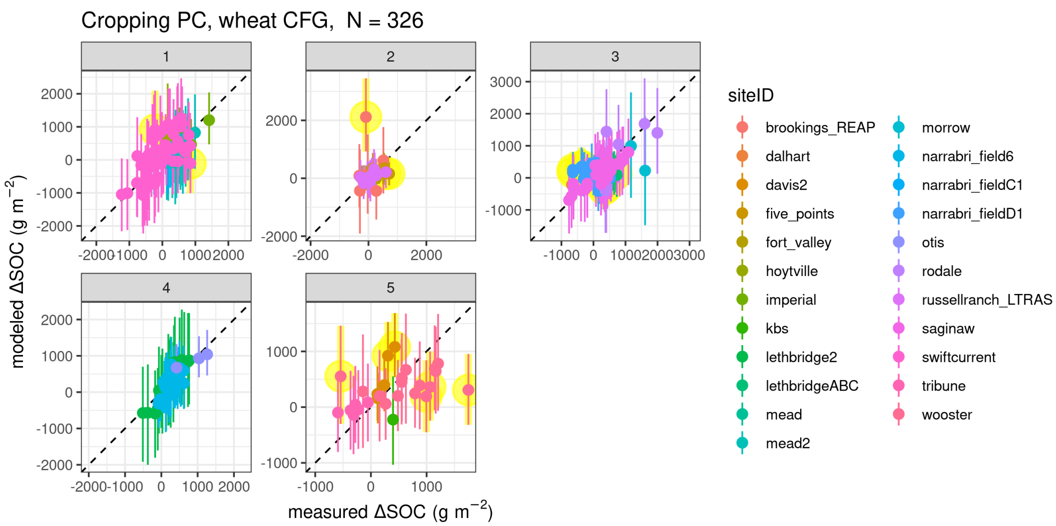

Figure 29: Model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the wheat-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 44 for coverage rates).

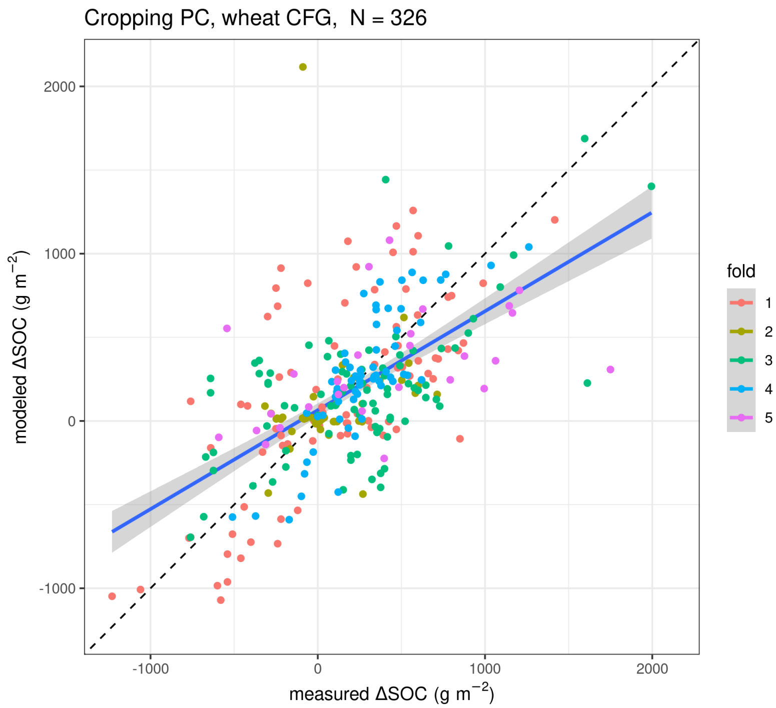

Figure 30: Scatterplot of model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the wheat-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

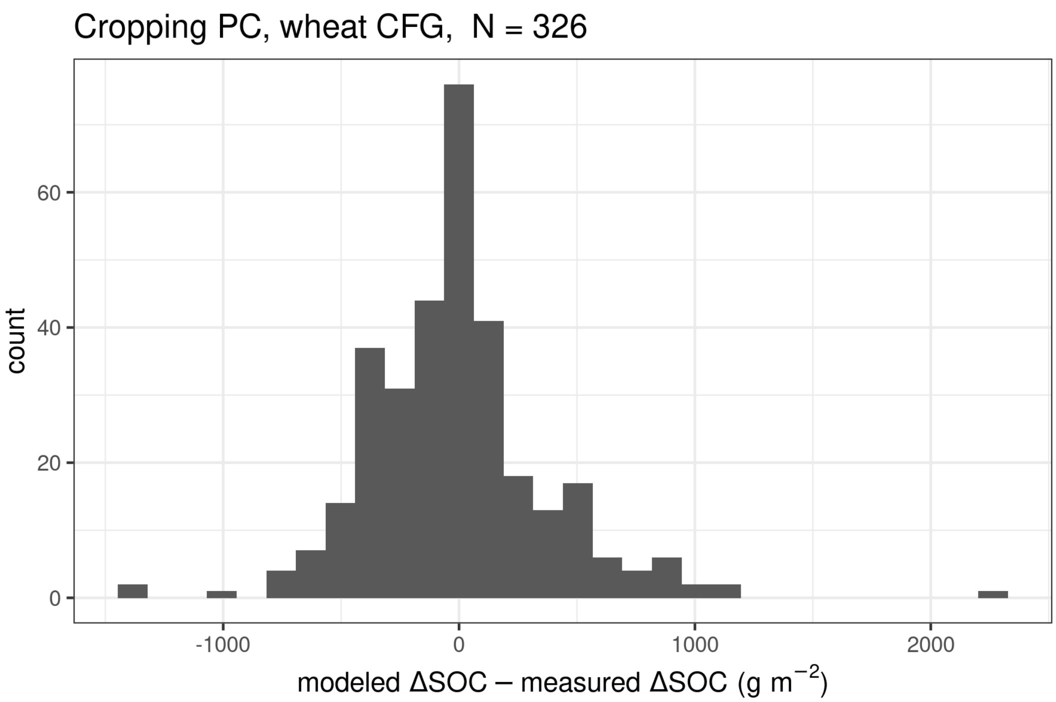

Figure 31: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, cropping practices involving crops from the wheat-type CFG.

Table 44: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed cropping practices involving wheat-type CFG.

fold	n	n in	n out	% coverage
1	83	81	2	98
2	68	65	3	96
3	84	75	9	89
4	63	63	0	100
5	28	22	6	79
All folds	326	306	20	94

Mean squared error: 425356 ± 58137; RMSE: 651 ± 44

Do 90% prediction intervals cover observed data 90% of the time? Yes

10.7 DISTURB x corn x SOC

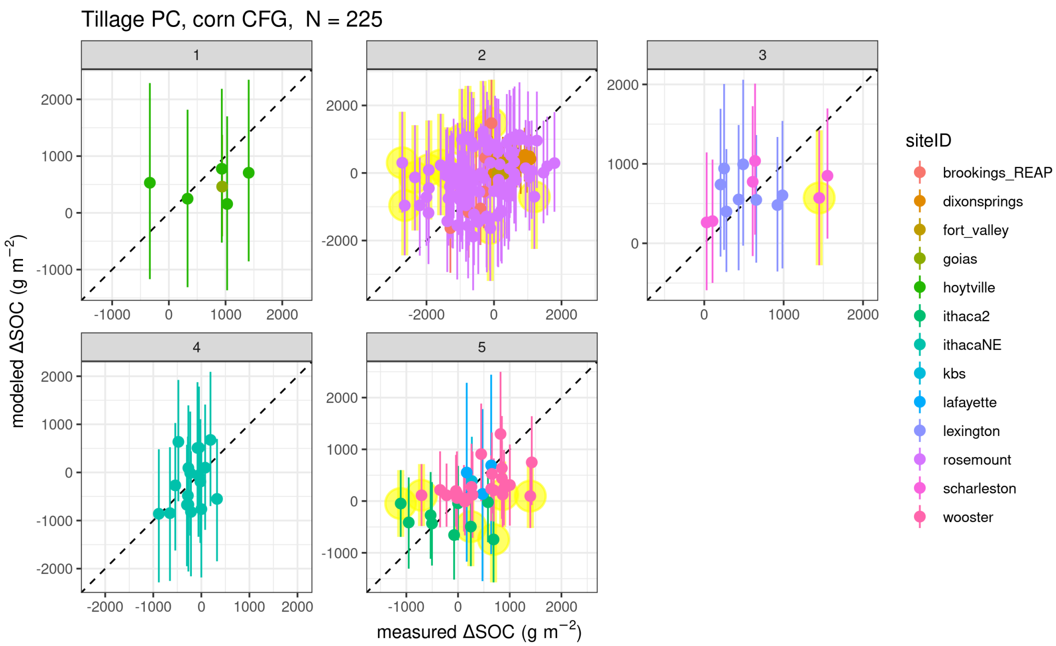

Figure 32: Model predictions versus measurements of SOC change in response to changed tillage practices involving crops from the corn-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 45 for coverage rates).

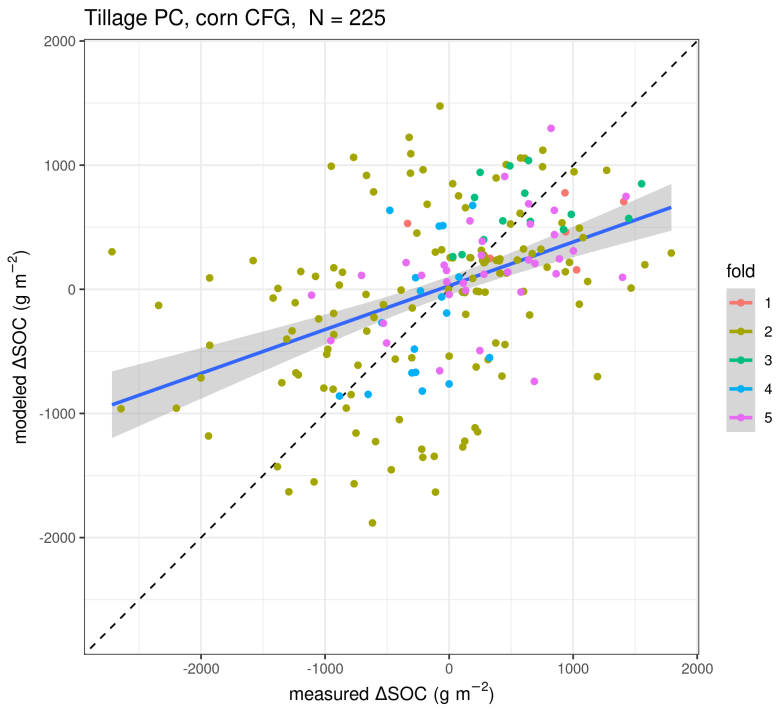

Figure 33: Scatterplot of model predictions versus measurements of SOC change in response to changed tillage practices involving crops from the corn-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

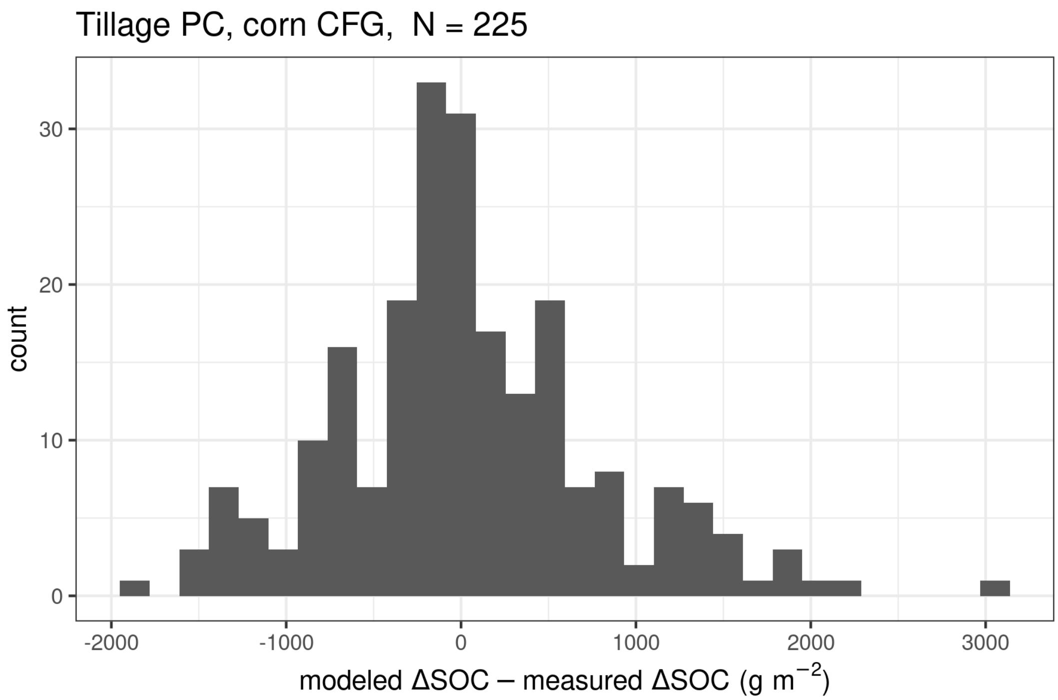

Figure 34: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, tillage practices involving crops from the corn-type CFG.

Table 45: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed tillage practices involving corn-type CFG.

fold	n	n in	n out	% coverage
1	6	6	0	100
2	152	139	13	91
3	14	13	1	93
4	18	18	0	100
5	35	29	6	83
All folds	225	205	20	91

Mean squared error: 1112412 ± 158687; RMSE: 1052 ± 75

Do 90% prediction intervals cover observed data 90% of the time? Yes

10.8 DISTURB x cotton x SOC

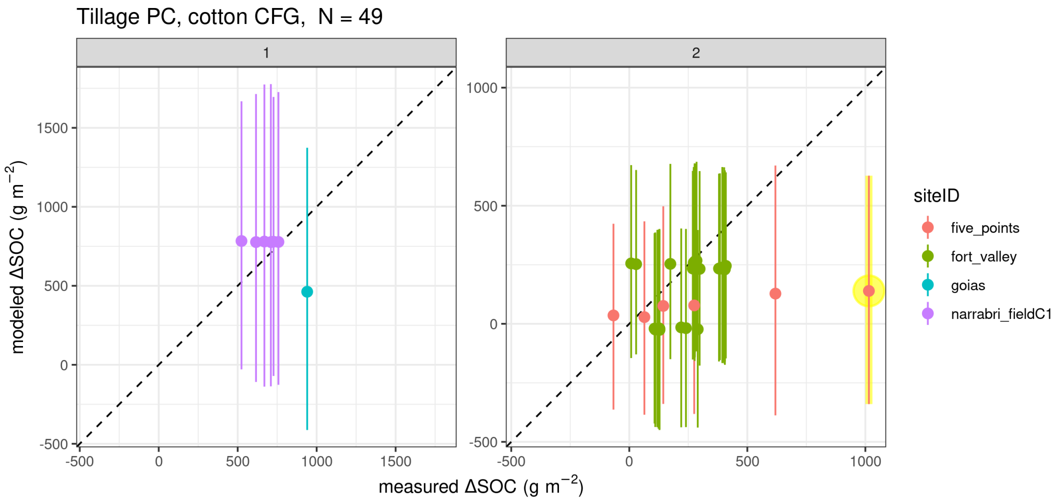

Figure 35: Model predictions versus measurements of SOC change in response to changed tillage practices involving crops from the cotton-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 46 for coverage rates).

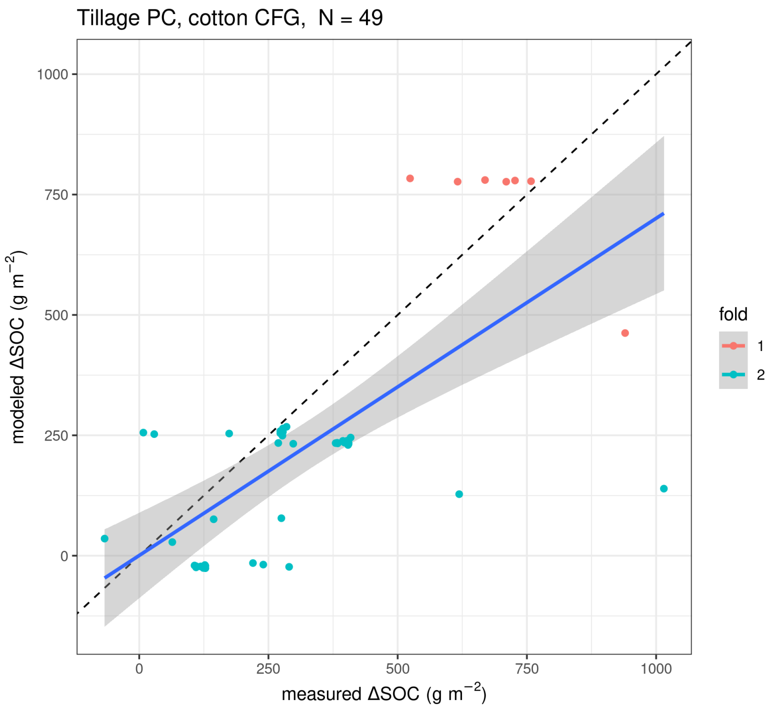

Figure 36: Scatterplot of model predictions versus measurements of SOC change in response to changed tillage practices involving crops from the cotton-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

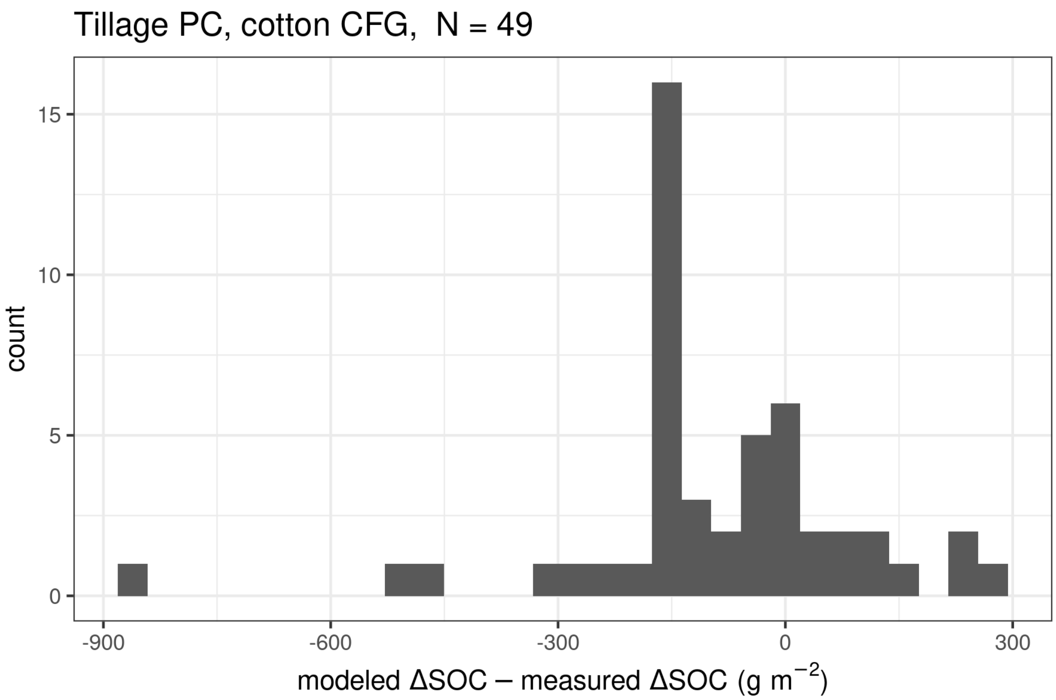

Figure 37: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, tillage practices involving crops from the cotton-type CFG.

Table 46: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed tillage practices involving cotton-type CFG.

fold	n	n in	n out	% coverage
1	7	7	0	100
2	42	41	1	98
All folds	49	48	1	98

Mean squared error: 144049 ± 37990; RMSE: 376 ± 49

Do 90% prediction intervals cover observed data 90% of the time? Yes

10.9 DISTURB x soy x SOC

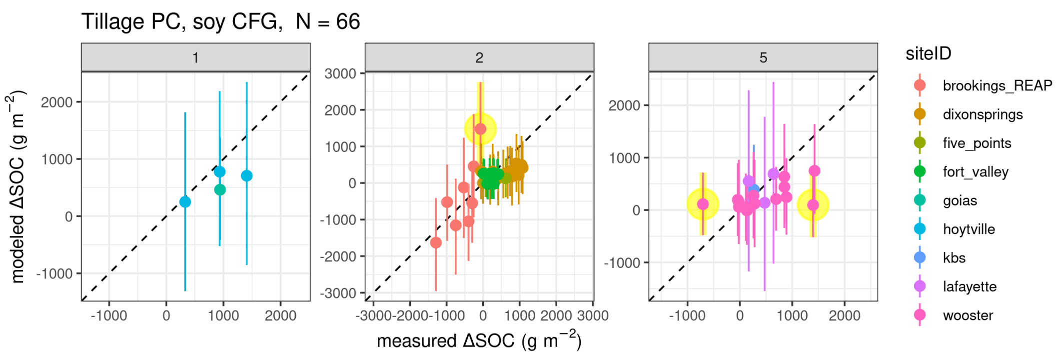

Figure 38: Model predictions versus measurements of SOC change in response to changed tillage practices involving crops from the soy-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 47 for coverage rates).

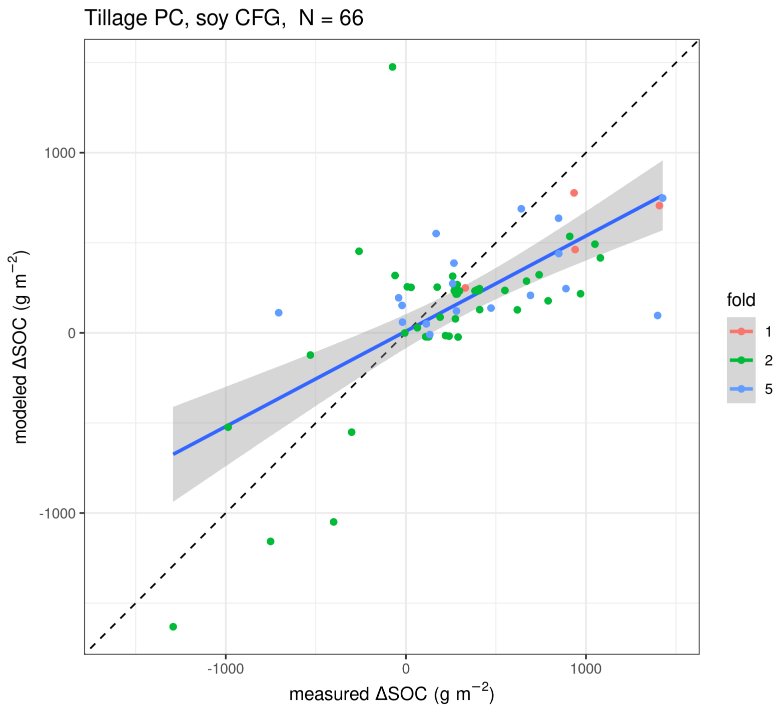

Figure 39: Scatterplot of model predictions versus measurements of SOC change in response to changed tillage practices involving crops from the soy-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

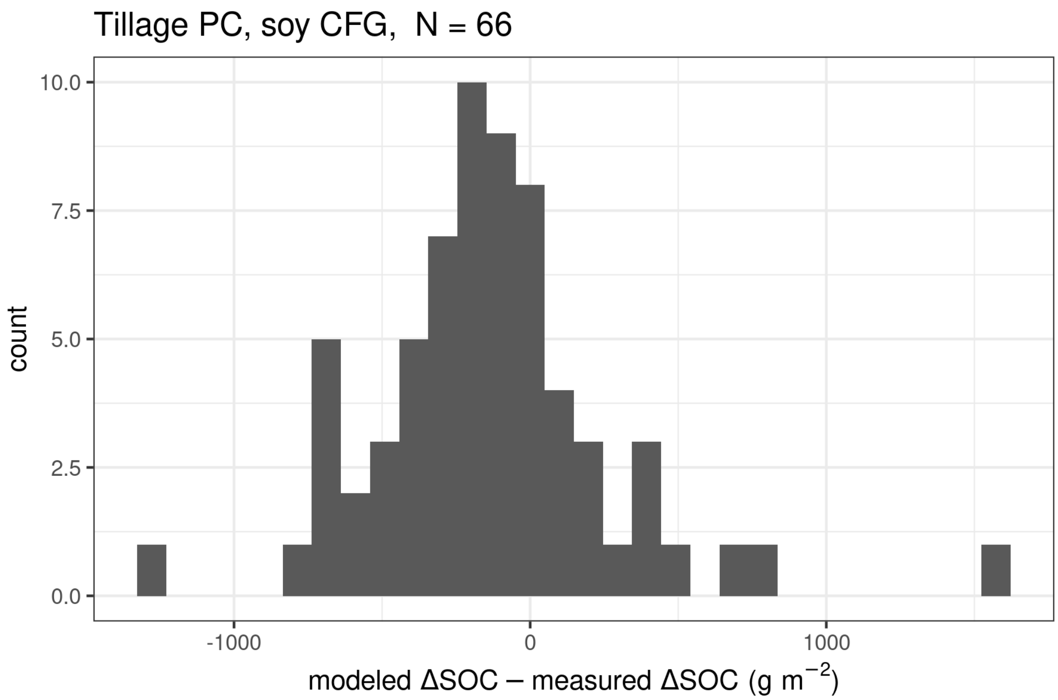

Figure 40: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, tillage practices involving crops from the soy-type CFG.

Table 47: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed tillage practices involving soy-type CFG.

fold	n	n in	n out	% coverage
1	4	4	0	100
2	44	43	1	98
5	18	16	2	89
All folds	66	63	3	95

Mean squared error: 456619 ± 109703; RMSE: 671 ± 79

Do 90% prediction intervals cover observed data 90% of the time? Yes

10.10 DISTURB x wheat x SOC

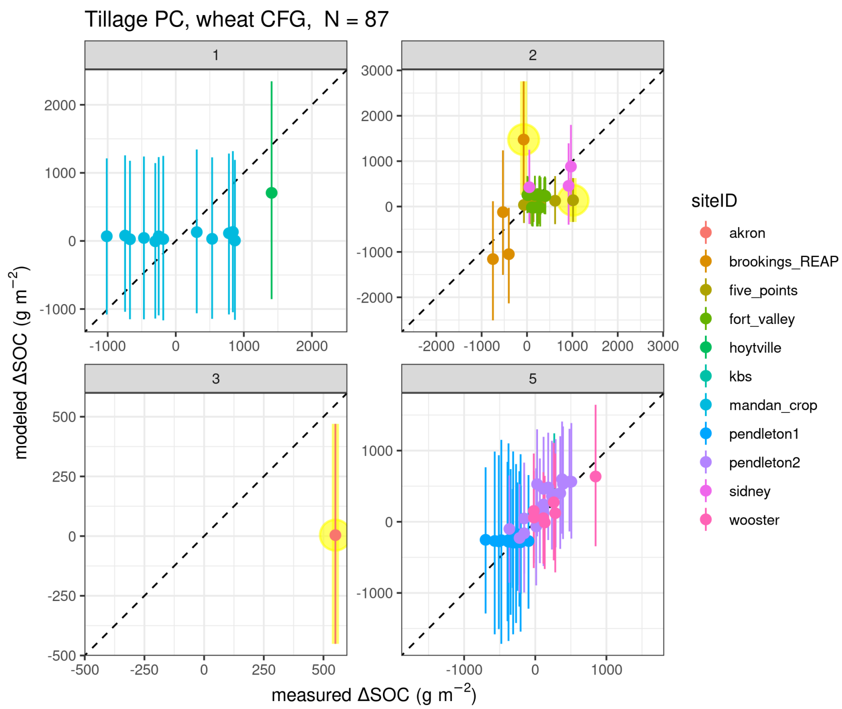

Figure 41: Model predictions versus measurements of SOC change in response to changed tillage practices involving crops from the wheat-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 48 for coverage rates).

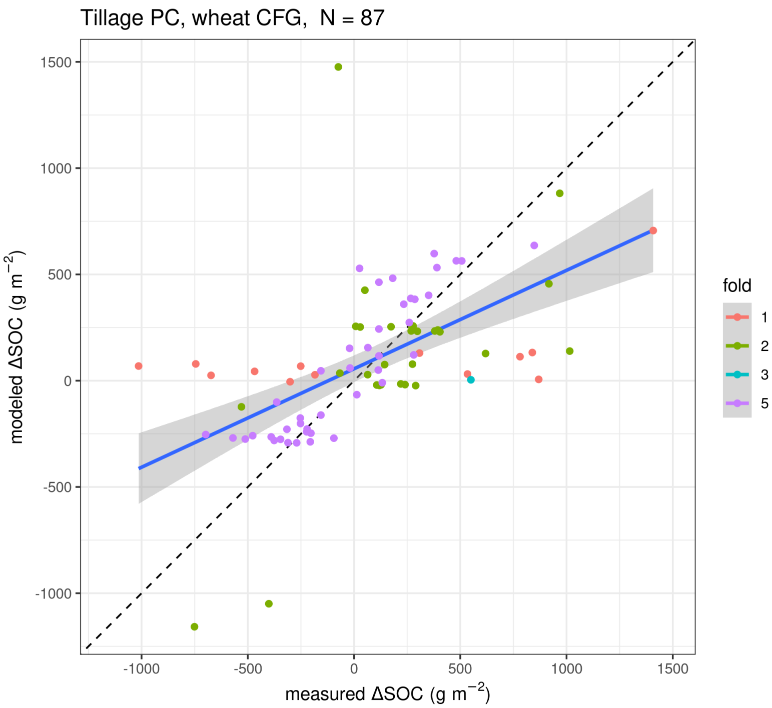

Figure 42: Scatterplot of model predictions versus measurements of SOC change in response to changed tillage practices involving crops from the wheat-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

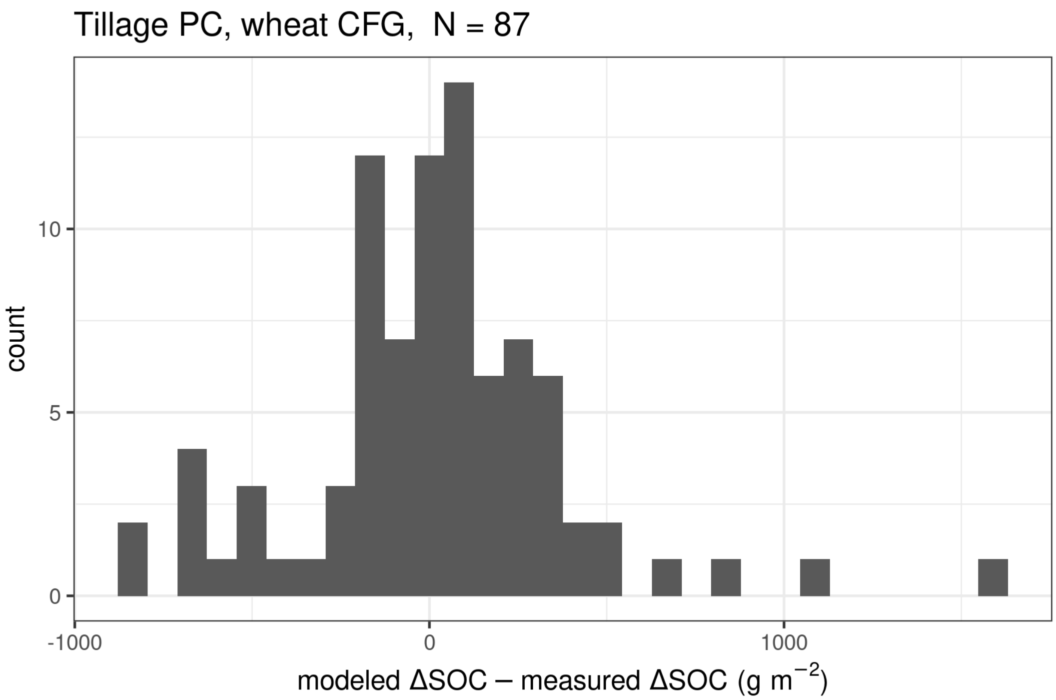

Figure 43: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, tillage practices involving crops from the wheat-type CFG.

Table 48: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed tillage practices involving wheat-type CFG.

fold	n	n in	n out	% coverage
1	13	13	0	100
2	31	29	2	94
3	1	0	1	0
5	42	42	0	100
All folds	87	84	3	97

Mean squared error: 423539 ± 91906; RMSE: 647 ± 70

Do 90% prediction intervals cover observed data 90% of the time? Yes

10.11 NFERT x corn x SOC

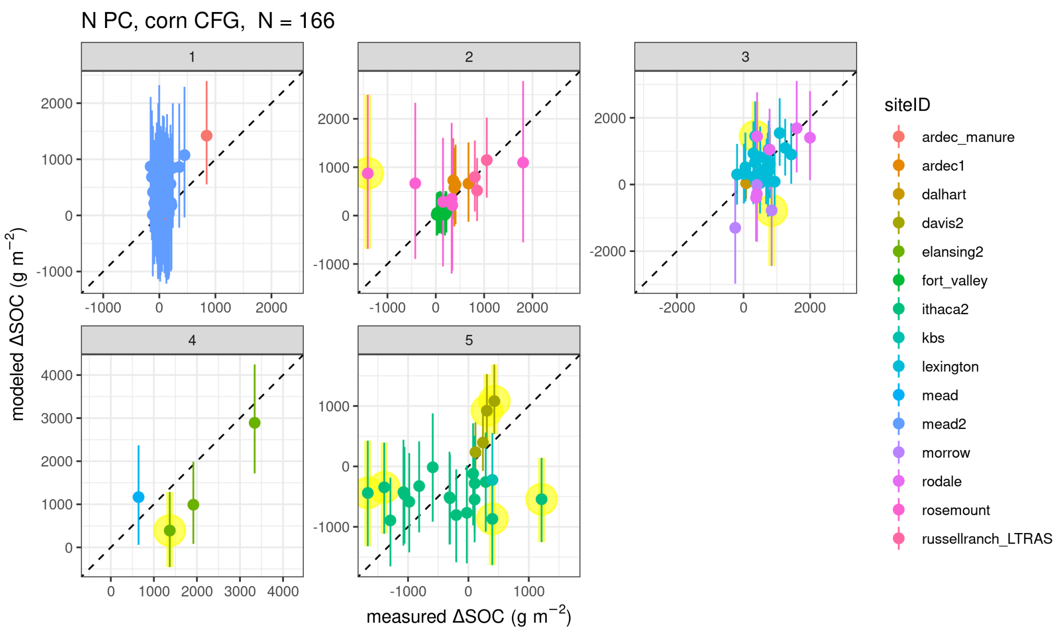

Figure 44: Model predictions versus measurements of SOC change in response to changed nitrogen practices involving crops from the corn-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 49 for coverage rates).

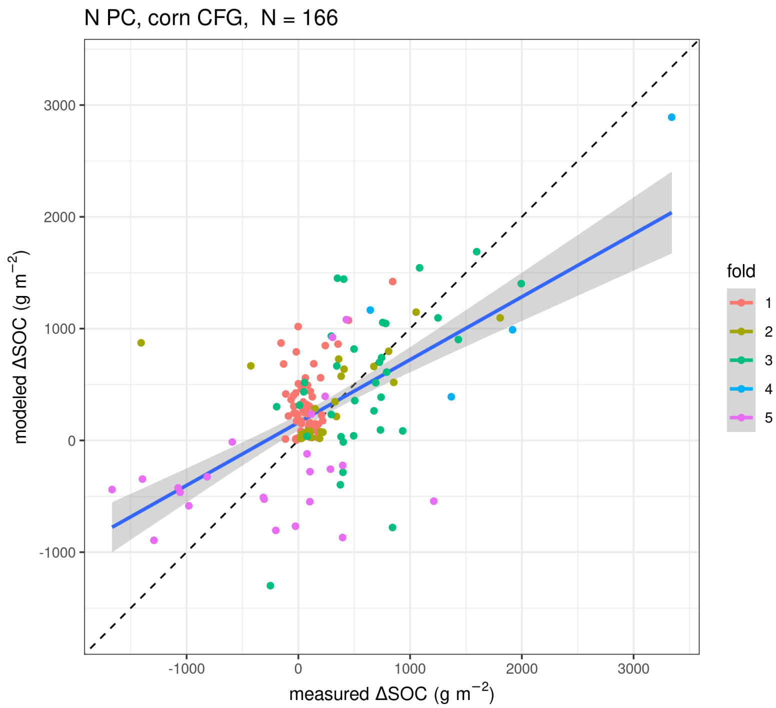

Figure 45: Scatterplot of model predictions versus measurements of SOC change in response to changed nitrogen practices involving crops from the corn-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

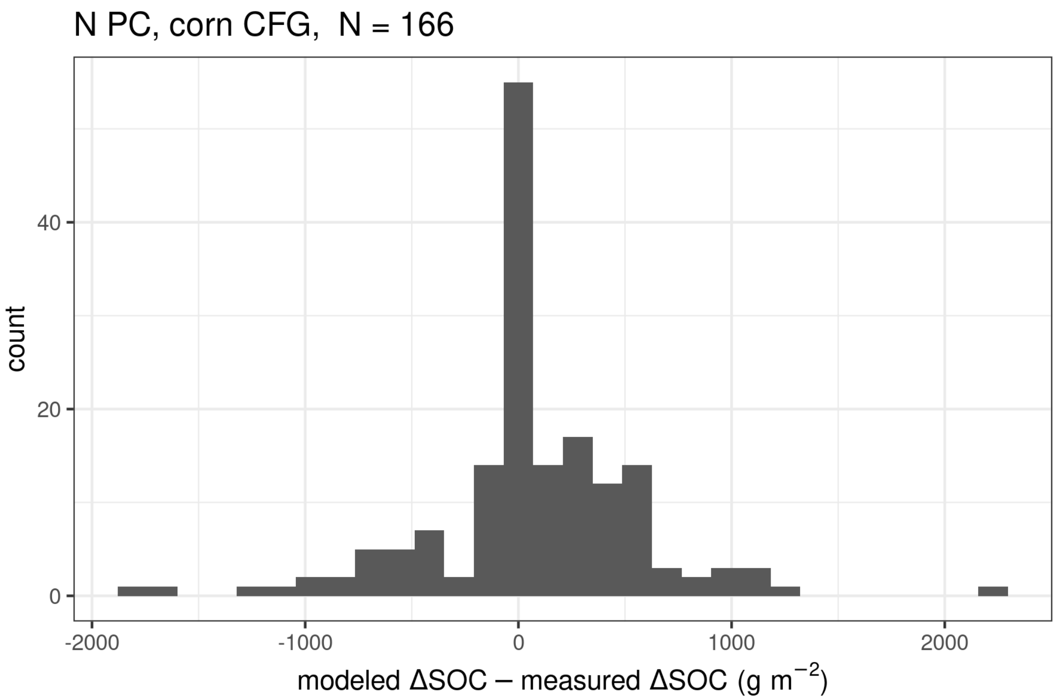

Figure 46: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, nitrogen practices involving crops from the corn-type CFG.

Table 49: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed nitrogen practices involving corn-type CFG.

fold	n	n in	n out	% coverage
1	56	56	0	100
2	49	48	1	98
3	34	32	2	94
4	4	3	1	75
5	23	17	6	74
All folds	166	156	10	94

Mean squared error: 587484 ± 99368; RMSE: 764 ± 64

Do 90% prediction intervals cover observed data 90% of the time? Yes

10.12 NFERT x soy x SOC

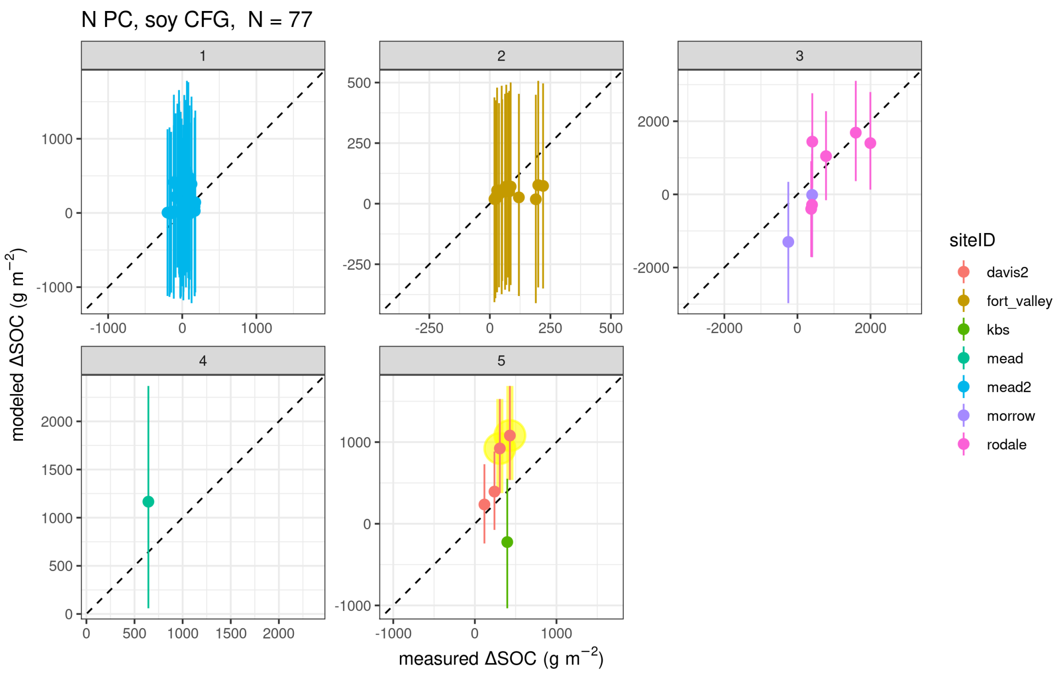

Figure 47: Model predictions versus measurements of SOC change in response to changed nitrogen practices involving crops from the soy-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 50 for coverage rates).

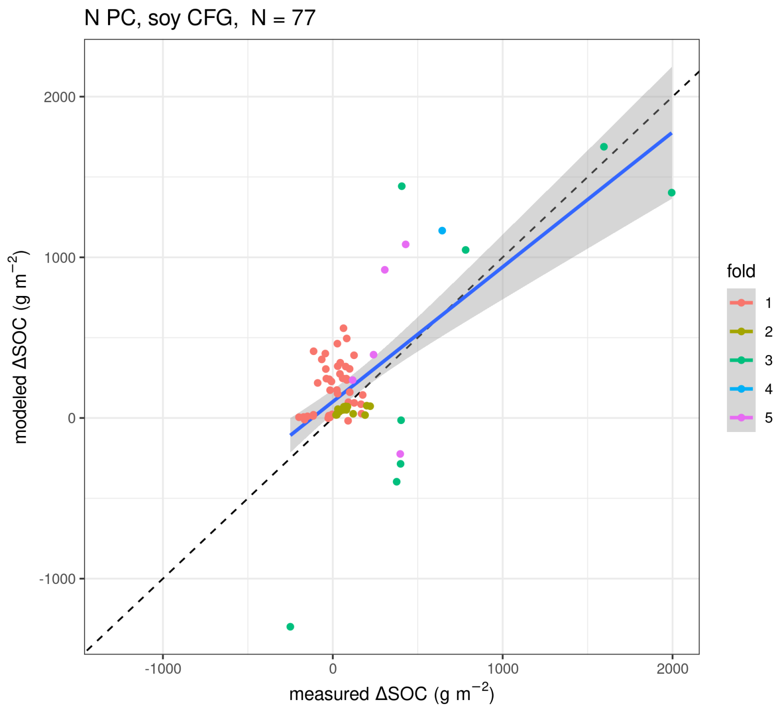

Figure 48: Scatterplot of model predictions versus measurements of SOC change in response to changed nitrogen practices involving crops from the soy-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

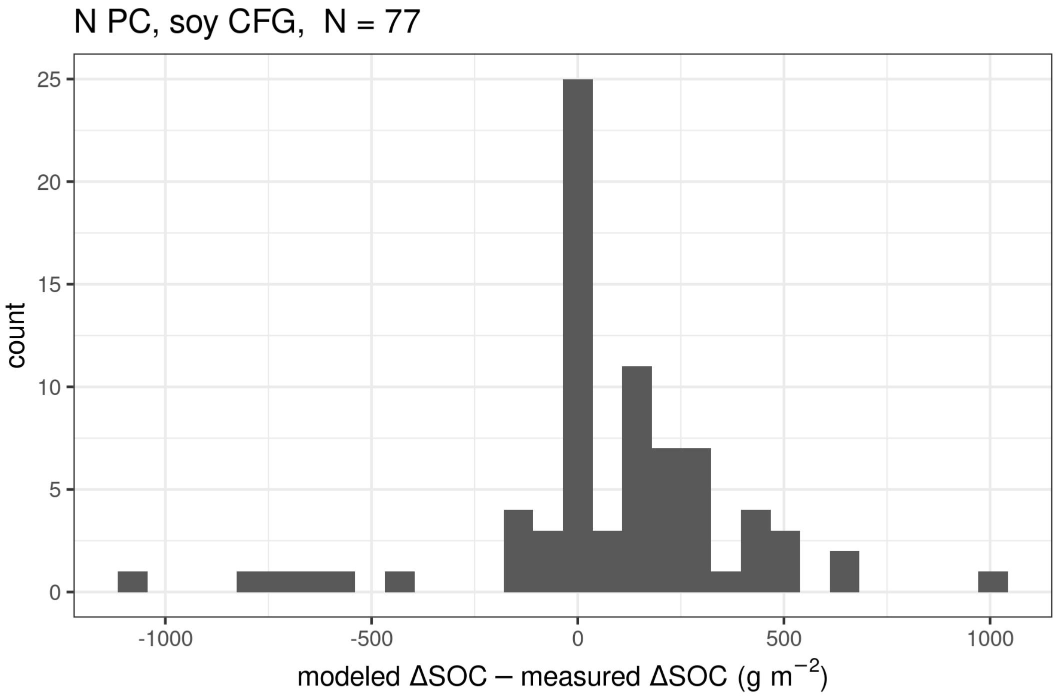

Figure 49: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, nitrogen practices involving crops from the soy-type CFG.

Table 50: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed nitrogen practices involving soy-type CFG.

fold	n	n in	n out	% coverage
1	45	45	0	100
2	18	18	0	100
3	8	8	0	100
4	1	1	0	100
5	5	3	2	60
All folds	77	75	2	97

Mean squared error: 487636 ± 132496; RMSE: 692 ± 93

Do 90% prediction intervals cover observed data 90% of the time? Yes

10.13 NFERT x wheat x SOC

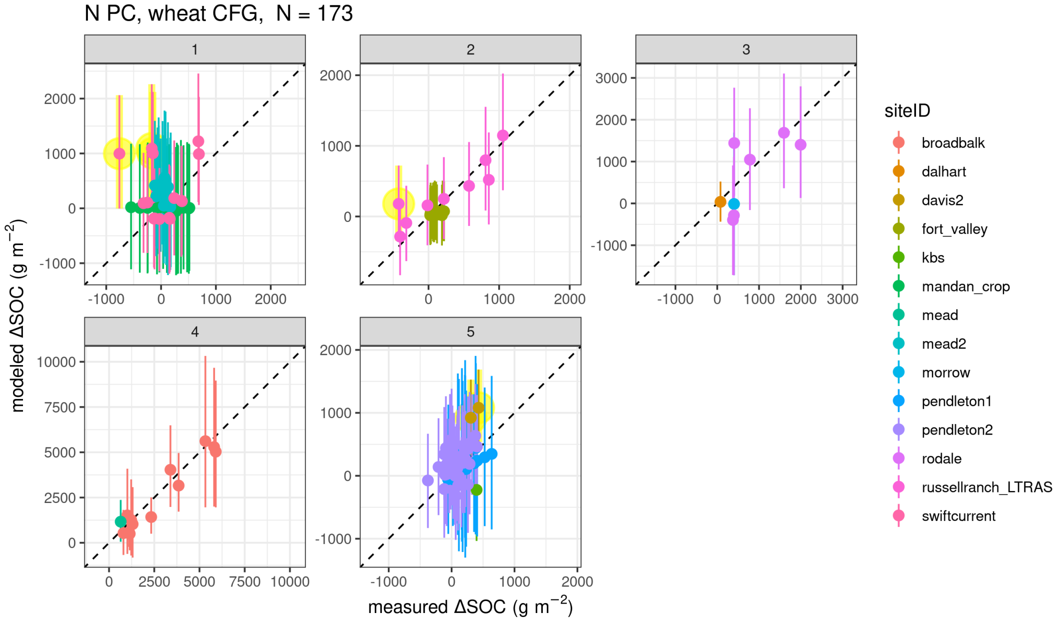

Figure 50: Model predictions versus measurements of SOC change in response to changed nitrogen practices involving crops from the wheat-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 51 for coverage rates).

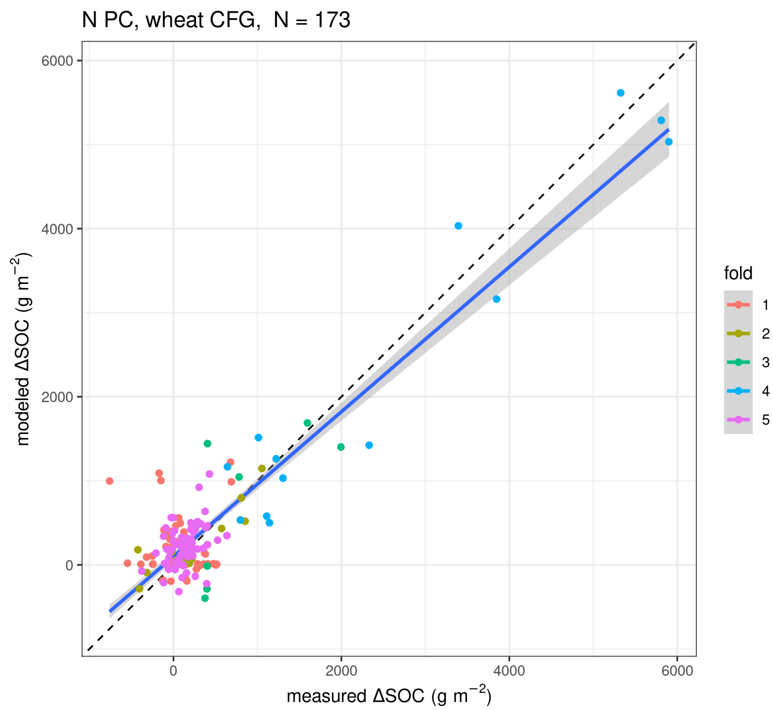

Figure 51: Scatterplot of model predictions versus measurements of SOC change in response to changed nitrogen practices involving crops from the wheat-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

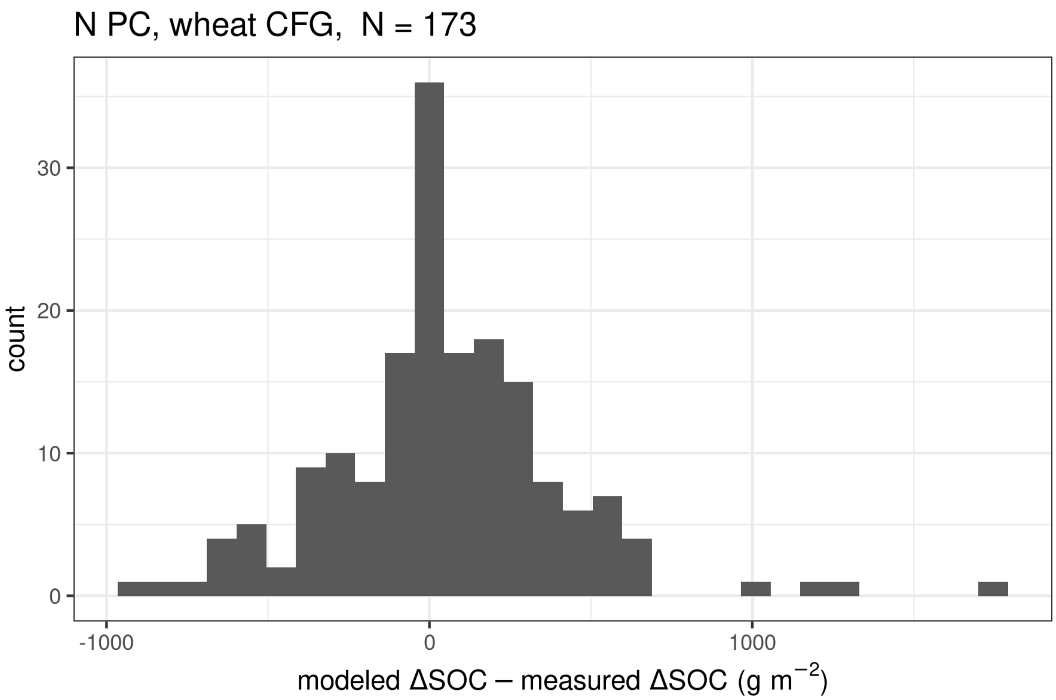

Figure 52: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, nitrogen practices involving crops from the wheat-type CFG.

Table 51: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed nitrogen practices involving wheat-type CFG.

fold	n	n in	n out	% coverage
1	51	48	3	94
2	27	26	1	96
3	8	8	0	100
4	13	13	0	100
5	74	72	2	97
All folds	173	167	6	97

Mean squared error: 601354 ± 141192; RMSE: 771 ± 86

Do 90% prediction intervals cover observed data 90% of the time? Yes

10.14 ORG x All x SOC

Follows Model Requirements section 3.3.1, paragraph 5

The uncertainty intervals presented in this section are those computed directly from the raw model calibration. In Appendix E we analyze how these coverage rates may vary over time and find evidence that these intervals may be too narrow (anti-conservative) for the first few years of ORG practice changes only. To ensure that the model uncertainty is estimated conservatively for all durations, we will apply an additional variance inflation factor of 1.36 (Appendix F) when using the model for crediting of organic amendment practices; see Section 13 "Restrictions on application of model" for details.

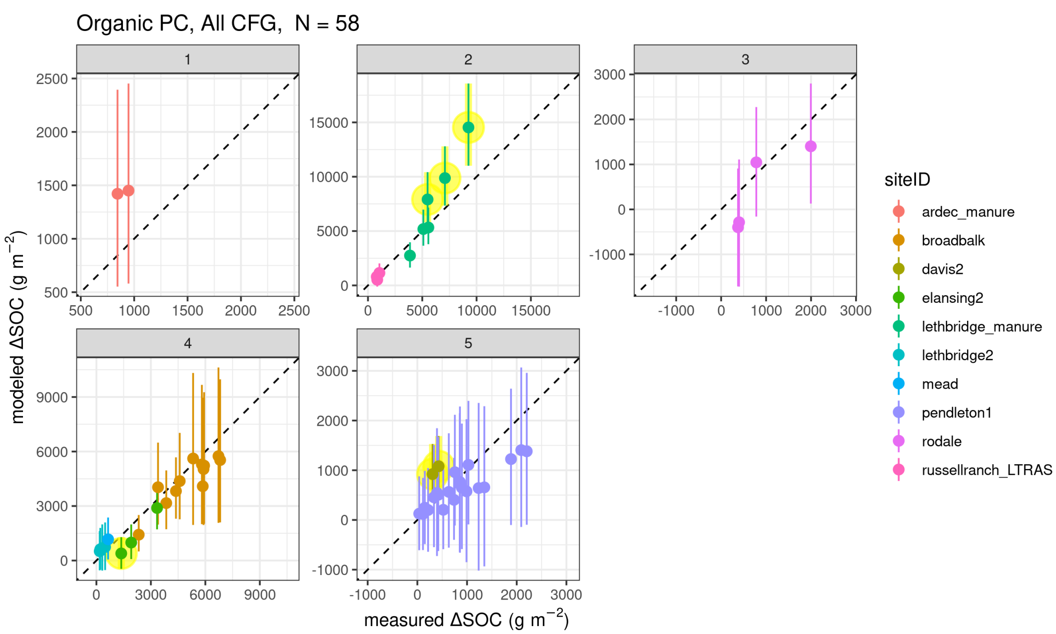

Figure 53: Model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the All-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 52 for coverage rates).

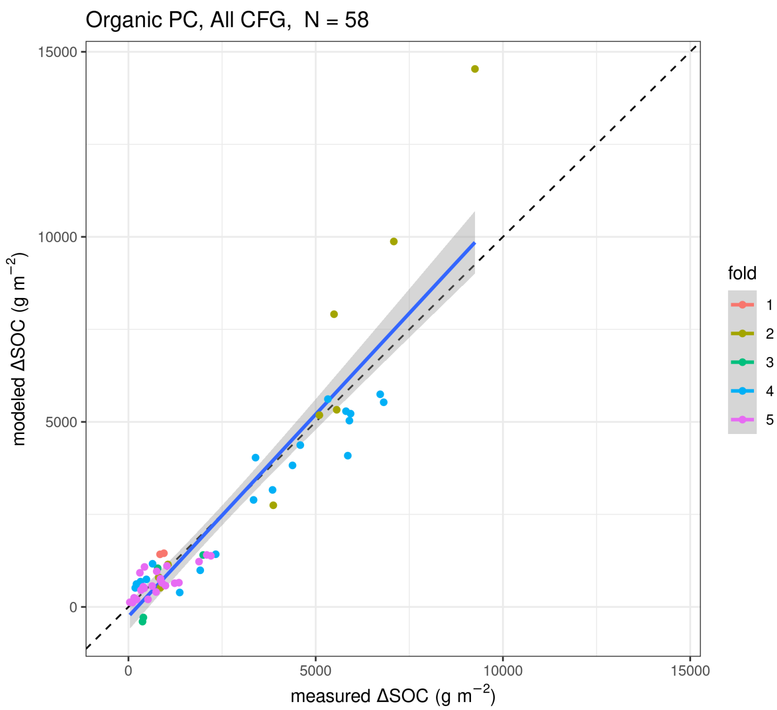

Figure 54: Scatterplot of model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the All-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

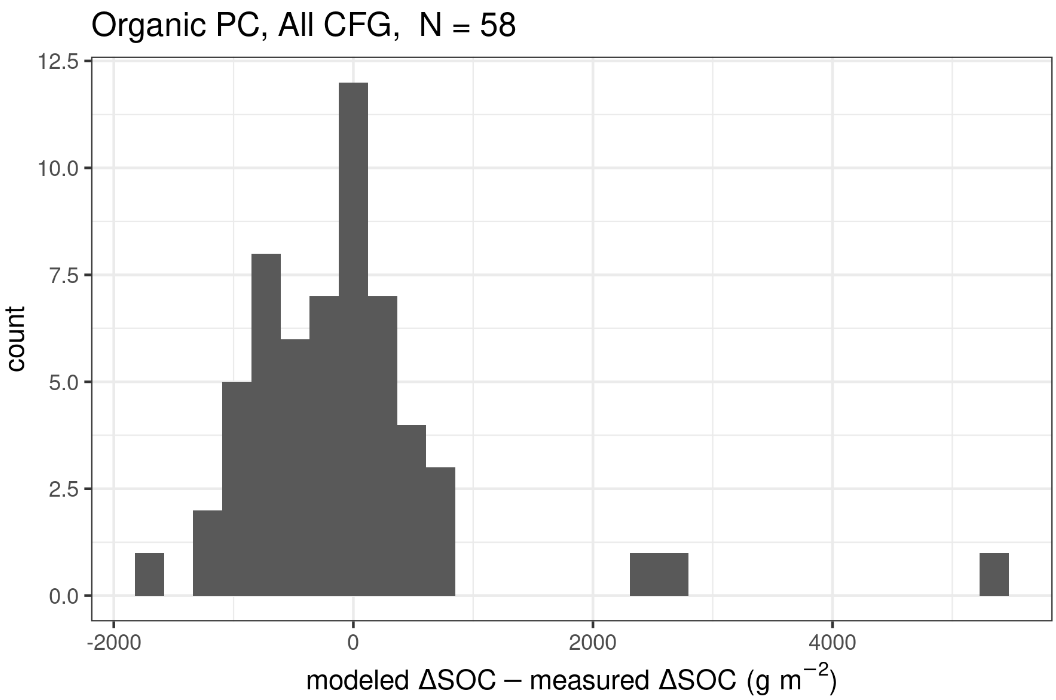

Figure 55: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, organic amendment practices involving crops from the All-type CFG.

Table 52: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed organic amendment practices involving any annual CFG.

fold	n	n in	n out	% coverage
1	2	2	0	100
2	9	6	3	67
3	4	4	0	100
4	20	19	1	95
5	23	21	2	91
All folds	58	52	6	90

Mean squared error: 2364456 ± 859092; RMSE: 1515 ± 264

Do 90% prediction intervals cover observed data 90% of the time? Yes

10.15 ORG x corn x SOC

This category is presented for context only to support the use of its data as part of Section 10.14 "ORG x All x SOC".

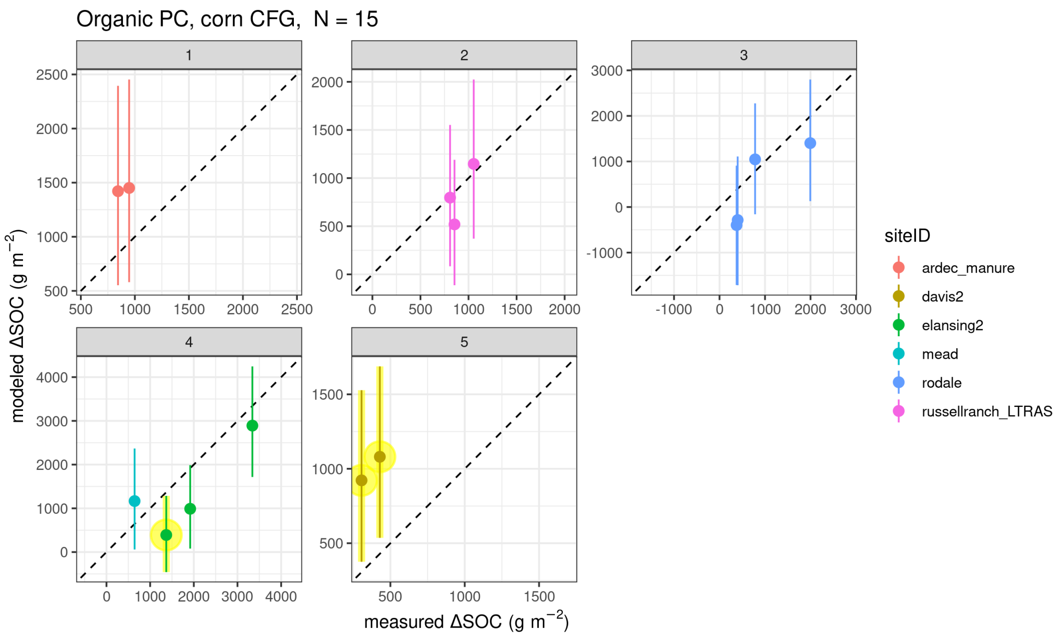

Figure 56: Model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the corn-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 53 for coverage rates).

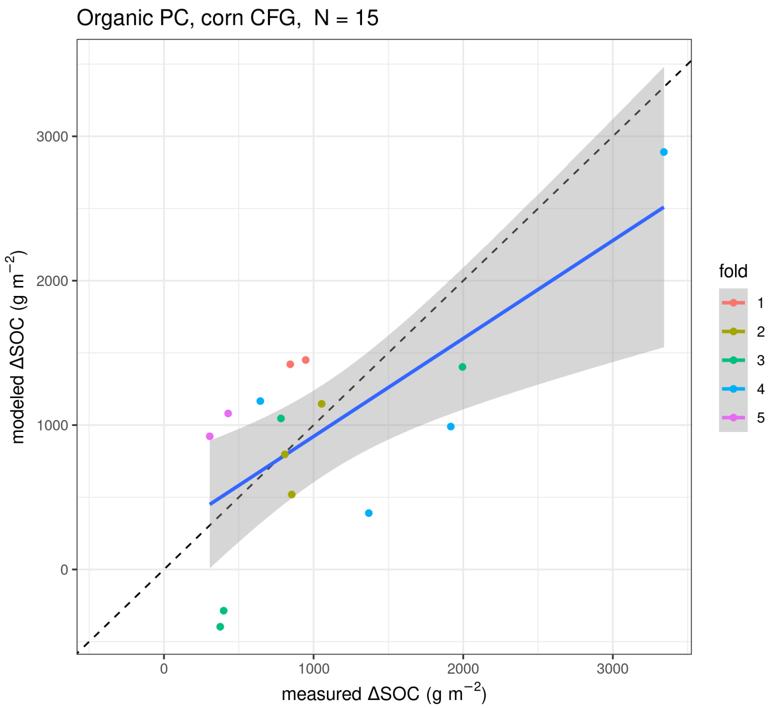

Figure 57: Scatterplot of model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the corn-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

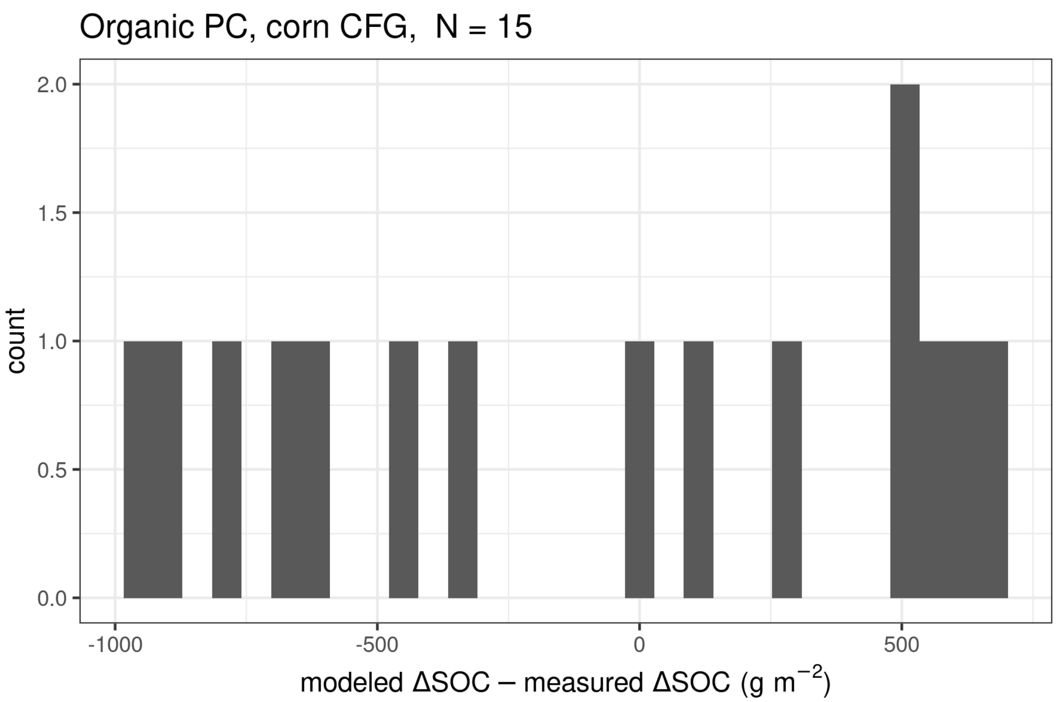

Figure 58: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, organic amendment practices involving crops from the corn-type CFG.

Table 53: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed organic amendment practices involving corn-type CFG.

fold	n	n in	n out	% coverage
1	2	2	0	100
2	3	3	0	100
3	4	4	0	100
4	4	3	1	75
5	2	0	2	0
All folds	15	12	3	80

Mean squared error: 740286 ± 293775; RMSE: 844 ± 167

Do 90% prediction intervals cover observed data 90% of the time? No

10.16 ORG x soy x SOC

This category is presented for context only to support the use of its data as part of Section 10.14 "ORG x All x SOC".

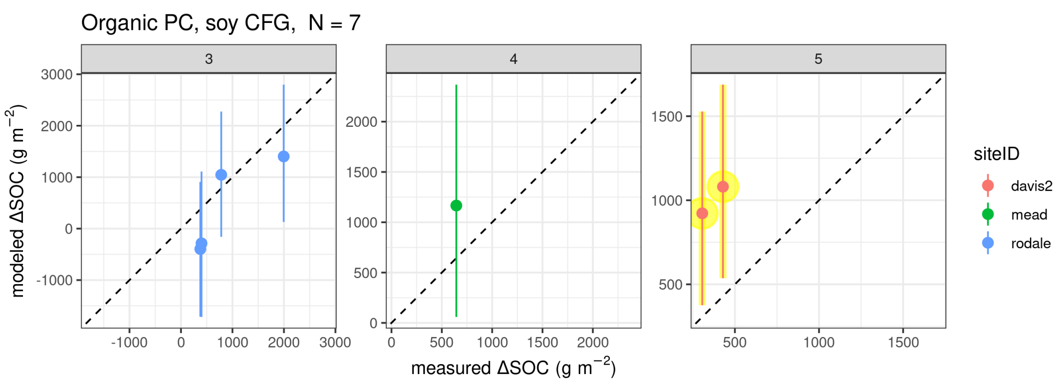

Figure 59: Model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the soy-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 54 for coverage rates).

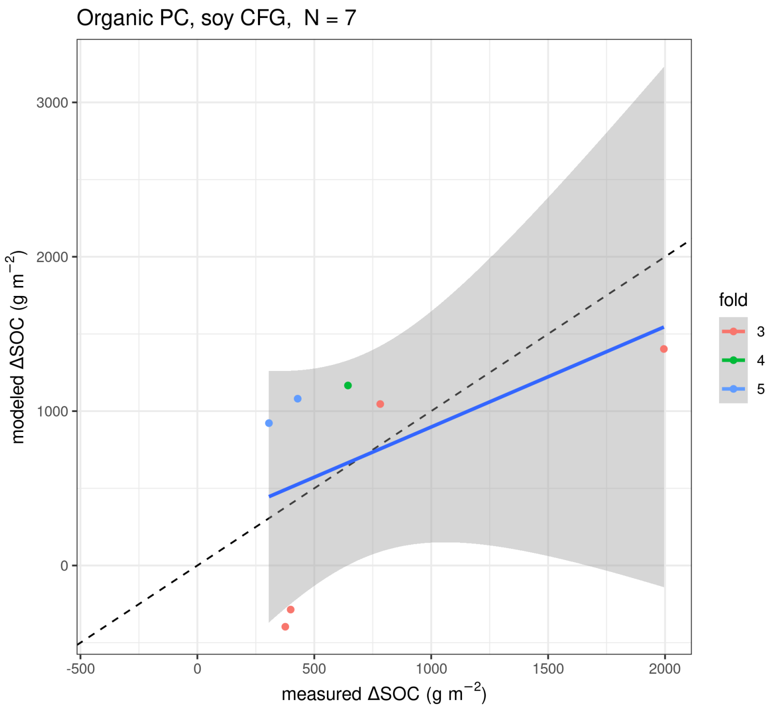

Figure 60: Scatterplot of model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the soy-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

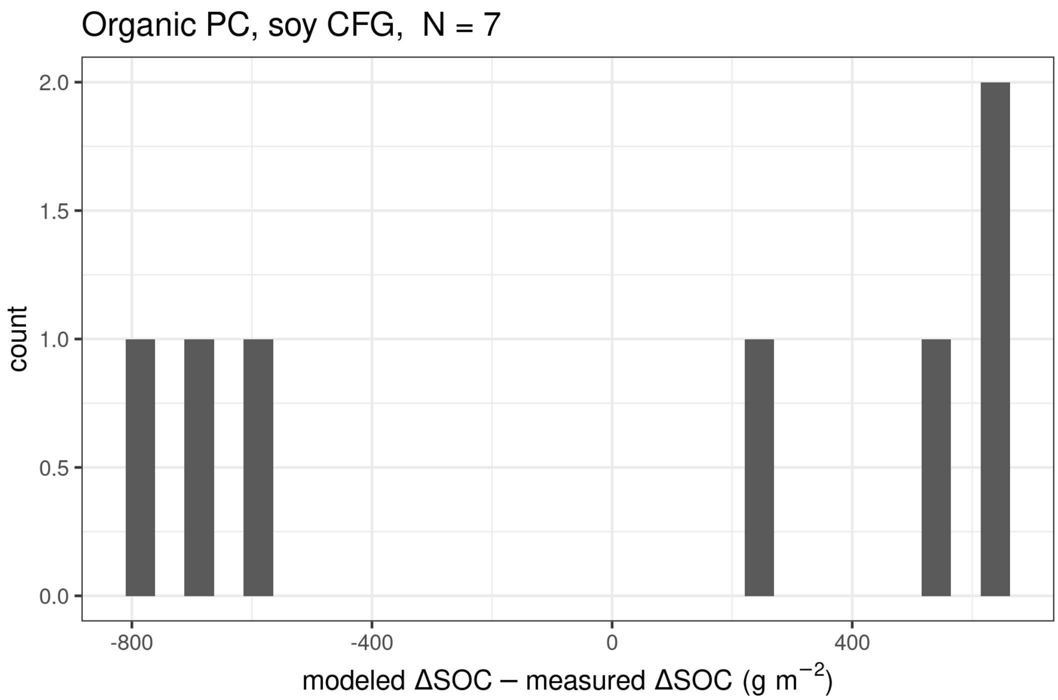

Figure 61: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, organic amendment practices involving crops from the soy-type CFG.

Table 54: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed organic amendment practices involving soy-type CFG.

fold	n	n in	n out	% coverage
3	4	4	0	100
4	1	1	0	100
5	2	0	2	0
All folds	7	5	2	71

Mean squared error: 843553 ± 475809; RMSE: 884 ± 250

Do 90% prediction intervals cover observed data 90% of the time? No

10.17 ORG x wheat x SOC

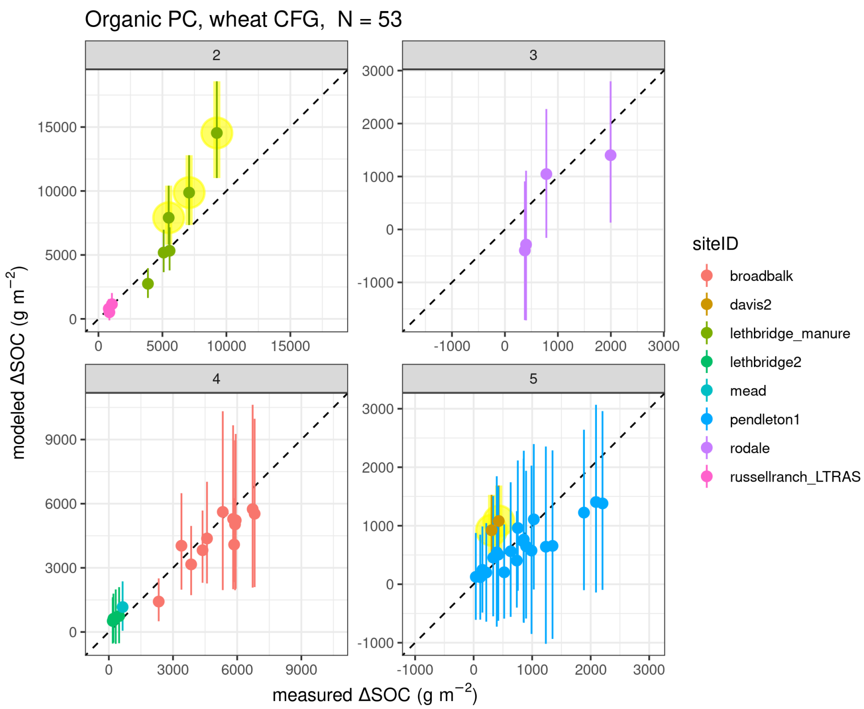

Figure 62: Model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the wheat-type CFG, faceted by fold. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value (see Table 55 for coverage rates).

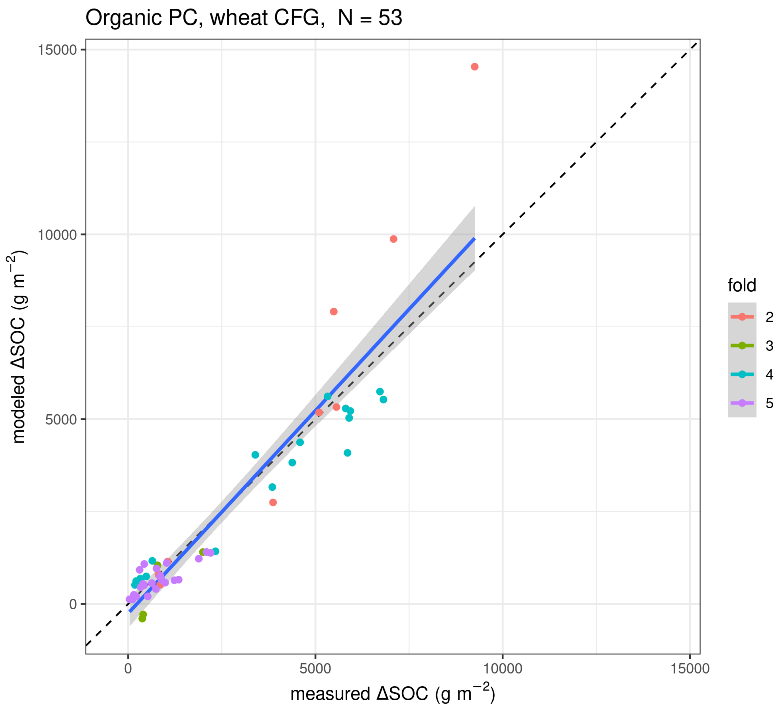

Figure 63: Scatterplot of model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the wheat-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

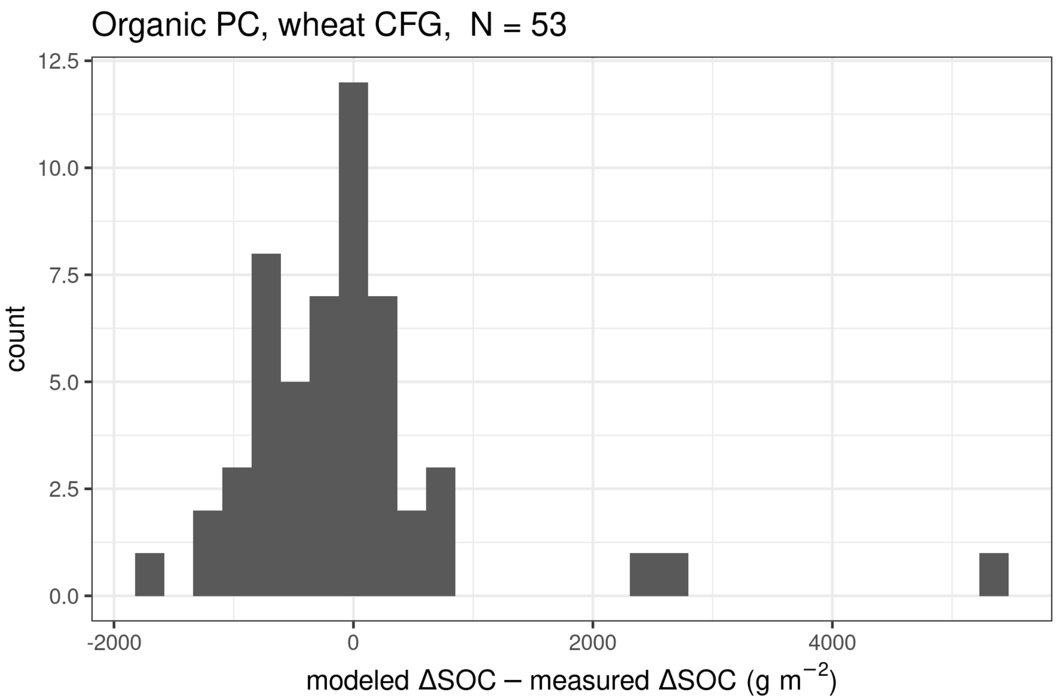

Figure 64: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed, organic amendment practices involving crops from the wheat-type CFG.

Table 55: Number of observed datapoints falling inside and outside of modeled 90% prediction intervals for SOC predictions for each fold of the calibration/validation process for changed organic amendment practices involving wheat-type CFG.

fold	n	n in	n out	% coverage
2	9	6	3	67
3	4	4	0	100
4	17	17	0	100
5	23	21	2	91
All folds	53	48	5	91

Mean squared error: 2503166 ± 938532; RMSE: 1557 ± 281

Do 90% prediction intervals cover observed data 90% of the time? Yes

11 Model validation outputs for use in SEP uncertainty calculations

Follows Model Requirements, Section 3.5

When the model is used for crediting in project CAR1459 according to SEP requirements, an uncertainty deduction will be computed using the methods described in SEP Appendix D, using the same model outputs used in this validation report. At the time of writing this report, a revision to SEP Appendix D was under consideration by CAR, so we present here an outline of what values from this report will be used under both the original and the proposed SEP requirements.

11.1 Original SEP

SEP v1.0 assumes an analytical approach to computing model prediction errors. When computing uncertainty deductions using this approach, the only quantity needed from the validation data will be $s^2_{model,\Delta G}$ (SEP Equation D.5), which can be computed as the mean squared error of out-of-sample model predictions from the cross-validation as computed across all folds, PCs and CFGs. This is the same MSE reported in Section 10.2: Mean squared error: 719393 $\pm$ 71026 (g C m^{-2})^2; RMSE: 847 $\pm$ 41 g C m^{-2} .

11.2 Revised SEP

The revision of SEP Appendix D under consideration by CAR for use in Project CAR1459 (details in Supporting Document `Appendix_D_revision_v4.docx`) introduces more complete guidelines for Monte Carlo approaches to uncertainty estimation and allow projects to use either of two approaches: analytical or Monte Carlo. These two approaches are summarized below.

11.2.1 SEP Appendix D.1: Analytical approach

The analytical approach from SEP v1.0 is retained as SEP Appendix D.1, with $s^2_{model,\Delta G}$ retained as SEP Equation D.2. If reporting under this approach, we would proceed as described in Section 11.1 "Original SEP" above.

11.2.2 SEP Appendix D.2: Monte Carlo approach

When using a Monte Carlo approach per revised SEP Appendix D.2 (Supporting Document `Appendix_D_revision_v4.docx`), model error for predicting SOC stocks in baseline and project scenarios will be computed on the natural log scale by sampling from the posterior distributions of the parameters that were adjusted during cross-validation and from variance parameters capturing prediction error (both summarized in Table A3, Appendix A). These SOC stock prediction errors will then be propagated to obtain model prediction error for emission reductions by following the procedures described in SEP Appendix D.2.

This is the same error propagation approach already used to demonstrate adequate uncertainty coverage in Section 10 "Model prediction error" of this report. 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 Monte Carlo prediction for each unique combination of parameter states in the stored posterior. The variance inflation factor described in Appendix F will also be applied to points with an ORG practice change. These are then summarized to quantify uncertainty in the estimate of total emission reductions for the project.

12 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, we took random draws of the random site and site-by-year random intercepts, which is aligned with our approach for making out-of-sample predictions; we did not use the best unbiased linear predictors (BLUPs) for the random intercepts.

The resulting posterior distributions from this final step (Table A1) are very similar to the distributions obtained during cross-validation (Figure 65; Tables 56, 57, and 58) and are saved for use when running the model for credits. We report here on the performance of the model when fitting the validation data using the final parameter set, but we emphasize that this is an evaluation against the training data and may not be representative of model performance at other sites; in particular the RMSE of the final parameter set should not be used as an estimate of model prediction error during crediting. 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.

12.1 Model bias across all PCs and CFGs

- During cross-validation: Mean bias across all studies and PC x CFG combinations: -3.87 g C m^{-2}
- With final parameter set: Mean bias across all studies and PC x CFG combinations: 4.95 g C m^{-2}

12.2 MSE and RMSE across all PCs and CFGs

- During cross-validation: Mean squared error: 719393 $\pm$ 71026 (g C $m^{-2})^2$; RMSE: 847 $\pm$ 41 g C m^{-2}
- With final parameter set: Mean squared error: 710973 $\pm$ 78614 (g C $m^{-2})^2$; RMSE: 842 $\pm$ 46 g C m^{-2}

12.3 Model fit

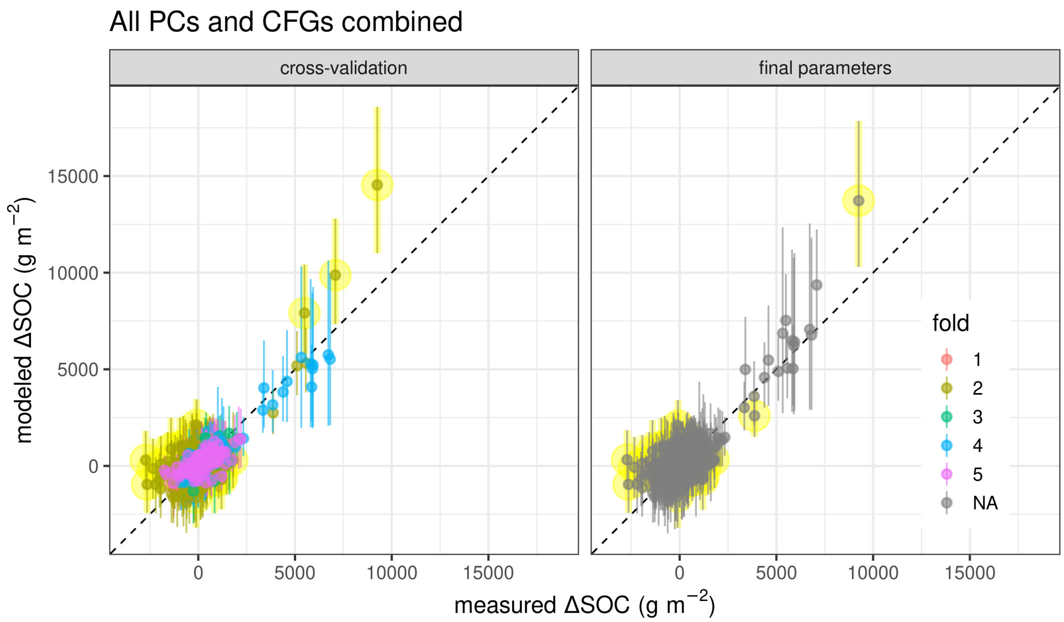

Figure 65: Model predictions versus measurements of SOC change in all practice changes and crop types, obtained during cross-validation (left) and with final parameter set (right). Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

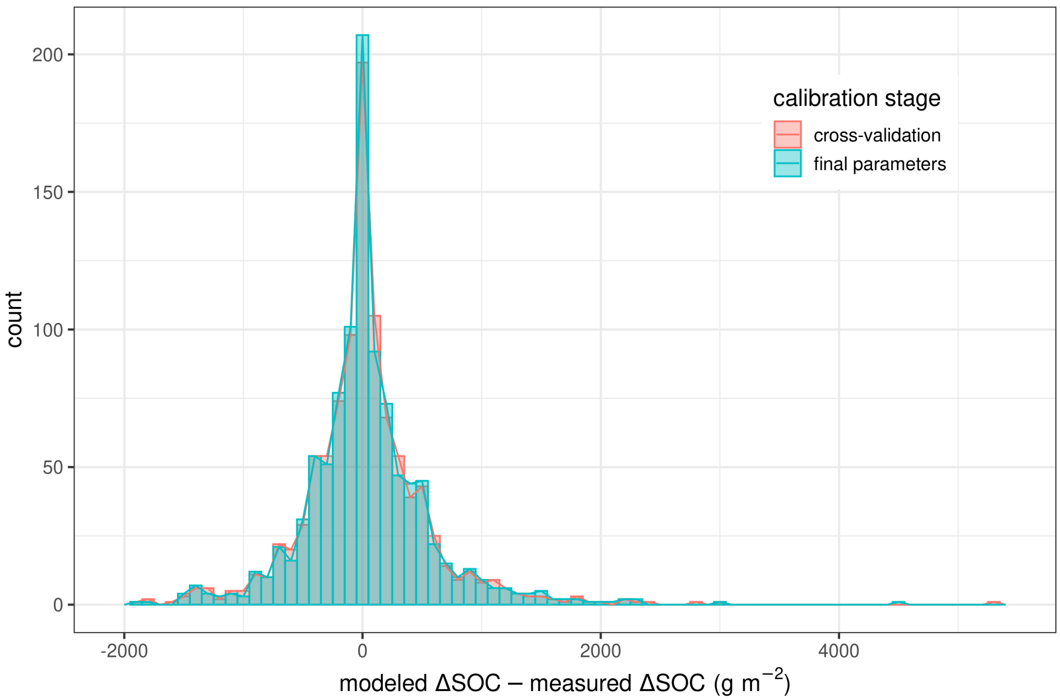

Figure 66: Histogram of model residuals (predicted - observed) for change in SOC in all studies used for model validation across all practices and crop types, obtained during cross-validation (red) and with final parameter set (blue).

Table 56: Comparison of model bias in each PC x CFG category during cross-validation and with the final parameter set.

PC	CFG	PMU	n obs	n sites	bias k-fold	bias final params	final bias smaller?
ORG	soy	529.6	7	3	236.6	250.7	No
ORG	wheat	372.8	53	8	208.9	295.1	No
ORG	All	372.8	58	10	142.6	217.2	No
ORG	corn	529.6	15	6	63.45	71.6	No
CROP	corn	399.2	210	17	8.79	7.42	Yes
NFERT	wheat	306.1	173	14	0.98	42.39	No
CROP	soy	399.2	295	20	-25.59	-29.04	Yes
CROP	wheat	342.4	326	23	-35.06	-34.4	No
NFERT	corn	868.6	166	15	-44.69	-38.99	No
CROP	cotton	240.2	162	6	-48.93	-46.03	No
DISTURB	corn	656.8	225	13	-57.95	-68.11	Yes
NFERT	soy	366.8	77	7	-60.05	-42.84	No
DISTURB	wheat	378.6	87	11	-94.13	-95.82	Yes
DISTURB	soy	467.5	66	9	-139.1	-139.4	Yes
DISTURB	cotton	209.3	49	4	-180.4	-182	Yes

Table 57: Comparison of MSE and RMSE in each PC x CFG category during cross-validation and with the final parameter set.

PC	CFG	MSE k-fold	RMSE k-fold	MSE final params	RMSE final params	final RMSE smaller?
Organic	wheat	2503166 ± 938532	1557 ± 281	2463340 ± 1070169	1538 ± 315	Yes
Organic	All	2364456 ± 859092	1515 ± 264	2321217 ± 979241	1494 ± 297	Yes
Tillage	corn	1112412 ± 158687	1052 ± 75	1127319 ± 166440	1059 ± 77	No
Organic	soy	843553 ± 475809	884 ± 250	922952 ± 547436	922 ± 270	No
Organic	corn	740286 ± 293775	844 ± 167	752258 ± 309888	850 ± 174	No
N	wheat	601354 ± 141192	771 ± 86	630835 ± 196503	786 ± 112	No
N	corn	587484 ± 99368	764 ± 64	572063 ± 95010	754 ± 62	Yes
N	soy	487636 ± 132496	692 ± 93	450418 ± 124128	665 ± 91	Yes
Tillage	soy	456619 ± 109703	671 ± 79	455322 ± 109109	670 ± 79	Yes
Cropping	corn	438595 ± 74591	660 ± 56	440885 ± 74628	662 ± 56	No
Cropping	wheat	425356 ± 58137	651 ± 44	430305 ± 57077	655 ± 43	No
Tillage	wheat	423539 ± 91906	647 ± 70	406465 ± 85538	634 ± 66	Yes
Cropping	soy	407857 ± 55371	637 ± 43	409304 ± 55233	638 ± 43	No
Cropping	cotton	215426 ± 37481	462 ± 40	218133 ± 40525	465 ± 43	No
Tillage	cotton	144049 ± 37990	376 ± 49	134505 ± 33781	364 ± 45	Yes

Table 58: Comparison of posterior 90% prediction interval coverage in each PC x CFG category during cross-validation and with the final parameter set.

PC	CFG	n k-fold	% coverage k-fold	n final params	% coverage final params
DISTURB	cotton	48 in, 1 out	98	48 in, 1 out	98
DISTURB	wheat	84 in, 3 out	97	83 in, 4 out	95
NFERT	soy	75 in, 2 out	97	75 in, 2 out	97
NFERT	wheat	167 in, 6 out	97	167 in, 6 out	97
CROP	corn	201 in, 9 out	96	202 in, 8 out	96
DISTURB	soy	63 in, 3 out	95	63 in, 3 out	95
CROP	wheat	306 in, 20 out	94	306 in, 20 out	94
NFERT	corn	156 in, 10 out	94	157 in, 9 out	95
All	All	952 in, 65 out	94	955 in, 62 out	94
CROP	soy	274 in, 21 out	93	275 in, 20 out	93
DISTURB	corn	205 in, 20 out	91	206 in, 19 out	92
ORG	wheat	48 in, 5 out	91	49 in, 4 out	92
CROP	cotton	146 in, 16 out	90	146 in, 16 out	90
ORG	All	52 in, 6 out	90	53 in, 5 out	91
ORG	corn	12 in, 3 out	80	12 in, 3 out	80
ORG	soy	5 in, 2 out	71	5 in, 2 out	71

13 Restrictions on application of model

In the previous validation report, we observed that the model underestimated uncertainty for very large changes in SOC, and therefore restricted the valid range of the model to changes smaller than 5000 g C m^{-2}. In this report, uncertainty coverage appears adequate for the full range of validation data, so this restriction is no longer needed for DayCent-CR Version 1.0.2.

The validation of model uncertainty for ORG x All x SOC (Section 10.14) passes the 90% coverage threshold specified by the Model Requirements, but our additional assessments of the heterogeneous variance approach showed evidence that the coverage rate may be time-dependent in this category (Appendix E). To ensure that model uncertainty is estimated conservatively for ORG practice changes of short duration, we apply an empirical correction for the residual variance (Appendix F) sufficient to observe coverage of at least 90% across 3, 5, and 10-year durations as well as across the entire ORG x All dataset. Therefore we **consider the model validated for ORG x All x SOC only when applied with a variance inflation factor that multiplies the calibrated residual variance by 1.36**.

References

[1] A. Agresti and B. A. Coull. "Approximate is better than "exact" for interval estimation of binomial proportions". In: *The American Statistician* 52.2 (1998), pp. 119–126.

[2] L.E. Allison. "Organic carbon". In: *Methods of soil analysis: Part 2 chemical and microbiological properties*. Vol. 9. John Wiley & Sons, Ltd, 1965. Chap. 90, pp. 1367–1378.

[3] Pierre Barré et al. "Quantifying and isolating stable soil organic carbon using long-term bare fallow experiments". In: *Biogeosciences* 7.11 (2010), pp. 3839–3850.

[4] P. Bista et al. "Simulating Soil Organic Carbon in a Wheat-Fallow System Using the Daycent Model". In: *Agronomy Journal* 108.6 (2016), pp. 2554–2565. DOI: 10.2134/agronj2016.04.0202. URL: http://doi.wiley.com/10.2134/agronj2016.04.0202.

[5] R.L. Blevins et al. "Changes in Soil Properties after 10 Years Continuous Non-Tilled and Conventionally Tilled Corn". In: *Soil and Tillage Research* 3.2 (1983), pp. 135–146. ISSN: 01671987. DOI: 10.1016/0167-1987(83)90004-1. URL: https://linkinghub.elsevier.com/retrieve/pii/0167198783900041.

[6] C.A. Campbell and R.P. Zentner. "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*. CRC Press, 1997, pp. 317–334.

[7] C.A. Campbell et al. "Quantifying Carbon Sequestration in a Minimum Tillage Crop Rotation Study in Semiarid Southwestern Saskatchewan". In: *Canadian Journal of Soil Science* 87.3 (2007), pp. 235–250. ISSN: 0008-4271, 1918-1841. DOI: 10.4141/S06-018. URL: http://www.nrcresearchpress.com/doi/10.4141/S06-018.

[8] B.P. Carlin and T.A. Louis. *Bayesian methods for data analysis*. 3rd edition. CRC Press, 2009.

[9] G. Casella. "An introduction to empirical Bayes data analysis". In: *The American Statistician* 39.2 (1985), pp. 83–87.

[10] D.R. Christenson. "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*. CRC Press, 1997, pp. 151–159.

[11] C.E Clapp et al. "Soil Organic Carbon and 13C Abundance as Related to Tillage, Crop Residue, and Nitrogen Fertilization under Continuous Corn Management in Minnesota". In: *Soil and Tillage Research* 55.3 (2000), pp. 127–142. ISSN: 01671987. DOI: 10.1016/S0167-1987(00)00110-0. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167198700001100.

[12] M.S. Clark et al. "Changes in Soil Chemical Properties Resulting from Organic and Low-Input Farming Practices". In: *Agronomy Journal* 90.5 (1998), pp. 662–671. ISSN: 0002-1962, 1435-0645. DOI: 10.2134/agronj1998.00021962009000050016x. URL: https://onlinelibrary.wiley.com/doi/abs/10.2134/agronj1998.00021962009000050016x.

[13] H.P. Collins et al. "Soil Carbon Dynamics in Corn-Based Agroecosystems: Results from Carbon-13 Natural Abundance". In: *Soil Science Society of America Journal* 63.3 (1999), pp. 584–591. ISSN: 03615995. DOI: 10.2136/sssaj1999.03615995006300030022x. URL: http://doi.wiley.com/10.2136/sssaj1999.03615995006300030022x.

[14] A.C. Davison and D.V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, 1997.

[15] S.J. Del Grosso et al. "DayCent model simulations for estimating soil carbon dynamics and greenhouse gas fluxes from agricultural production systems". In: *Managing agricultural greenhouse gases: coordinated agricultural research through GRACEnet to address our changing climate* (2012), pp. 241–250.

[16] S.J. Del Grosso et al. "DAYCENT national-scale simulations of nitrous oxide emissions from cropped soils in the United States". In: *Journal of Environmental quality* 35.4 (2006), pp. 1451–1460.

[17] K. Denef et al. "Does Long-Term Center-Pivot Irrigation Increase Soil Carbon Stocks in Semi-Arid Agro-Ecosystems?" In: *Geoderma* 145.1 (2008), pp. 121–129. ISSN: 00167061. DOI: 10.1016/j.geoderma.2008.03.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S0016706108000748.

[18] W.A. Dick, W.M. Edwards, and E.L. McCoy. "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*. CRC Press, 1997, pp. 171–182.

[19] M.S. Dolan et al. "Soil Organic Carbon and Nitrogen in a Minnesota Soil as Related to Tillage, Residue and Nitrogen Management". In: *Soil and Tillage Research* 89.2 (2006), pp. 221–231. ISSN: 01671987. DOI: 10.1016/j.still.2005.07.015. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167198705002187 (visited on 04/28/2021).

[20] E.T. Elliott et al. "Terrestrial Carbon Pools: Preliminary Data from the Corn Belt and Great Plains Regions". In: *SSSA Special Publications*. Madison, WI, USA: Soil Science Society of America and American Society of Agronomy, 1994, pp. 179–191. ISBN: 978-0-89118-930-5 978-0-89118-807-0. DOI: 10.2136/sssaspecpub35.c12. URL: http://doi.wiley.com/10.2136/sssaspecpub35.c12.

[21] A.C.B Ferreira et al. "Soil carbon accumulation in cotton production systems in the Brazilian Cerrado". In: *Acta Scientiarum. Agronomy* 42 (2019).

[22] A. Gelman and D.B. Rubin. "Inference from iterative simulation using multiple sequences". In: *Statistical science* 7.4 (1992), pp. 457–472.

[23] Andrew Gelman et al. *Bayesian data analysis*. 3rd. CRC press, 2014.

[24] R. Ghimire, S. Machado, and P. Bista. "Soil pH, Soil Organic Matter, and Crop Yields in Winter Wheat-Summer Fallow Systems". In: *Agronomy Journal* 109.2 (2017), pp. 706–717. ISSN: 00021962. DOI: 10.2134/agronj2016.08.0462. URL: http://doi.wiley.com/10.2134/agronj2016.08.0462.

[25] Rajan Ghimire, Stephen Machado, and Karl Rhinhart. "Long-Term Crop Residue and Nitrogen Management Effects on Soil Profile Carbon and Nitrogen in Wheat-Fallow Systems". In: *Agronomy Journal* 107.6 (2015), pp. 2230–2240. ISSN: 00021962. DOI: 10.2134/agronj14.0601. URL: http://doi.wiley.com/10.2134/agronj14.0601.

[26] Benedikt Graler, Edzer Pebesma, and Gerard Heuvelink. "Spatio-Temporal Interpolation using gstat". In: *The R Journal* 8 (1 2016), pp. 204–218. URL: https://journal.r-project.org/archive/2016/RJ-2016-014/index.html.

[27] J. Guillaume and F. Andrews. *dream: DiffeRential Evolution Adaptive Metropolis*. R package version 0.4-2. 2012. URL: http://CRAN.R-project.org/package=dream.

[28] R.B. Gurung et al. “Bayesian calibration of the DayCent ecosystem model to simulate soil organic carbon dynamics and reduce model uncertainty”. In: *Geoderma* 376 (2020), p. 114529.

[29] A.D. Halvorson and C.P. Jantalia. “Nitrogen Fertilization Effects on Irrigated No-Till Corn Production and Soil Carbon and Nitrogen”. In: *Agronomy Journal* 103.5 (2011), pp. 1423–1431. ISSN: 00021962. DOI: 10.2134/agronj2011.0102. URL: http://doi.wiley.com/10.2134/agronj2011.0102.

[30] A.D. Halvorson and A.J. Schlegel. “Crop rotation effect on soil carbon and nitrogen stocks under limited irrigation”. In: *Agronomy Journal* 104.5 (2012), pp. 1265–1273.

[31] A.D. Halvorson, C.E. Stewart, and S.J. Del Grosso. “Manure and Inorganic Nitrogen Affect Irrigated Corn Yields and Soil Properties”. In: *Agronomy Journal* 108.2 (2016), pp. 519–531. ISSN: 00021962. DOI: 10.2134/agronj2015.0402. URL: http://doi.wiley.com/10.2134/agronj2015.0402.

[32] A.D. Halvorson, B.J. Wienhold, and A.L. Black. “Tillage, Nitrogen, and Cropping System Effects on Soil Carbon Sequestration”. In: *Soil Science Society of America Journal* 66.3 (2002), pp. 906–912. ISSN: 03615995. DOI: 10.2136/sssaj2002.9060. URL: http://doi.wiley.com/10.2136/sssaj2002.9060.

[33] A.D. Halvorson et al. “Soil organic carbon and nitrogen sequestration in irrigated cropping systems of the Central Great Plains”. In: *Soil carbon sequestration and the greenhouse effect* 57 (2009), pp. 141–157.

[34] X. Hao et al. “Soil Carbon and Nitrogen Response to 25 Annual Cattle Manure Applications”. In: *Journal of Plant Nutrition and Soil Science* 166.2 (2003), pp. 239–245. ISSN: 14368730, 15222624. DOI: 10.1002/jpln.200390035. URL: http://doi.wiley.com/10.1002/jpln.200390035.

[35] N.R. Hulugalle et al. “Soil organic carbon concentrations and storage in irrigated cotton cropping systems sown on permanent beds in a Vertosol with restricted subsoil drainage”. In: *Crop and Pasture Science* 64.8 (2013), pp. 799–805.

[36] IPCC. *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*. Intergovernmental Panel on Climate Change, Switzerland, 2019. ISBN: 978-4-88788-232-4.

[37] I. Ismail, R.L. Blevins, and W.W. Frye. “Long-Term No-tillage Effects on Soil Properties and Continuous Corn Yields”. In: *Soil Science Society of America Journal* 58.1 (1994), pp. 193–198. ISSN: 03615995. DOI: 10.2136/sssaj1994.03615995005800010028x. URL: http://doi.wiley.com/10.2136/sssaj1994.03615995005800010028x.

[38] H.H Janzen et al. “Soil organic matter dynamics in long-term experiments in Southern Alberta”. In: *Soil Organic Matter in Temperate Agroecosystems: Long-Term Experiments in North America*. CRC Press, 1997, pp. 283–296.

[39] M.K. Jarecki and R. Lal. “SOIL ORGANIC CARBON SEQUESTRATION RATES IN TWO LONG-TERM NO-TILL EXPERIMENTS IN OHIO”. In: *Soil Science* 170.4 (2005), pp. 280–291. ISSN: 0038-075X. DOI: 10.1097/00010694-200504000-00005. URL: http://journals.lww.com/00010694-200504000-00005.

[40] V. Jin and G. Varvel. *All; 1998–2016. ver. GPSR_NATRES. Fort Collins, CO: USDA-ARS Natural Resources Database. Project: NEMERREM*. 2018. URL: https://gpsr.ars.usda.gov/NEMERREM.

[41] V. Jin and G. Varvel. *All; 2001–2016. ver. GPSR_NATRES. Fort Collins, CO: USDA-ARS Natural Resources Database. Project: NEMEIRR*. 2018. URL: https://gpsr.ars.usda.gov/NEMEIRR.

[42] V.L. Jin et al. “Twelve Years of Stover Removal Increases Soil Erosion Potential without Impacting Yield”. In: *Soil Science Society of America Journal* 79.4 (2015), pp. 1169–1178. ISSN: 03615995. DOI: 10.2136/sssaj2015.02.0053. URL: http://doi.wiley.com/10.2136/sssaj2015.02.0053.

[43] S.A. Khan et al. “The Myth of Nitrogen Fertilization for Soil Carbon Sequestration”. In: *Journal of Environmental Quality* 36.6 (2007), pp. 1821–1832. ISSN: 00472425. DOI: 10.2134/jeq2007.0099. URL: http://doi.wiley.com/10.2134/jeq2007.0099.

[44] A.Y.Y. Kong et al. “The Relationship between Carbon Input, Aggregation, and Soil Organic Carbon Stabilization in Sustainable Cropping Systems”. In: *Soil Science Society of America Journal* 69.4 (2005), pp. 1078–1085. ISSN: 03615995. DOI: 10.2136/sssaj2004.0215. URL: http://doi.wiley.com/10.2136/sssaj2004.0215.

[45] M. Kuhn and K. Johnson. “Over-fitting and model tuning”. In: *Applied predictive modeling*. Springer, 2013, pp. 61–92.

[46] L.J. Mebius. “A rapid method for the determination of organic carbon in soil”. In: *Analytica Chimica Acta* 22 (1960), pp. 120–124.

[47] S.A. Mestelan. “Impact of Long-Term No till and Plow till on Soil Properties and Soil Nutrient Cycling”. PhD thesis. Ohio State University, 2008. 150 pp. URL: http://rave.ohiolink.edu/etdc/view?acc_num=osu1199221756.

[48] M.M. Mikha, M.F. Vigil, and J.G. Benjamin. “Long-Term Tillage Impacts on Soil Aggregation and Carbon Dynamics under Wheat-Fallow in the Central Great Plains”. In: *Soil Science Society of America Journal* 77.2 (2013), pp. 594–605. ISSN: 03615995. DOI: 10.2136/sssaj2012.0125. URL: http://doi.wiley.com/10.2136/sssaj2012.0125.

[49] J.P. Mitchell et al. “Cover Cropping and No-Tillage Improve Soil Health in an Arid Irrigated Cropping System in California’s San Joaquin Valley, USA”. In: *Soil and Tillage Research* 165 (2017), pp. 325–335. ISSN: 01671987. DOI: 10.1016/j.still.2016.09.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S0167198716301854.

[50] J.P. Mitchell et al. “Tillage and Cover Cropping Affect Crop Yields and Soil Carbon in the San Joaquin Valley, California”. In: *Agronomy Journal* 107.2 (2015), pp. 588–596. ISSN: 00021962. DOI: 10.2134/agronj14.0415. URL: http://doi.wiley.com/10.2134/agronj14.0415.

[51] C.M. Monreal and H.H. Janzen. “Soil Organic-Carbon Dynamics after 80 Years of Cropping a Dark Brown Chernozem”. In: *Canadian Journal of Soil Science* 73.1 (1993), pp. 133–136. ISSN: 0008-4271, 1918-1841. DOI: 10.4141/cjss93-014. URL: http://www.nrcresearchpress.com/doi/10.4141/cjss93-014.

[52] K.R. Olson, S.A. Ebelhar, and J.M. Lang. “Cover crop effects on crop yields and soil organic carbon content”. In: *Soil Science* 175.2 (2010), pp. 89–98.

[53] S. Osborne and R. Lehman. *All; 2000-2015. ver. GPSR_NATRES. Fort Collins, CO. USDA-ARS Natural Resources Database. Project: SDBRREAP*. 2018. URL: https://gpsr.ars.usda.gov/SDBRREAP.

[54] W.J. Parton et al. “Generalized model for NO_x and N_2O emissions from soils”. In: *Journal of Geophysical Research: Atmospheres* 106.D15 (2001), pp. 17403–17419.

[55] D. Pimentel et al. “Environmental, Energetic, and Economic Comparisons of Organic and Conventional Farming Systems”. In: *BioScience* 55.7 (2005), p. 573.

[56] J.C. Pinheiro and D.M. Bates. "Linear mixed-effects models: basic concepts and examples". In: *Mixed-effects models in S and S-Plus* (2000), pp. 3–56.

[57] J.C. Pinheiro et al. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-155. 2022. URL: https://CRAN.R-project.org/package=nlme.

[58] M. Plummer et al. "CODA: Convergence Diagnosis and Output Analysis for MCMC". In: *R News* 6.1 (2006), pp. 7–11. URL: https://journal.r-project.org/archive/.

[59] W. Powers et al. "Quantifying greenhouse gas sources and sinks in animal production systems". In: *Quantifying Greenhouse Gas Fluxes in Agriculture and Forestry: Methods for Entity-Scale Inventory*. Ed. by M. Eve et al. USDA Technical Bulletin 1939. U.S. Department of Agriculture, 2014.

[60] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2021. URL: https://www.R-project.org/.

[61] P.E. Rasmussen and R.W. Smiley. "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*. CRC Press, 1997, pp. 353–360.

[62] Rothamsted Research. *Dataset: Broadbalk soil organic carbon content 1843-2010*. 2014. URL: http://www.era.rothamsted.ac.uk/dataset/rbk1/01-BKSOC1843/.

[63] D.R. Roberts et al. "Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure". In: *Ecography* 40.8 (2017), pp. 913–929.

[64] I.J. Rochester. "Sequestering carbon in minimum-tilled clay soils used for irrigated cotton and grain production". In: *Soil and tillage research* 112.1 (2011), pp. 1–7.

[65] B. A. Rosner. *Fundamentals of Biostatistics*. 6th. 2006.

[66] U.M. Sainju, W.F. Whitehead, and B.P. Singh. "Carbon accumulation in cotton, sorghum, and underlying soil as influenced by tillage, cover crops, and nitrogen fertilization". In: *Plant and Soil* 273.1 (2005), pp. 219–234.

[67] M.R. Schmer et al. "Tillage and Residue Management Effects on Soil Carbon and Nitrogen Under Irrigated Continuous Corn". In: *Soil Science Society of America Journal* 78.6 (2014), pp. 1987–1996. ISSN: 03615995. DOI: 10.2136/sssaj2014.04.0166. URL: http://doi.wiley.com/10.2136/sssaj2014.04.0166.

[68] N. Senapati et al. "Modelling soil organic carbon storage with RothC in irrigated Vertisols under cotton cropping systems in the sub-tropics". In: *Soil and Tillage Research* 143 (2014), pp. 38–49.

[69] I.M. Sobol. "Sensitivity estimates for nonlinear mathematical models". In: *MMCE* 4 (1993), pp. 407–414.

[70] Soil Survey Staff. *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*. 2022. URL: https://gdg.sc.egov.usda.gov/.

[71] Kellog Biological Station. *Main Cropping System Experiment (MCSE) downloadable tables 164 and 308*. 2021. URL: https://lter.kbs.msu.edu/datatables/.

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

[73] G. E. Varvel. “Soil Organic Carbon Changes in Diversified Rotations of the Western Corn Belt”. In: *Soil Science Society of America Journal* 70.2 (2006), pp. 426–433. ISSN: 03615995. DOI: 10.2136/sssaj2005.0100. URL: http://doi.wiley.com/10.2136/sssaj2005.0100.

[74] J.J. Veenstra et al. “Conservation Tillage and Cover Cropping Influence Soil Properties in San Joaquin Valley Cotton-Tomato Crop”. In: *California Agriculture* 60.3 (2006), pp. 146–153. ISSN: 0008-0845. DOI: 10.3733/ca.v060n03p146. URL: http://californiaagriculture.ucanr.org/landingpage.cfm?articleid=ca.v060n03p146.

[75] M.L. Vitosh, J.F. Davis, and B.D. Knezek. “Long-Term Effects of Manure, Fertilizer, and Plow Depth on Chemical Properties of Soils and Nutrient Movement in a Monoculture Corn System”. In: *Journal of Environmental Quality* 2.2 (1973), pp. 296–299. ISSN: 00472425. DOI: 10.2134/jeq1973.00472425000200020029x. URL: http://doi.wiley.com/10.2134/jeq1973.00472425000200020029x.

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

[77] J.A. Vrugt and C.J.F. Ter Braak. “DREAM$_{(D)}$: an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems”. In: *Hydrology and Earth System Sciences* 15.12 (2011), pp. 3701–3713.

[78] J.A Vrugt et al. “Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling”. In: *International journal of nonlinear sciences and numerical simulation* 10.3 (2009), pp. 273–290.

[79] B.R. Wegner et al. “Response of soil surface greenhouse gas fluxes to crop residue removal and cover crops under a corn-soybean rotation”. In: *Journal of environmental quality* 47.5 (2018), pp. 1146–1154.

[80] Y. Zhang, M. Arabi, and K. Paustian. “Analysis of parameter uncertainty in model simulations of irrigated and rainfed agroecosystems”. In: *Environmental Modelling & Software* (2020).

[81] Y.A. Zhang, A. Suyker, and K. Paustian. “Improved crop canopy and water balance dynamics for agroecosystem modeling using DayCent”. In: *Agronomy Journal* 110.2 (2018), pp. 511–524.