Validation Report for DayCent-CR version 1.0

Soil Metrics, LLC
Indigo Ag

Reviewed by Brian McConkey, Viresco Solutions & Michael Dietze, Boston University

First submission: February 12, 2021
Revised to address reviewer comments: March 22, 2021; July 13, 2021; September 21, 2021; October 29. 2021; November 3, 2021

Contents

Report type

Report Type

Type 1 (Project-specific, for project CAR1459)

SEP version

Version 1.0, accessed on 3 Dec 2020

SEP Model Requirements version

Requirements and Guidance for Model Calibration, Validation, Uncertainty, and Verification For Soil Enrichment Projects Version 1.0a, accessed on 3 Dec 2020 (referred to hereafter as the "Model Requirements")

For this report we incorporate the following changes from version 1.0a of the Model Requirements. The following proposed changes in requirements were submitted to CAR, reviewed with CAR and 1 – 2 external experts, and written guidance was provided by CAR approving these for use in Indigo US Project #1, first monitoring report:

1. Background irrigation exemption (Appendix A): allows cropping systems using irrigation as a background practice to not require validation of the Water Management/Irrigation PC on the condition that at least one study in the validation dataset uses irrigation as a management practice. The range of precipitation regimes included in the validation dataset (173-1233 mm yr^{-1}), covering at least 3 LRRs, are considered an adequate proxy for testing the effects of artificial rainfall;
2. Validation of N-fixers x Fertilizer (Appendix B): allows the Soy CFG to be considered validated for the N Fertilizer PC on the condition of N Fertilizer being validated for another annual CFG and Soy being validated for the Cropping, Planting and Harvesting PC;
3. Use of stacked practices in validation (Appendix C): allows validation datasets containing "stacked" practices to be used more than once per validation report, on the

condition that each PC x CFG combination has at least one isolated treatment comparison;

4. Validation of organic amendment applications (Appendix D): allows any project CFGs to be aggregated together when validating the organic amendment PC, rather than requiring that each be validated individually.

In addition, the following proposal was submitted to CAR as of July 29, 2021 and is currently under review with CAR and external experts. It is included in this report to address the gap in detailed Model Requirements when statistical methods are used to demonstrate independence of calibration and validation data.

5. Cross-validation methods for model calibration and validation (Appendix H): includes specific reporting requirements when using cross-validation methods.

Copies of the above proposals were filed with CAR as Appendices for this report, as listed above. For more information about the proposals and their review process, contact CAR directly.

Model version

DayCent-CR version 1.0.

This model version consists of the following components (collectively the "model files"). Each of these components are version-controlled independently from each other, but only the following component versions shall be considered the validated DayCent-CR version 1.0:

6. 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.
7. Version 1.0 build 1.0 of the DayCent-CR model parameters, corresponding to Git commit 237eddd1f176ea7eeb4a5a8fc0a3712623800a85 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.

Note that all components currently have the same version number, but this will not be true in future versions. A future validation report might hypothetically describe a "DayCent-CR version 1.5" by declaring that it consists of "DayCent-CR executable version 1.1 build 2.0" together with "DayCent-CR parameters version 2.6 build 88.9"

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 or validation and is not included in the model version described here. The validation

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

described here should be applicable to any result obtained from DayCent-CR 1.0 whether it is run directly or accessed through any technically compatible version of the DayCent-CR API.

Version confirmation materials

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

- Copies of the model executable and all model run files used in the simulations for this report (in their initial state, prior to model calibration and validation); these have been submitted to CAR as `DayCent-CR _1.0_Simulation_files.zip`
- Validation dataset (`CAR1459_model_val_DayCentCR1.0_dataset.xlsx`)
- Calibrated model predictions (`CAR1459_model_val_DayCentCR_1.0_model_ests.csv`)
- Diagnostic plots of MCMC sampler performance (`CAR1459_model_val_DayCentCR_1.0_sampler_diagnostics.docx`)
- Appendix E – documentation of calibrated parameter sets
- Appendix F - declaration of practices and crops
- Appendix G – errata.

The executable and parameter sets were renamed to the above versioning to support final submission of this report. All code and files loaded into a new private GitLab repository for dedicated tracking are separate from the model API, which may have independent version updates that do not change the validated model files.

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 US Project #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. Parton et al. 2001, DelGrosso et al. 2006, 2012, Zhang et al. 2018) has been used extensively for more than two decades by researchers worldwide to simulation 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 as for estimating 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 (Eve et al. 2014) and the model is implemented as part of the Climate Action

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

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 (version 1.0, build 1.0.) 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 Model setup section 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.

Responsible parties

Calibration, validation, and running of DayCent-CR for this project were all performed by Soil Metrics, LLC. As required in Section 5 of the SEP Model Requirements, Soil Metrics LLC has the requisite expertise to calibrate and validate DayCent-CR for model performance and uncertainty, as approved by CAR on February 12, 2021 (https://soilmetrics.eco/our-team/). Y. Zhang performed modification to the model version to use direct soil measurements to initialize the model and implemented the Bayesian calibration and cross-validation 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. J. Soong and K. Paustian supervised the project and provided review of all model development and data collection.

Indigo Ag is the project developer of CAR1459. C. Black provided assistance to Soil Metrics in aligning available datasets for model calibration and validation with SEP Model Requirements, performed data analysis of validation datasets and model outputs, and led the writing of this report. R. Gurung performed data analysis of model output and contributed to the writing of the report. As a former contractor to Soil Metrics LLC, R. Gurung also provided technical guidance on Bayesian calibration of DayCent-CR model. M. Motew helped with interpretation of the SEP Model Requirements criteria and provided review of the report. N. Campbell was the project supervisor and provided review to all Validation Report materials.

Model Calibration

<u>Follows Model Requirements Section 2 Summary of Requirements (p8)</u>

Description of Model Calibration

DayCent-CR model 1.0, build 1.0 was calibrated using an empirical Bayesian method that employs the DiffeRential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt 2016, Vrugt & Ter Braak, 2011), which follows a Markov Chain Monte Carlo (MCMC) approach and simultaneously updates model parameters using a likelihood-maximizing optimization algorithm. It has advantages over traditional MCMC algorithms due to its efficiency in mitigating issues with high-dimensionality, multimodality, nonlinearity, and local optima with proved ergodicity. The method has been previously implemented in DayCent by Zhang et al (2020) to calibrate crop growth/production. To calibrate DayCent-CR for modeling SOC stock and stock change, optimization was performed over the likelihood function proposed by Gurung et al. (2020). 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 as follows:

$$\mathrm{L}(\mathrm{O}|\theta) = (2\pi)^{-n/2}|\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(\mathrm{O}-\mathrm{M})^{\mathrm{T}}\Sigma^{-1}(\mathrm{O}-\mathrm{M})\right\},$$

where θ is a given set of model parameters, n is the number of observations, Σ is the variance-covariance matrix, and O and M are natural log transformed vectors of observed and modeled SOC values. The variance-covariance matrix partitions model error into three random effects: variance between experimental sites σ^2_{site}, variance between years within sites $\sigma^2_{site\times year}$, and unexplained residual variance σ^2. These hyperparameters were estimated empirically by fitting the model residuals from each MCMC iteration using a linear mixed effect model with two levels of random effects, of the form (log(O)-log(M)) ~ N(0, σ^2_{site} + $\sigma^2_{site\times year}$ + σ^2) (Pinhiero and Bates, 2000; see supplement to Gurung et al. 2020 for additional discussion of this approach). These models were fitted in the R package `lme4` using the `lmer` function as part of the likelihood evaluation for each MCMC iteration and were then used after calibration for estimating model prediction uncertainty.

The calibration of DayCent-CR was implemented in R (R Core Team, 2018) software using the DREAM package (Guillaume and Andrews, 2012) and the DREAM algorithm is provided in detail by Vrugt et. Al., (2009). The calibration was run implementing 11 independent MCMC

chains, all of which were run until the $\widehat{R}$-statistic of Gelman and Rubin (1992) dropped to a value of 1.2 (450-720 iterations), suggesting a convergence of the posterior distribution of model parameters (Vrugt and Ter Braak, 2011). The first 50% of each chain was discarded as the "burn-in" period (Vrugt and Ter Braak, 2011) and the remaining 50% of simulations were used to summarize the marginal posterior probability of model parameters. Traceplots and parameter shrinkage factors are provided in supporting document CAR1459_DayCentCR_1.0_sampler_diagnostics.docx.

Calibration and validation of the model were conducted simultaneously using a k-fold cross-validation procedure with a choice of 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 "Justification for splitting of experimental data" for details). Splitting the dataset by experiment-IDs in order to construct each fold thus resulted in independent calibration and validation datasets by meeting the condition that "datasets used for calibration and validation do not overlap in experimental research locations and are not taken from the same experimental study" (Model Requirements section 2, page 4). The fold configuration used in this report is provided in Table 48 and a map of study sites in Figure 1.
2. Second, for each fold (fold = 1, 2, …, 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 and hyperparameters estimated from the calibration data for that fold.
4. Fourth, Monte Carlo simulations of the validation dataset were run, with each simulation step invoking DayCent-CR on sites reserved for validation using model parameters from the final 10% of samples from the joint posterior distribution that was calculated in step three from sites used for calibration. The resulting range of model predictions (245-396 samples, depending on fold) was then used to estimate the posterior predictive distribution of SOC differences between the experimental treatments, similarly 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 these each 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 the final 20 draws from each MCMC chain for a total of 220 joint posterior samples. See section "Model validation outputs for use in SEP uncertainty calculations" for a description of how the saved posteriors are used during crediting.

The prior distributions of parameters adjusted during the calibration process (Appendix E, Table 46) and summary statistics of marginal posterior distributions of model parameters (Appendix E, Table 47) 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 `DayCent_CR_1.0_Simulation_files.zip`, filed with this report.

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. To the authors' knowledge, the approach described here is a first-of-its-kind application in the field of validating soil biogeochemical modeling with use of Bayesian calibration. 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. This approach therefore departs from the requirement that "parameter sets used when validating the model are the same used when the model is applied to simulate baselines and project practices" (Model Requirements, page 4 bullet 1). This wording does not seem to anticipate approaches that use multiple cross-validation steps to produce multiple sets of parameters during the validation process, which suggests a gap in the Model Requirements. This gap is discussed in detail and revised Requirements are proposed in Appendix H: Proposal for Cross-Validation Methods for Model Calibration and Validation. We propose for this report that the validated parameter sets should be considered *equivalent to* the set used for crediting, and we further 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). Additionally, we completed a comparison between the parameter distributions obtained from cross-validation and from fitting the full dataset (Figure 55), as well as between the distributions of model outputs (Figure 54 and section "Evaluation of final parameter set") to ensure differences between the validated and final parameterizations, particularly for the most sensitive parameters, are not such as would materially change model results.

Model setup

For calibration and validation, we ran DayCent-CR for all treatments and sites (See site-level summaries in Table 4 through Table 11, and full dataset in data file `CAR1459_model_val_DayCentCR_1.0_dataset.xlsx`). 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 "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 2020), management information estimated from typical agronomic practice in the region (see sections "Management information" and "Procedures for missing data" for details), and climate data

(minimum and maximum daily temperature, precipitation) from the PRISM database (PRISM Climate Group) for the experiments located in the United States, and nearest weather station 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).

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 (USEPA, 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 "Description of Model Calibration" below).

10 of the 33 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: The period between experiment start and first SOC measurement was treated as part of the historic period, then the model SOC was initialized to match the first SOC measurement as described above, then the period between first SOC measurement and experiment end 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 "Description of data requirements", below). 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 will meet SEP requirements and be described in the CAR1459 Monitoring Plan.

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 (USEPA, 2020), and not all model parameters have an impact on SOC dynamics. Therefore we selected 28 parameters to consider in the calibration exercise (Table 46; Appendix E). 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 Soil Metrics 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 the modeled SOC, but in an indirect manner mediated by the selected parameters, and were left as constants assigned values according to defaults used in COMET-Farm and the US GHG Inventory.

Prior distributions for all 28 parameters were each assigned independent uniform distributions defined by lower and upper bounds (Table 46). The initial list of parameters and their prior range was similar to those reported in Gurung et al. (2020). The priors were developed 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 these previous studies.

A variance based Global Sensitivity Analysis (GSA) was performed on the calibration dataset 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 significant 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 its prior to uncertainty in model outputs—particularly the likelihood. Similarly 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 (Gurung et al., 2020). The total sensitivity indices for each of the 28 parameters are plotted in Figure 15 of the Supporting Document `CAR1459_model_val_DayCentCR_1.0_sampler_diagnostics.docx`. From this analysis, we identified as most influential a final set of all 10 parameters (Table 46) that contributed more than 0.5% of variance. This very inclusive cutoff was chosen to reduce the dataset dependence of

the GSA by including more parameters that might be influential in the general case even though they contribute little variance to the present dataset.

Bayesian calibration was performed on these 10 most influential parameters, whereas the rest of the parameters were fixed to their default values. The Bayesian calibration also included estimates of three additional hyperparameters (σ^2_{site}, $\sigma^2_{site \times year}$, and σ^2) in the joint posterior distribution (Table 47) associated with model uncertainty addressing the spatiotemporal correlation presented in the calibration dataset (Gurung et al., 2020). The prior ranges for all model parameters used for both GSA and Bayesian calibration, along with descriptions of each parameter, are provided in Table 46 (Appendix E). The summary statistics of the marginal posterior distribution are provided in Table 47 (Appendix E) and Figure 55. Sampler diagnostics, marginal posterior plots, and parameter correlation plots are available in supporting file `CAR1459_model_val_DayCentCR_1.0_sampler_diagnostics.docx`. Only the parameters adjusted during the calibration process are shown in Appendix E; for the full parameter set and auxiliary files needed to reproduce the validation, please see supporting file `DayCent_CR_1.0_Simulation_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 estimated them using the joint posterior distribution. By using this joint posterior as the parameter input to 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.

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 these 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 simultaneously using a 5-fold cross-validation method. Cross-validation retains statistical independence of calibration and validation data (see Appendix H: Proposal for Cross-Validation Methods for Model Calibration and Validation for detail) 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 is limited.

To retain independence while dividing the available datasets 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 not correlated 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 semivariograms of initial SOC stock, measured and modeled SOC change, site effects assigned during calibration, and model residuals after calibration (Figures 1 & 2 in supporting file `CAR1459_model_val_DayCentCR_1.0_sampler_diagnostics.docx`). All plots showed very low overall correlation between sites and no evidence of higher correlation between sites geographically nearer each other, supporting our decision to treat sites as if they were independent.

Project Domain

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

Practice Categories

The project intends to credit 14 practices (filed with CAR in Appendix F, Table 49) falling into four Practice Categories (PCs):

a. "Inorganic N fertilizer application" (NFERT),
b. "Organic amendments application" (ORG),
c. "Soil disturbance and/or residue management" (DISTURB),
d. "Cropping practices" (CROP)

Crop Functional Groups

The project includes crops spanning three crop functional groups (CFGs): "Corn-Group", "Soy-Group", and "Wheat-group". Details about crops included in CFGs are filed with CAR in Appendix F.

Land resource regions

The project includes fields in 13 LRRs:

F: Northern Great Plains
G: Western Great Plains
H: Central Great Plains
I: Southwest Plateaus
J: Southwestern Prairies
K: Northern Lake States
L: Lake States
M: Central Feed Grains
N: East and Central Farming
O: Mississippi Delta
P: South Atlantic
R: Northeastern Forage
T: Atlantic and Gulf

These sites span four IPCC climate zones: Cool temperate dry, cool temperate moist, warm temperate dry, and warm temperate moist.

Soils

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

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%)

Emissions Sources

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

Domain covered by this validation

The domain validated in this report includes a total of 9 combinations of CFG, PC, and emissions source (ES), as summarized in Table 1: Combinations of CFG and PC that are validated for SOC in this project**.** Additionally, we consider NFERT x Soy x SOC to be valid per the proposal in Appendix B requiring the successful validation of CROP x Soy x SOC and NFERT x Any Non-Soy CFG x SOC, with the Non-Soy CFG in this case being both corn and wheat. Further, following the proposal in Appendix D allowing multiple project CFGs to be aggregated when validating the ORG PC, we present a combined dataset for the ORG PC from all crops combined. Data for ORG x Corn x SOC and ORG x Soy x SOC are provided for context only, as both had insufficient data to pass the geographic distribution requirements on their own.

Table 1: Combinations of CFG and PC that are validated for SOC in this project

	Corn	**Soy**	**Wheat**
NFERT	x	Via Appendix B	x
ORG	Info only	Info only	With cap on dSOC (see page 105)
DISTURB	x	x	x
CROP	x	x	x

Table 2: Biophysical attribute ranges across which each PC/CFG was validated for SOC, meeting minimum requirements outlined in Requirement 2 of Section 3.3. All PC/CFG categories pass the "stacking" requirement (Requirement 1 of Section 3.3, amended according to written guidance from the Registry allowing stacked practice studies to be used more than once per report – see Appendix C) by containing at least one study that isolates the effect of the PC change being validated. See data declaration table for each PC x CFG combination in "Documentation of Validation and Calibration Datasets, per CFG-PC-ES combination" for counts of stacked and unstacked observations.

PC/CFG	**N sites**	**N obs**	**LRRs (bold = in project)**	**IPCC zones***	**Soil Textures****	**Clay range**
CROP/Corn	12	113	C, **H, L, M**, S	CTM,CTD, WTM, WTD	Lo, SiCl, SiClLo, SiLo	32%
CROP/soy	13	109	C, **H, L, M**, S +Canada	CTM,CTD, WTM, WTD	Lo, SiCl, SiClLo, SiLo	32%
CROP/wheat	16	121	C, **H, L, M**, S +Canada	CTM,CTD, WTM, WTD	ClLo, Lo, SaClLo, SiLo, SiCl, SiClLo	32%
DISTURB/Corn	10	86	**K, L, M, N**	CTM,CTD, WTM, WTD	Lo, SiLo, SiClLo	25%
DISTURB/Soy	5	30	**L, M, N**	CTM,WTM	Lo, SiLo, SiClLo	25%
DISTURB/Wheat	9	58	B, C, **F, G, H, L, M**	CTM,CTD, WTD	ClLo, Lo, SiLo, SiClLo	25%

NFERT/Corn	10	80	C, E, **K, L, M, N**, S	CTM,CTD, WTM, WTD	ClLo, Lo, SiClLo, SiLo	17%
NFERT/Wheat	12	102	B, C, **F, L, M**, S +Canada +England	CTM,CTD, WTM, WTD	Lo, SiClLo, SiLo	18%
ORG/Wheat	6	39	B, C, **M** +Canada +England	CTM,CTD, WTD	ClLo, Lo, SiClLo, SiLo	22%
ORG/Corn (info only)	6	13	C, E, **L, M,** S	CTM, CTD, WTM, WTD	ClLo, Lo, LoSa, SiLo, SiClLo	25
Org/Soy (info only)	3	5	C, **M**, S	WTM, WTD	Lo, SiLo, SiClLo	18
ORG all crops	10	49	B, C, **E, L, M**, S +Canada +England	CTM, CTD, WTM, WTD	ClLo, Lo, LoSa, SiLo, SiClLo	29

* IPCC zones: WTM = Warm Temperate Moist, CTM = Cool Temperate Moist, WTM = Warm Temperate Moist, WTD = Warm Temperate Dry. All zones present in the validation dataset are also present in the declared project domain.
** Soil Textures: ClLo = clay loam, Lo = loam, LoSa = loamy sand, SaClLo = sandy clay loam, SiClLo = silty clay loam, SiL = silt loam, SiCl = silty clay. All soil textures present in the validation dataset are also present in the declared project domain.

Description of data requirements

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

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

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

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)

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.

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 (USEPA, 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 "Description of data requirements" were reported, or could be inferred according to the procedures reported above in section "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 2020 were considered for inclusion.

For this validation and calibration, data was evaluated from 135 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 'Description of Data Requirements', or missing data could be inferred according to procedures reported above in '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 spreadsheet `CAR1459_model_val_DayCentCR_1.0_dataset.xlsx` 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 one of the four included in the Project's domain (Section 3 of this report).

After this evaluation process, 33 sites (Figure 1) were identified as usable for the calibration and validation process, collectively containing 220 treatments and 486 measurements that could be combined into 495 pairs of observed practice-change effects. Seven of these observation pairs were from PC x CFG combinations not validated in this report (six NFERT x Soy, one H2O x

Wheat), and these were included in the calibration and validation runs (therefore allowing the final parameter set to be informed by these observations), but are not reported here.

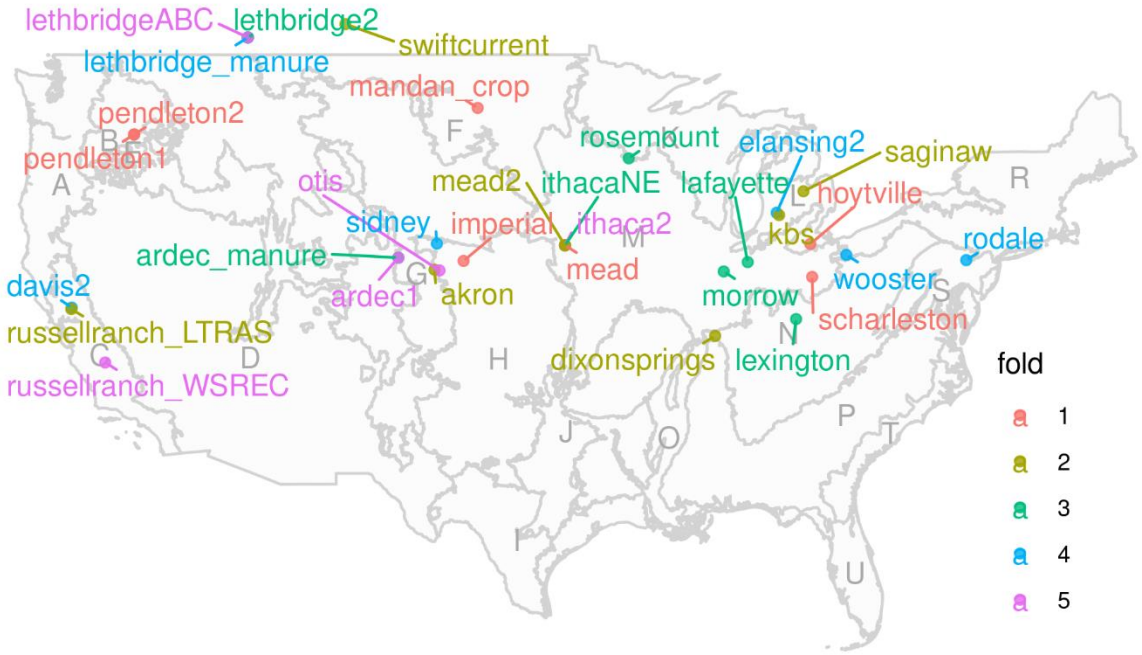

Figure 1: Locations of experimental sites used for calibration and validation of DayCent-CR. Land Resource Regions are shown in grey. Colors indicate which fold of the cross-validation held this site out for validation. Not shown:Broadbalk, England (fold 5).

Where information from multiple publications was combined for a single validation point, all publications used are included in the citation list for that site (Table 4 through Table 11). 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 and Appendix C.

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 individual 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.

Documentation of Validation and Calibration Datasets, per CFG-PC-ES combination

Follows Model Requirements Section 3.3 Summary of Requirements (p14)

CROP x Corn x SOC

This category's validation is usable in all project LRRs and soil textures (see Section 3) because:
• The selected studies span five LRRs (C, H, L, M, S), three of which (H, L, M) are in the declared project domain.
• Four soil textures are included, all of which are included in the declared project domain: loam, silt loam, silty clay, silty clay loam.
• Clay content spans 32 percentage points, from a low of 15% to a high of 47%..
• At least one study isolates effects, i.e. only 15 of 110 observations compare stacks of PC changes.

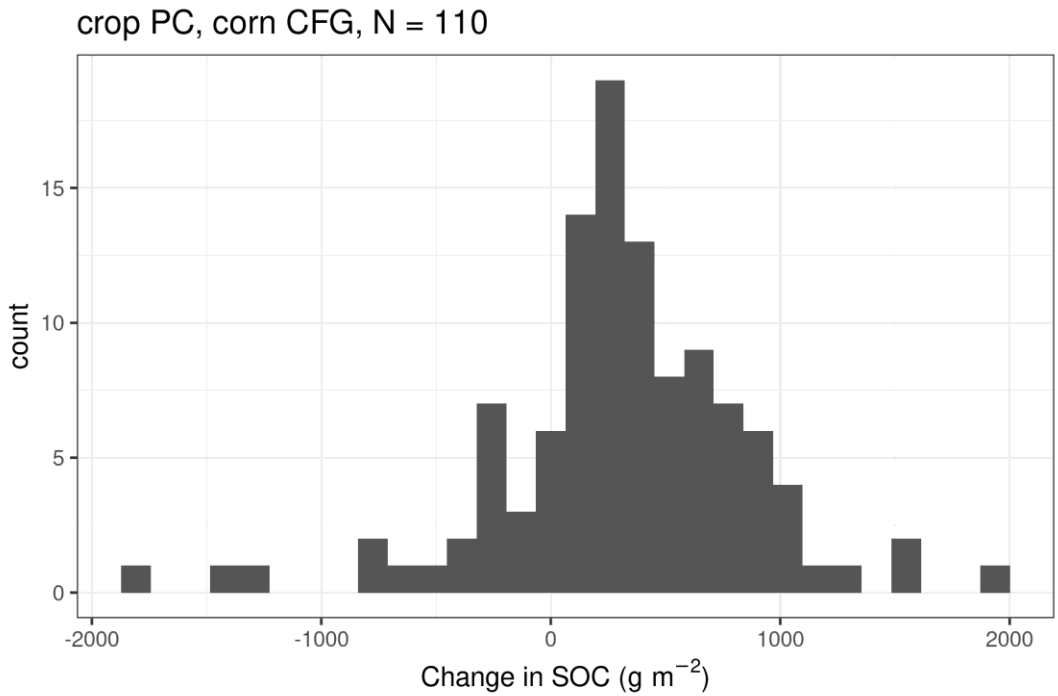

Figure 2: Histogram of changes in SOC observed by the studies used for model validation in response to changed cropping practices involving crops from the corn-type CFG.

Table 3: 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
davis2	Clark et al. 1998	Davis, CA	1988	1996	C	warm temperate dry	loam	17	Walkley-Black method	5 (4 stack PCs)
hoytville	Collins et al. 1999	Hoytville, OH	1963	1993	L	cool temperate moist	silty clay loam	40	dry combustion	3
imperial	Denef et al. 2008	Imperial, NE	1970	2012	H	cool temperate dry	loam	24	dry combustion	3 (3 stack PCs)
kbs	Collins et al 2000	Hickory Corners KBS, MI	1993	2001	L	cool temperate moist	loam	19	dry combustion	1 (1 stack PCs)
mead	Elliott et al. 1994	Mead, NE	1975	1992	M	warm temperate dry	silty clay loam	35	dry combustion	1
mead2	Varvel 2006	Mead, NE	1982	1992; 1998; 2002	M	warm temperate dry	silty clay loam	31	dry combustion	54
morrow	Khan et al. 2007	Champaign -Urbana, IL	1955	2005	M	warm temperate moist	silt loam	25	dichromate oxidation technique ofMebius (1960)	4
otis	Denef et al. 2008	Otis, CO	1966	2012	H	cool temperate dry	loam	26	dry combustion	3 (3 stack PCs)
rodale	Elliott et al. 1994; Pimental et al. 2005	Kutztown, PA	1981	1992; 2002	S	warm temperate moist	silt loam	30	not reported	4 (4 stack PCs)
russellranch _LTRAS	Kong et al. 2005	Winter, CA	1993	1997; 2003; 2012	C	warm temperate dry	silt loam	18	dry combustion	6
saginaw	Christenson 1997	Saginaw, MI	1972	1981; 1991	L	cool temperate moist	silty clay	47	dry combustion	10

wooster	Collins et al. 1999; Dick et al. 1997	Wooster, OH	1962	1971; 1980; 1992	M	cool temperate moist	silt loam	15		16

CROP x Soy x SOC

This category's validation is usable in all project LRRs and soil textures (see Section 3) because:

- The selected studies span five LRRs (C, H, L, M, S), three of which (H, L, M) are in the declared project domain. Additionally, one site is used from outside the US in a declared project climate zone (cool temperate dry).
- Four soil textures are included, all of which are included in the declared project domain: loam, silt loam, silty clay, silty clay loam
- Clay content spans 32 percentage points, from a low of 15% to a high of 47%.
- At least one study isolates effects, i.e. only 15 of 106 observations compare stacked PC changes.

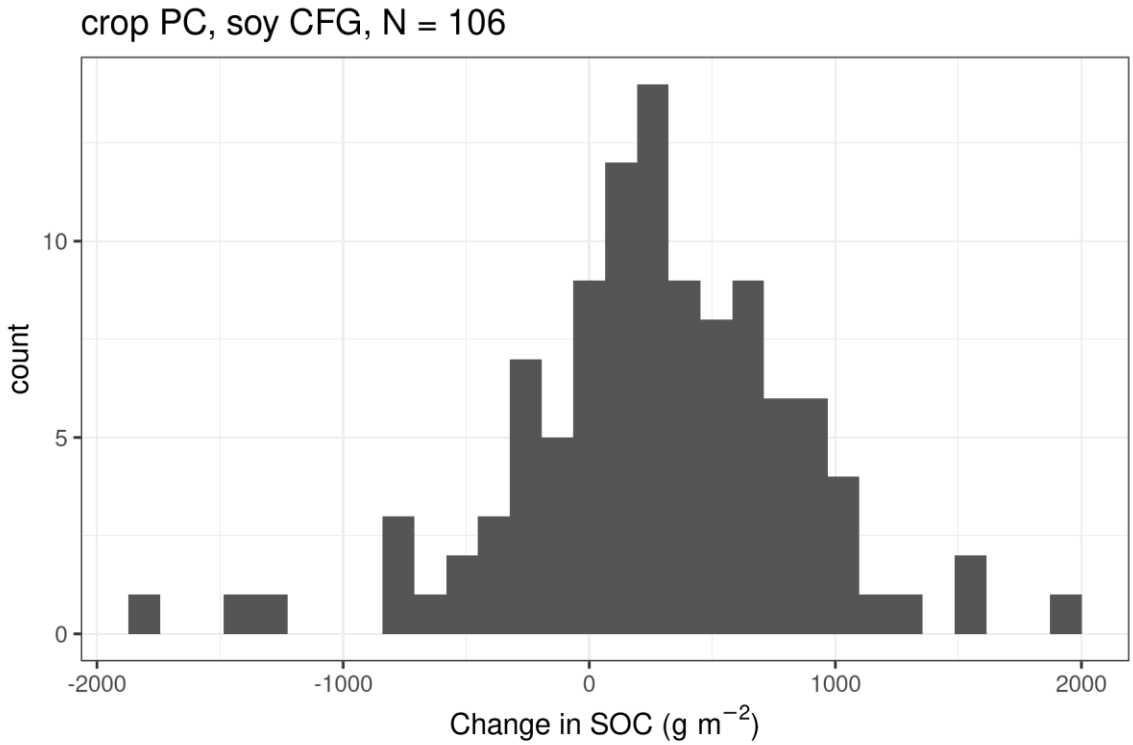

Figure 3: Histogram of changes in SOC observed by the studies used for model validation in response to changed cropping practices involving crops from the soy-type CFG.

Table 4: 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
davis2	Clark et al. 1998	Davis, CA	1988	1996	C	warm temperate dry	loam	17	Walkley-Black method	5 (4 stack PCs)
hoytville	Collins et al. 1999	Hoytville, OH	1963	1993	L	cool temperate moist	silty clay loam	40	dry combustion	3
imperial	Denef et al. 2008	Imperial, NE	1970	2012	H	cool temperate dry	loam	24	dry combustion	3 (3 stack PCs)
kbs	Collins et al 2000	Hickory Corners KBS, MI	1993	2001	L	cool temperate moist	loam	19	dry combustion	1 (1 stack PCs)
mead	Elliott et al. 1994	Mead, NE	1975	1992	M	warm temperate dry	silty clay loam	35	dry combustion	1
mead2	Varvel 2006	Mead, NE	1982	1992; 1998; 2002	M	warm temperate dry	silty clay loam	31	dry combustion	45
morrow	Khan et al. 2007	Champaign-Urbana, IL	1955	2005	M	warm temperate moist	silt loam	25	dichromate oxidation technique ofMebius (1960)	4
otis	Denef et al. 2008	Otis, CO	1966	2012	H	cool temperate dry	loam	26	dry combustion	3 (3 stack PCs)
rodale	Elliott et al. 1994; Pimental et al. 2005	Kutztown, PA	1981	1992; 2002	S	warm temperate moist	silt loam	30	not reported	4 (4 stack PCs)
russellranch_LTRAS	Kong et al. 2005	Winter, CA	1993	1997; 2003; 2012	C	warm temperate dry	silt loam	18	dry combustion	8
saginaw	Christenson 1997	Saginaw, MI	1972	1981; 1991	L	cool temperate moist	silty clay	47	dry combustion	8

swiftcurrent	Campbell and Zentner 1997; Campbell et al. 2007	Swift Current, SK, Canada	1966	1981; 1984; 1990; 1993; 1996		cool temperate dry	silt loam	20	dry combustion	5
wooster	Collins et al. 1999; Dick et al. 1997	Wooster, OH	1962	1971; 1980; 1992	M	cool temperate moist	silt loam	15		16

CROP x Wheat x SOC

This category's validation is usable in all project LRRs and soil textures (see Section 3) because:

- The selected studies span five LRRs (C, H, L, M, S), three of which (H, L, M) are in the declared project domain. Additionally, three sites are used from outside the US in a declared project climate zone (cool temperate dry).
- Six soil textures are included, all of which are included in the declared project domain: clay loam; loam; sandy clay loam; silt loam; silty clay; silty clay loam.
- Clay content spans 32 percentage points, from a low of 15% to a high of 47%.
- At least one study isolates effects, i.e. only 15 of 120 observations compare stacked PC changes.

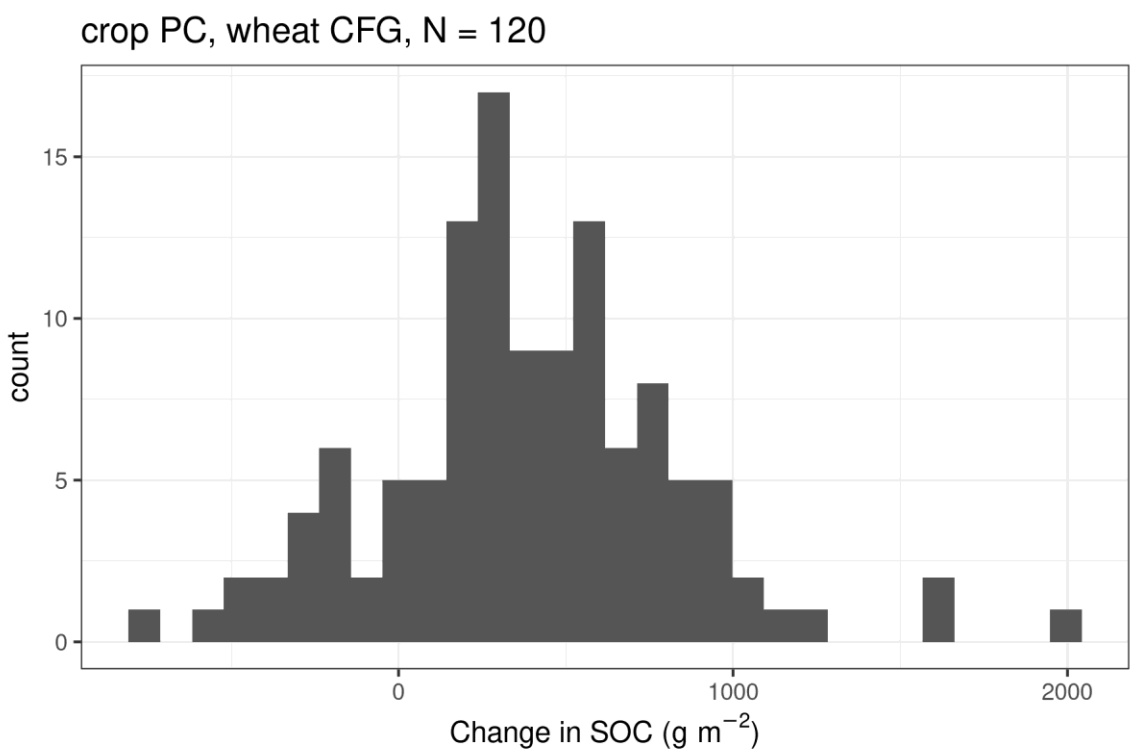

Figure 4: Histogram of changes in SOC observed by the studies used for model validation in response to changed cropping practices involving crops from the wheat-type CFG.

Table 5: 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
davis2	Clark et al. 1998	Davis, CA	1988	1996	C	warm temperate dry	loam	17	Walkley-Black method	5 (4 stack PCs)
hoytville	Collins et al. 1999	Hoytville, OH	1963	1993	L	cool temperate moist	silty clay loam	40	dry combustion	1
imperial	Denef et al. 2008	Imperial, NE	1970	2012	H	cool temperate dry	loam	24	dry combustion	3 (3 stack PCs)
kbs	Collins et al 2000	Hickory Corners KBS, MI	1993	2001	L	cool temperate moist	loam	19	dry combustion	1 (1 stack PCs)
lethbridge2	Janzen et al. 1997	Lethbridge, AB, Canada	1951	1967; 1974; 1985; 1992		cool temperate dry	loam	25	dry combustion	12
lethbridgeABC	Monreal and Janzen 1993	Lethbridge, AB, Canada	1910	1922; 1940; 1953; 1967; 1990		cool temperate dry	sandy clay loam	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	warm temperate dry	silty clay loam	35	dry combustion	1
mead2	Varvel 2006	Mead, NE	1982	1992; 1998; 2002	M	warm temperate dry	silty clay loam	31	dry combustion	18
morrow	Khan et al. 2007	Champaign-Urbana, IL	1955	2005	M	warm temperate moist	silt loam	25	dichromate oxidation technique of Mebius (1960)	2
otis	Denef et al.	Otis, CO	1966	2012	H	cool	loam	26	dry	3 (3

	2008					temperate dry			combustion	stack PCs)
rodale	Elliott et al. 1994; Pimental et al. 2005	Kutztown, PA	1981	1992; 2002	S	warm temperate moist	silt loam	30	not reported	4 (4 stack PCs)
russellranch _LTRAS	Kong et al. 2005	Winter, CA	1993	1997; 2003; 2012	C	warm temperate dry	silt loam	18	dry combustion	11
russellranch _WSREC	Mitchell et al. 2015; Mitchell et al. 2017; Veenstra et al. 2006	Five Points, CA	1999	2004; 2007; 2013	C	warm temperate dry	clay loam	39	dry combustion	6
saginaw	Christenson 1997	Saginaw, MI	1972	1981; 1991	L	cool temperate moist	silty clay	47	dry combustion	10
swiftcurrent	Campbell and Zentner 1997; Campbell et al. 2007	Swift Current, SK, Canada	1966	1981; 1984; 1990; 1993; 1996		cool temperate dry	silt loam	20	dry combustion	25
wooster	Collins et al. 1999; Dick et al. 1997	Wooster, OH	1962	1971; 1980; 1992	M	cool temperate moist	silt loam	15		8

DISTURB x Corn x SOC

This category's validation is usable in all project LRRs and soil textures (see Section 3) because:

- The selected studies span four LRRs (K, L, M, N), all of which are in the declared project domain.
- Three soil textures are included, all of which are included in the declared project domain: loam; silt loam; silty clay loam.
- Clay content spans 25 percentage points, from a low of 15% to a high of 40%.
- At least one study isolates effects, i.e. 0 of 82 observations compare stacked PC changes.

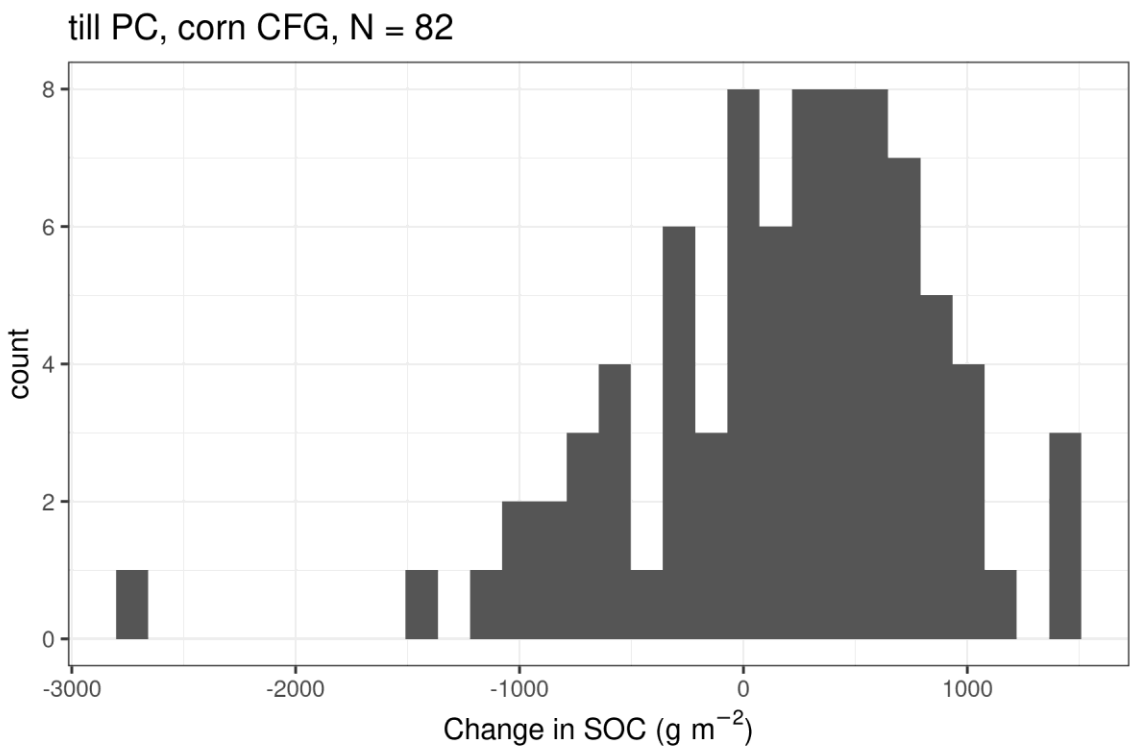

Figure 5: Histogram of changes in SOC observed by the studies used for model validation in response to changed tillage or residue management practices involving crops from the corn-type CFG.

Table 6: 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
dixonsprings	Olson et al. 2010	Dixon Springs, IL	1989	1992; 2000; 2003; 2007; 2009	N	warm temperate moist	silt loam	19	Walkley-Black method	10
Hoytville (One point with error; See Appendix G: Errata)	Collins et al. 1999; Jarecki and Lal 2005; Mestelan 2008	Hoytville, OH	1963	1993; 2003; 2005	L	cool temperate moist	silty clay loam	40	dry combustion	5
ithaca2	Jin et al. 2015; Jin and Varvel 2018a	Ithaca, NE	1998	2001; 2007; 2011	M	warm temperate dry	silty clay loam	32	dry combustion	9
ithacaNE	Schmer et	Ithaca, NE	2001	2010;	M	warm	silt	26	dry	14

	al. 2014; Jin and Varvel 2018b			2014		temperate dry	loam		combustion	
kbs	Collins et al 2000	Hickory Corners KBS, MI	1993	2001	L	cool temperate moist	loam	19	dry combustion	1
lafayette	Elliott et al. 1994	Lafayette, IN	1975	1992	M	warm temperate moist	silty clay loam	36	dry combustion	3
lexington	Blevins et al. 1983; Ismail et al. 1994	Lexington, KY	1970	1980; 1989	N	warm temperate moist	silt loam	23	dry combustion; sulfuric acid-permanganate method of Allison (1965)	8
rosemount	Clapp et al. 2000; Dolan et al. 2006	Rosemount, MN	1980	1992; 2002	K	cool temperate dry	silt loam	24	dry combustion	12
scharleston	Collins et al. 1999; Jarecki and Lal 2005	South Charleston, OH	1962	1992; 2003	M	warm temperate moist	silt loam	20	dry combustion	4
wooster	Collins et al. 1999; Dick et al. 1997; Mestelan 2008	Wooster, OH	1962	1971; 1980; 1992; 2005	M	cool temperate moist	silt loam	15		16

DISTURB x Soy x SOC

This category's validation is usable in all project LRRs and soil textures (see Section 3) because:

- The selected studies span three LRRs (L, M, N), all of which are in the declared project domain.
- Three soil textures are included, all of which are included in the declared project domain: loam; silt loam; silty clay loam.
- Clay content spans 25 percentage points, from a low of 15% to a high of 40%.
- At least one study isolates effects, i.e. 0 of 27 observations compare stacked PC changes.

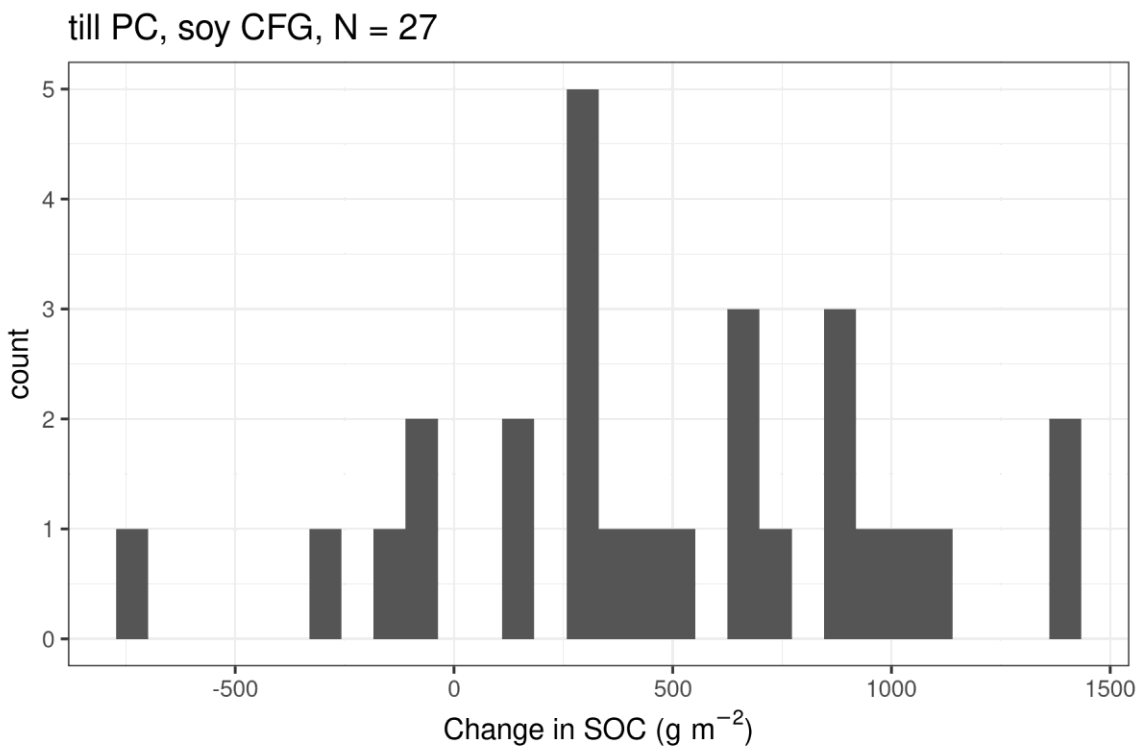

Figure 6: Histogram of changes in SOC observed by the studies used for model validation in response to changed tillage or residue management practices involving crops from the soy-type CFG.

Table 7: 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
dixonsprings	Olson et al. 2010	Dixon Springs, IL	1989	1992; 2000; 2003; 2007; 2009	N	warm temperate moist	silt loam	19	Walkley-Black method	10
hoytville	Collins et al. 1999; Jarecki and Lal 2005; Mestelan 2008	Hoytville, OH	1963	1993; 2003; 2005	L	cool temperate moist	silty clay loam	40	dry combustion	5
ithaca2	Jin et al. 2015; Jin and Varvel 2018a	Ithaca, NE	1998	2001; 2007; 2011	M	warm temperate dry	silty clay loam	32	dry combustion	9
ithacaNE	Schmer et	Ithaca, NE	2001	2010;	M	warm	silt	26	dry	14

	al. 2014; Jin and Varvel 2018b			2014		temperate dry	loam		combustion	
kbs	Collins et al 2000	Hickory Corners KBS, MI	1993	2001	L	cool temperate moist	loam	19	dry combustion	1
lafayette	Elliott et al. 1994	Lafayette, IN	1975	1992	M	warm temperate moist	silty clay loam	36	dry combustion	3
lexington	Blevins et al. 1983; Ismail et al. 1994	Lexington, KY	1970	1980; 1989	N	warm temperate moist	silt loam	23	dry combustion; sulfuric acid-permanganate method of Allison (1965)	8
rosemount	Clapp et al. 2000; Dolan et al. 2006	Rosemount, MN	1980	1992; 2002	K	cool temperate dry	silt loam	24	dry combustion	12
scharleston	Collins et al. 1999; Jarecki and Lal 2005	South Charleston, OH	1962	1992; 2003	M	warm temperate moist	silt loam	20	dry combustion	4
wooster	Collins et al. 1999; Dick et al. 1997; Mestelan 2008	Wooster, OH	1962	1971; 1980; 1992; 2005	M	cool temperate moist	silt loam	15		16

DISTURB x Wheat x SOC

This category's validation is usable in all project LRRs and soil textures (see Section 3) because:

- The selected studies span seven LRRs (B, C, F, G, H, L, M), five of which (F, G, H, L, M) are in the declared project domain.
- Four soil textures are included, all of which are included in the declared project domain: clay loam; loam; silt loam; silty clay loam.
- Clay content spans 25 percentage points, from a low of 15% to a high of 40%.
- At least one study isolates effects, i.e. 0 of 56 observations compare stacked PC changes.

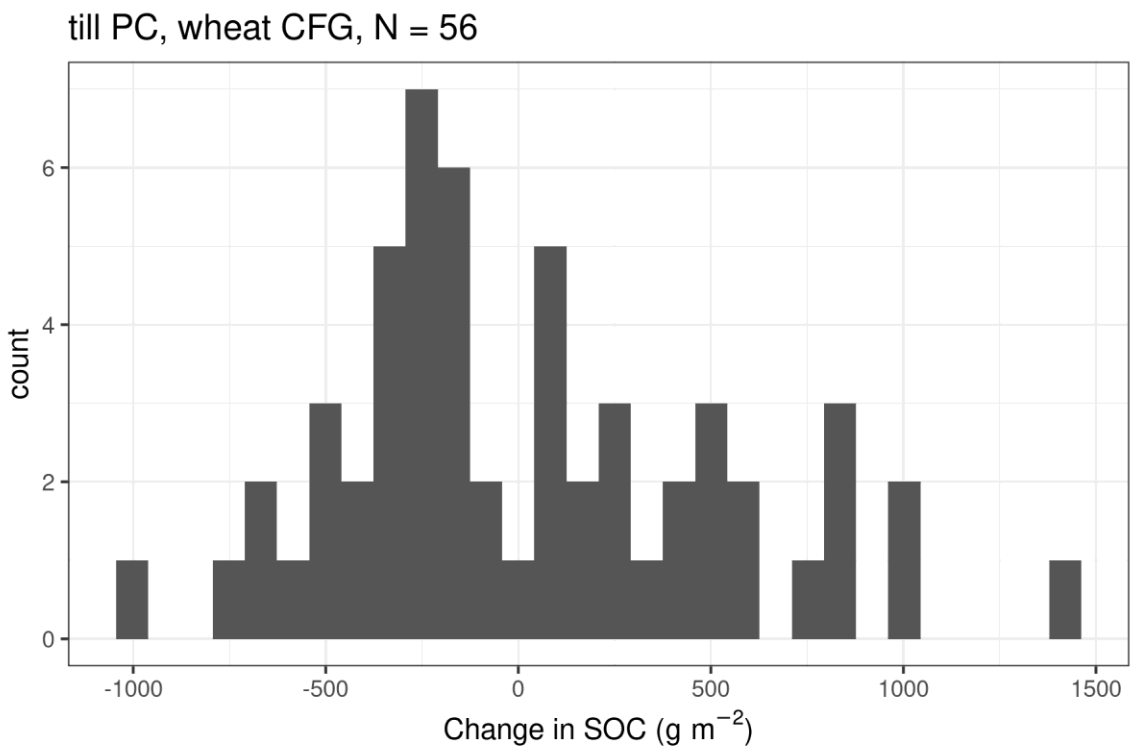

Figure 7: Histogram of changes in SOC observed by the studies used for model validation in response to changed tillage or residue management practices involving crops from the wheat-type CFG.

Table 8: 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 et al. 2013	Akron, CO	1992	2006	G	cool temperate dry	silt loam	25	dry combustion	1
hoytville	Collins et al. 1999	Hoytville, OH	1963	1993	L	cool temperate moist	silty clay loam	40	dry combustion	1
kbs	Collins et al 2000	Hickory Corners KBS, MI	1993	2001	L	cool temperate moist	loam	19	dry combustion	1
mandan_crop	Halvorsen et al. 2002	Mandan, ND	1984	1990	F	cool temperate dry	silt loam	20	dry combustion	12
pendleton1	Bista et al. 2016; Ghimire et	Pendleton, OR	1931	1941; 1951; 1976;	B	warm temperate dry	silt loam	22	dry combustion; Walkley-	16

	al. 2015; Rasmussen and Smiley 1997			1986; 1995; 2005; 2010					Black	
pendleton2	Ghimire et al. 2017	Pendleton, OR	1940	1995	B	warm temperate dry	silt loam	24	dry combustion	12
russellranch _WSREC	Mitchell et al. 2015; Mitchell et al. 2017; Veenstra et al. 2006	Five Points, CA	1999	2004; 2007; 2013	C	warm temperate dry	clay loam	39	dry combustion	6
sidney	Elliott et al. 1994	Sidney, NE	1970	1993	H	cool temperate dry	loam	25	dry combustion	2
wooster	Collins et al. 1999; Dick et al. 1997	Wooster, OH	1962	1971; 1980; 1992	M	cool temperate moist	silt loam	15		5

NFERT x Corn x SOC

This category's validation is usable in all project LRRs and soil textures (see Section 3) because:

- The selected studies span seven LRRs (C, E, K, L, M, N, S), four of which (K, L, M, N) are in the declared project domain.
- Five soil textures are included, all of which are included in the declared project domain: clay loam; loam; loamy sand; silt loam; silty clay loam.
- Clay content spans 25 percentage points, from a low of 10% to a high of 35%.
- At least one study isolates effects, i.e. only 17 of 91 observations compare stacked PC changes.

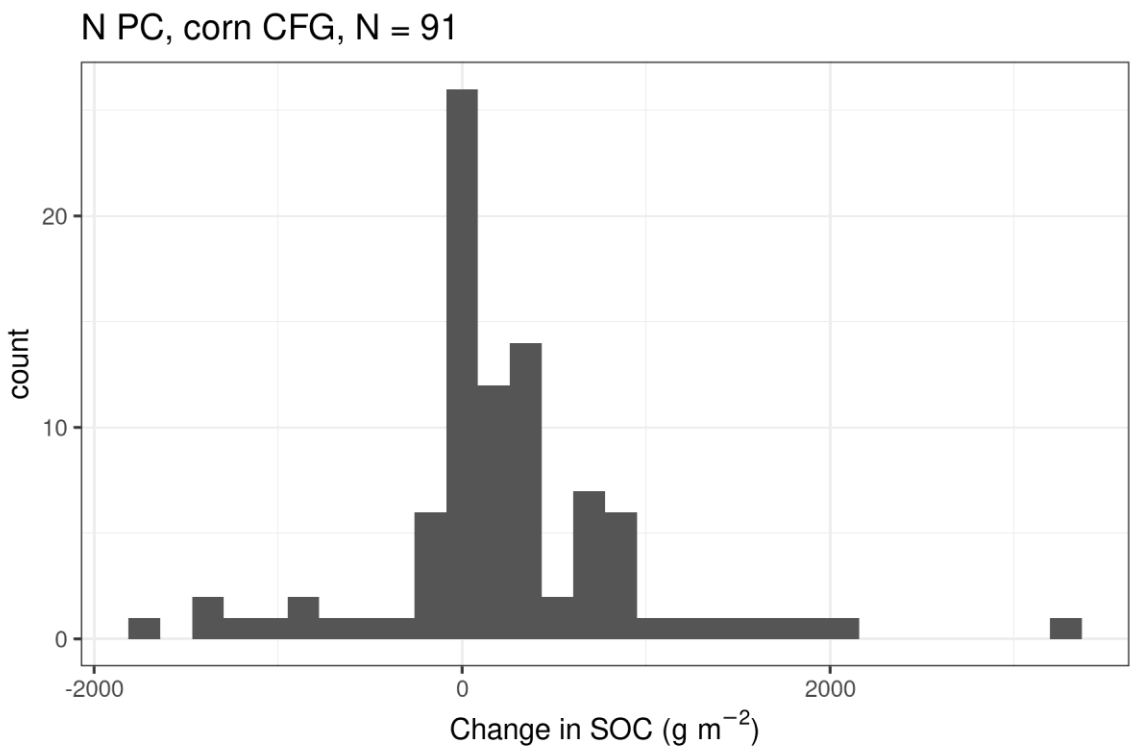

Figure 8: Histogram of changes in SOC observed by the studies used for model validation in response to changed inorganic N management practices involving crops from the corn-type CFG

Table 9: 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 et al. 2016	Fort Collins, CO	2011	2014	E	cool temperate dry	clay loam	34	dry combustion	2 (1 stack PCs)
ardec1	Halvorson and Jantalia 2011	Fort Collins, CO	1999	2009	E	cool temperate dry	clay loam	34	dry combustion	4
davis2	Clark et al. 1998	Davis, CA	1988	1996	C	warm temperate dry	loam	17	Walkley-Black method	4 (4 stack PCs)
elansing2	Vitosh et al. 1973	East Lansing, MI	1963	1971	L	cool temperate moist	loamy sand	10	high frequency induction furnace	3 (3 stack PCs)
ithaca2	Jin and Varvel	Ithaca, NE	1998	2001; 2007;	M	warm temperate	silty clay	32	dry combustion	12

	2018a; Jin et al. 2015			2011		dry	loam			
kbs	Collins et al 2000	Hickory Corners KBS, MI	1993	2001	L	cool temperate moist	loam	19	dry combustion	1 (1 stack PCs)
lexington	Blevins et al. 1983; Ismail et al. 1994	Lexington, KY	1970	1980; 1989	N	warm temperate moist	silt loam	23	dry combustion; sulfuric acid-permanganate method of Allison (1965)	12
mead	Elliott et al. 1994	Mead, NE	1975	1992	M	warm temperate dry	silty clay loam	35	dry combustion	1 (1 stack PCs)
mead2	Varvel 2006	Mead, NE	1982	1992; 1998; 2002	M	warm temperate dry	silty clay loam	31	dry combustion	36
morrow	Khan et al. 2007	Champaign -Urbana, IL	1955	2005	M	warm temperate moist	silt loam	25	dichromate oxidation technique ofMebius (1960)	3
rodale	Elliott et al. 1994; Pimental et al. 2005	Kutztown, PA	1981	1992; 2002	S	warm temperate moist	silt loam	30	not reported	4 (4 stack PCs)
rosemount	Dolan et al. 2006	Rosemount, MN	1980	2002	K	cool temperate dry	silt loam	24	dry combustion	6
russellranch _LTRAS	Kong et al. 2005	Winter, CA	1993	1997; 2003; 2012	C	warm temperate dry	silt loam	18	dry combustion	3 (3 stack PCs)

NFERT x Wheat x SOC

This category's validation is usable in all project LRRs and soil textures (see Section 3) because:

- The selected studies span six LRRs (B, C, F, L, M, S), three of which (F, L, M) are in the declared project domain, as well as sites outside the US that are within the declared project climate zones (cool temperate moist, cool temperate dry).
- Three soil textures are included, all of which are included in the declared project domain: loam; silt loam; silty clay loam.
- Clay content spans 18 percentage points, from a low of 17% to a high of 35%.

- At least one study isolates effects, i.e. only 19 of 102 observations compare stacked PC changes.

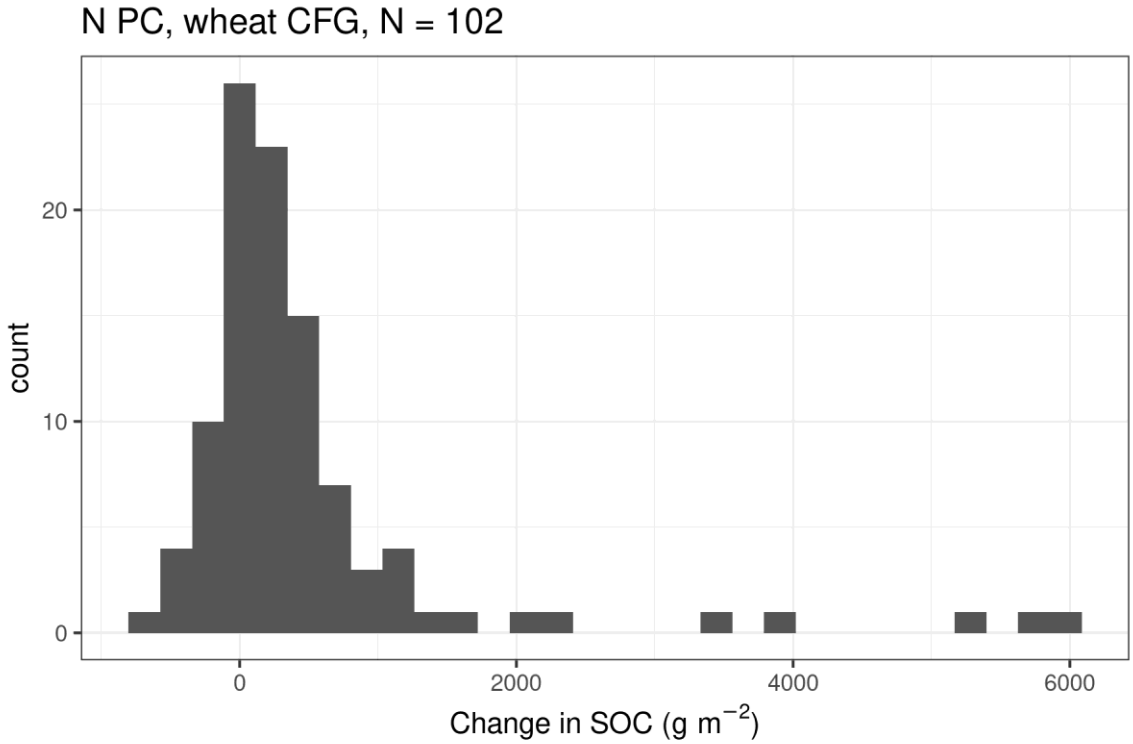

Figure 9: Histogram of changes in SOC observed by the studies used for model validation in response to changed inorganic N management practices involving crops from the wheat-type CFG.

Table 10: 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	Rothamsted Research (2014)	Rothamsted, England	1844	1893; 1914; 1944; 1992; 1997; 2005		cool temperate moist	silty clay loam	25	dry combustion	12 (6 stack PCs)
davis2	Clark et al. 1998	Davis, CA	1988	1996	C	warm temperate dry	loam	17	Walkley-Black method	4 (4 stack PCs)
kbs	Collins et al 2000	Hickory Corners KBS, MI	1993	2001	L	cool temperate moist	loam	19	dry combustion	1 (1 stack PCs)
mandan_crop	Halvorsen et al. 2002	Mandan, ND	1984	1990	F	cool temperate	silt loam	20	dry combustion	12

						dry				
mead	Elliott et al. 1994	Mead, NE	1975	1992	M	warm temperate dry	silty clay loam	35	dry combustion	1 (1 stack PCs)
mead2	Varvel 2006	Mead, NE	1982	1992; 1998; 2002	M	warm temperate dry	silty clay loam	31	dry combustion	12
morrow	Khan et al. 2007	Champaign -Urbana, IL	1955	2005	M	warm temperate moist	silt loam	25	dichromate oxidation technique ofMebius (1960)	1
pendleton1	Bista et al. 2016; Ghimire et al. 2015; Rasmussen and Smiley 1997	Pendleton, OR	1931	1941; 1951; 1976; 1986; 1995; 2005; 2010	B	warm temperate dry	silt loam	22	dry combustion; Walkley-Black	16
pendleton2	Ghimire et al. 2017	Pendleton, OR	1940	1995	B	warm temperate dry	silt loam	24	dry combustion	15
rodale	Elliott et al. 1994; Pimental et al. 2005	Kutztown, PA	1981	1992; 2002	S	warm temperate moist	silt loam	30	not reported	4 (4 stack PCs)
russellranch _LTRAS	Kong et al. 2005	Winter, CA	1993	1997; 2003; 2012	C	warm temperate dry	silt loam	18	dry combustion	9 (3 stack PCs)
swiftcurrent	Campbell and Zentner 1997; Campbell et al. 2007	Swift Current, SK, Canada	1966	1981; 1984; 1990; 1993; 1996		cool temperate dry	silt loam	20	dry combustion	15

ORG x Wheat x SOC

This category's validation is usable in all project LRRs and soil textures (see Section 3) because:

• The selected studies span four LRRs (B, C, M, S). Only one of these (M) is in the declared project domain, but the studies also include sites outside the US that are within the declared project climate zones (cool temperate dry, cool temperate moist). Collectively across US and international sites, the validation data are taken from all four declared project climate zones (cool temperate dry, cool temperate moist, warm temperate dry, warm temperate moist). Therefore the bioclimatic distribution requirements are met.

• Four soil textures are included, all of which are included in the declared project domain: clay loam; loam; silt loam; silty clay loam.
• Clay content spans 22 percentage points, from a low of 17% to a high of 39%.
• At least one study isolates effects, i.e. only 14 of 44 observations compare stacked PC changes.

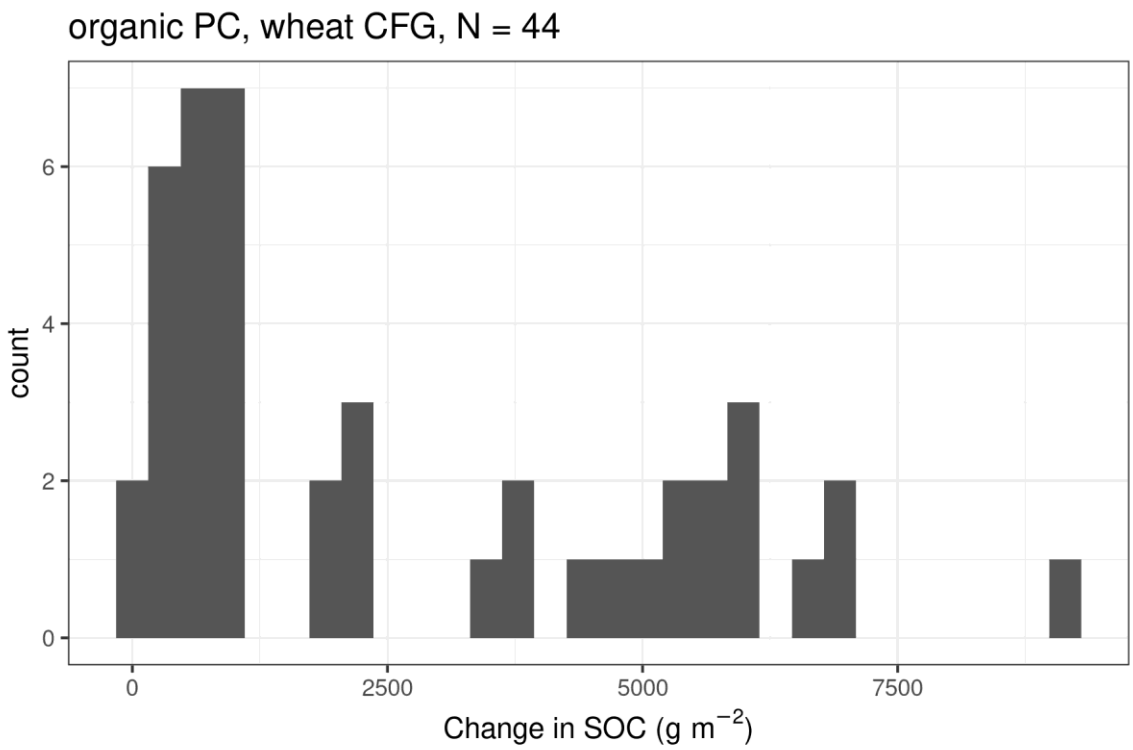

Figure 10: Histogram of changes in SOC observed by the studies used for model validation in response to changed organic amendment practices involving crops from the wheat-type CFG.

Table 11: 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	Rothamsted Research (2014)	Rothamsted, England	1844	1893; 1914; 1944; 1992; 1997; 2005		cool temperate moist	silty clay loam	25	dry combustion	12 (6 stack PCs)
davis2	Clark et al. 1998	Davis, CA	1988	1996	C	warm temperate dry	loam	17	Walkley-Black method	2 (2 stack PCs)
lethbridge_manure	Hao et al. 2003	Lethbridge, AB, Canada	1973	1998		cool temperate dry	clay loam	39	dry combustion	6

lethbridge2	Janzen et al. 1997	Lethbridge, AB, Canada	1951	1967; 1974; 1985; 1992		cool temperate dry	loam	25	dry combustion	4
mead	Elliott et al. 1994	Mead, NE	1975	1992	M	warm temperate dry	silty clay loam	35	dry combustion	1 (1 stack PCs)
pendleton1	Bista et al. 2016; Ghimire et al. 2015; Rasmussen and Smiley 1997	Pendleton, OR	1931	1941; 1951; 1976; 1986; 1995; 2005; 2010	B	warm temperate dry	silt loam	22	dry combustion; Walkley-Black	14
rodale	Elliott et al. 1994; Pimental et al. 2005	Kutztown, PA	1981	1992; 2002	S	warm temperate moist	silt loam	30	not reported	2 (2 stack PCs)
russellranch _LTRAS	Kong et al. 2005	Winter, CA	1993	1997; 2003; 2012	C	warm temperate dry	silt loam	18	dry combustion	3 (3 stack PCs)

ORG x Corn x SOC

This category is reported for information purposes to support the ORG x all crops x SOC category (below). ORG x Corn X SOC on its own does not fulfill the bioclimatic distribution requirements:

- The selected studies span five LRRs (C, E, L, M, S), but **only two** of these (L, M) are **in the declared project domain**. They do span all four declared project climate zones, but Model Requirements Section 3.3 only allows considering these for sites outside the US.
- Five soil textures are included, all of which are included in the declared project domain: clay loam; loam; loamy sand; silt loam; silty clay loam.
- Clay content spans 25 percent, from a low of 10% to a high of 35%.
- At least one study isolates effects, i.e. only 12 of 13 observations compare stacked PC changes.

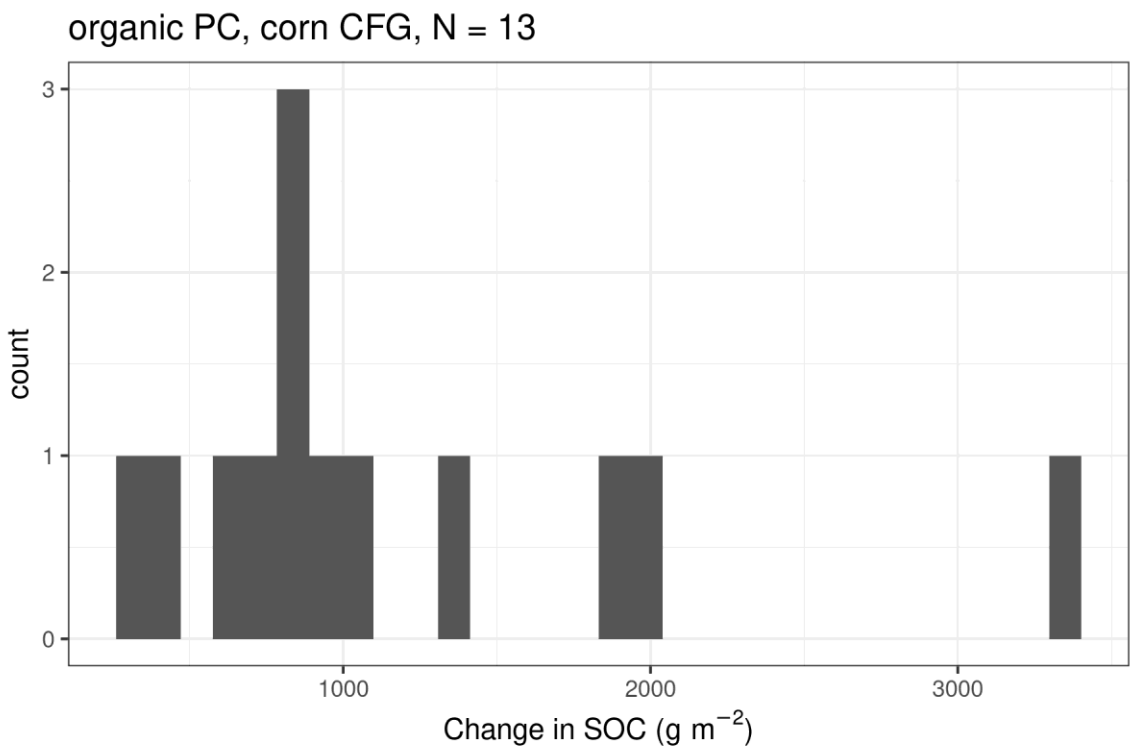

Figure 11: Histogram of changes in SOC observed by the studies used for model validation in response to changed organic amendment practices involving crops from the corn-type CFG.

Table 12: 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 et al. 2016	Fort Collins, CO	2011	2014	E	cool temperate dry	clay loam	34	dry combustion	2 (1 stack PCs)
davis2	Clark et al. 1998	Davis, CA	1988	1996	C	warm temperate dry	loam	17	Walkley-Black method	2 (2 stack PCs)
elansing2	Vitosh et al. 1973	East Lansing, MI	1963	1971	L	cool temperate moist	loamy sand	10	high frequency induction furnace	3 (3 stack PCs)
mead	Elliott et al. 1994	Mead, NE	1975	1992	M	warm temperate dry	silty clay loam	35	dry combustion	1 (1 stack PCs)
rodale	Elliott et al. 1994; Pimental et	Kutztown, PA	1981	1992; 2002	S	warm temperate moist	silt loam	30	not reported	2 (2 stack PCs)

	al. 2005									
russellranch _LTRAS	Kong et al. 2005	Winter, CA	1993	1997; 2003; 2012	C	warm temperate dry	silt loam	18	dry combustion	3 (3 stack PCs)

ORG x Soy x SOC

This category is reported for information purposes to support the ORG x all crops x SOC category. ORG x Soy x SOC on its own does not fulfill the bioclimatic distribution and stacked-practice requirements:

• The selected studies span three LRRs (C, M, S), **only one** of which (M) is **in the declared project domain**, and only two declared project climate zones (warm temperate moist, warm temperate dry).
• three soil textures are included, all of which are included in the declared project domain: loam; silt loam; silty clay loam.
• Clay content spans 18 percent, from a low of 17% to a high of 35%.
• **No studies isolate effects**, i.e. all 5 observations compare stacked PC changes.

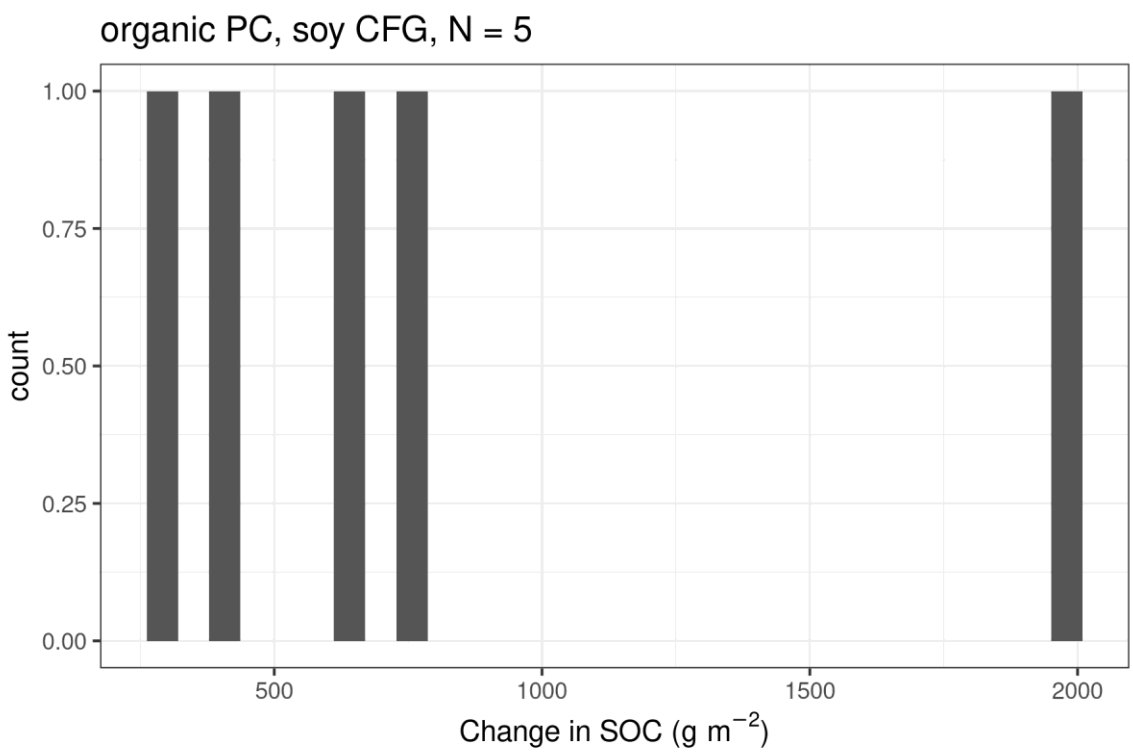

Figure 12: Histogram of changes in SOC observed by the studies used for model validation in response to changed organic amendment practices involving crops from the soy-type CFG.

Table 13: 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	warm temperate dry	loam	17	Walkley-Black method	2 (2 stack PCs)
mead	Elliott et al. 1994	Mead, NE	1975	1992	M	warm temperate dry	silty clay loam	35	dry combustion	1 (1 stack PCs)
rodale	Elliott et al. 1994; Pimental et al. 2005	Kutztown, PA	1981	1992; 2002	S	warm temperate moist	silt loam	30	not reported	2 (2 stack PCs)

ORG x all crops x SOC

This category is reported following the approach described in Appendix D: Proposal for validating organic amendment applications, the proposal for which is under consideration by CAR and would allow validation of ORG across multiple CFGs; if approved, this category will be usable in all project LRRs and soil textures because:

- The selected studies span six LRRs (B, C, E, L, M, S). Only two of these (L, M) are in the declared project domain, but studies also include sites outside the US that are within the declared project climate zones (cool temperate dry, cool temperate moist). Collectively across US and international sites, the validation data are taken from all four declared project climate zones (cool temperate dry, cool temperate moist, warm temperate dry, warm temperate moist). Therefore the bioclimatic distribution requirements are met.
- Five soil textures are included, all of which are included in the declared project domain: clay loam, loam, loamy sand, silt loam, silty clay loam.
- Clay content spans 29 percent, from a low of 10% to a high of 39%.
- At least one study isolates effects, i.e. only 18 of 31 observations compare stacked PC changes.

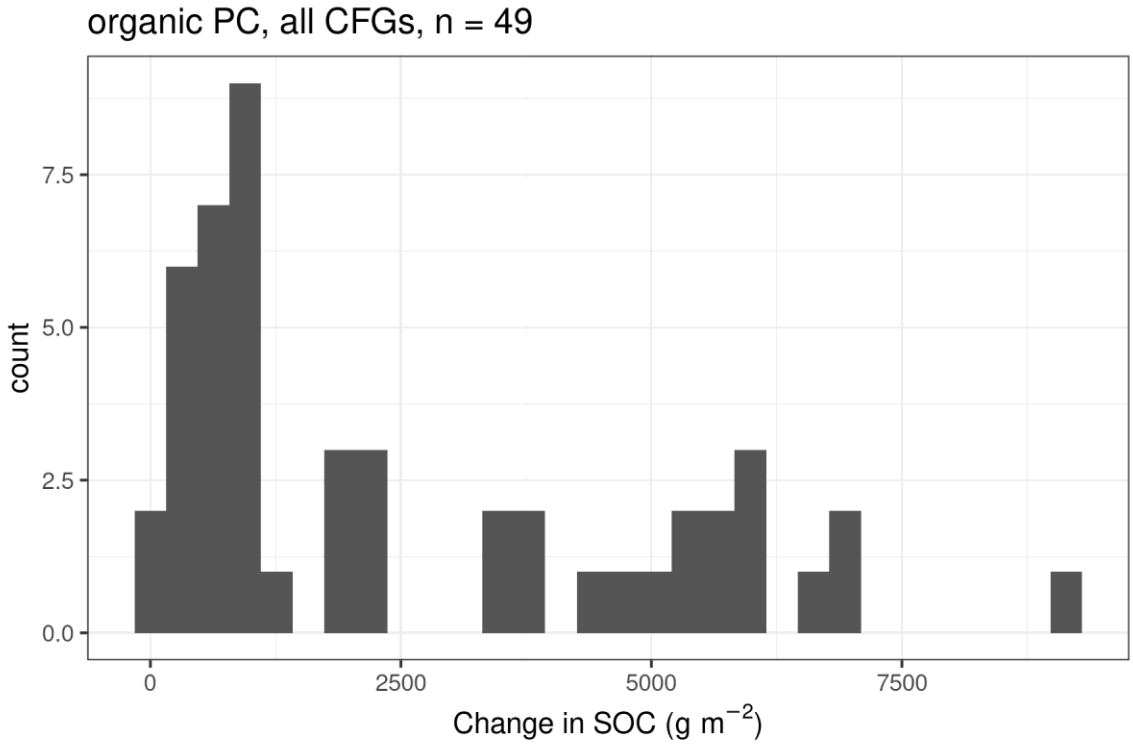

Figure 13: Histogram of changes in SOC observed by the studies used for model validation in response to changed organic amendment practices involving all crop types combined.

Descriptive dataset attributes are reported above in the sections for ORG x Wheat (Table 11), ORG x Corn (Table 12), and ORG x Soy (Table 13).

Bias evaluation

Follows Model Requirements Section 3.4 Summary of Requirements (p18)

Calculating bias

In all categories, bias was computed for each study as the mean difference between modeled and observed practice effects:

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

where $observed_i$ is the change in SOC observed from the ith experimental treatment (e.g. SOC_{notill} – SOC_{till}), $modeled_i$ is the change in SOC predicted by modeling the ith experimental treatment, and n is the number of treatments used from the study (per Equation 3.1 of the Model Requirements). When a study reported observations fitting multiple categories, only the observations matching the category of interest were included in the calculation.

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 Equation 3.2 of the Model Requirements):

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

Where k is the number of observations with uncertainty reported, σ_j is the standard error of the jth observation of differences between the treatments, and n_j is the number of replicates included in the observation.

Example PMU calculation

The CROP x Corn validation dataset contained k = nine observations of practice changes that reported uncertainty 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 14: 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
hoytville	3	3	419	37	420.6
hoytville	3	3	670	509	841.4
hoytville	3	3	670	419	790.2
kbs	30	30	126	119	173.3
mead	4	4	455	288	538.5
wooster	3	3	127	215	249.7
wooster	3	3	179	80	196.1
wooster	3	3	127	171	213.0
wooster	3	3	179	154	236.1

(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 $\sqrt{sd_1^2/n_1 + sd_2^2/n_2}$.)
Thus to compute PMU the needed summations over n-1 and se_diff are:

Table 15: Computing pooled measurement uncertainty for CROP x corn from the standard errors of differences shown in Table 14.

site	n trt1	n trt2	se trt1	se trt2	se diff	se^2	n-1	se^2*(n-1)
hoytville	3	3	419	37	420.6	176930	2	353860
hoytville	3	3	670	509	841.4	707981	2	1415962
hoytville	3	3	670	419	790.2	624461	2	1248922
kbs	30	30	126	119	173	30037	29	871073
mead	4	4	455	288	538	289969	3	869907
wooster	3	3	127	215	250	62354	2	124708
wooster	3	3	179	80	196	38441	2	76882
wooster	3	3	127	171	213	45370	2	90740
wooster	3	3	179	154	236	55757	2	111514
						sum	46	5163568
						sqrt(5163568/46) = **PMU**		**335.03**

PMU coverage by category

Table 16: 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 2) with uncertainty available. % sites: percentage of the sites in the full dataset (Table 2) 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.

PC	CFG	N obs	N stacked	PMU	N sites	% obs	% sites	Citations
CROP	corn	9	1	335.0395	4	8	33	a,b,c
CROP	soy	9	1	335.0395	4	8	31	a,b,c
CROP	wheat	5	1	289.8045	4	4	25	a,b,c
DISTURB	corn	21	0	635.3432	6	26	60	a,b,c,d,e
DISTURB	soy	8	0	400.4813	4	30	80	a,b,c
DISTURB	wheat	6	0	282.1985	4	11	44	a,b,f
NFERT	corn	14	2	764.6667	3	15	23	b,c,d,e
NFERT	wheat	6	2	212.8265	3	6	25	b,c,f
ORG	wheat	3	1	376.1941	2	7	25	c,f
ORG	corn	1	1	529.59	1	8	17	c
ORG	soy	1	1	529.59	1	20	33	c
ORG	all crops	3	1	376.1941	2	6	20	c,f
All categories combined		52	2	624.22	8	11	24	a,b,c,d,e,f

Citations:
a: Collins et al. 1999
b: Collins et al. 2000
c: Elliott et al. 1994
d: Denef et al. 2008
e: Jin and Varvel 2018a
f: Ghimire et al. 2015

Bias across all categories

Pooled measurement uncertainty for the entire dataset: 624 g C m^{-2}

Table 17: Model bias per study (g C m^2 across entire time measured) 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 48

Table 48 for study fold assignments.

Study	n observations	Bias
lethbridge_manure	7	1164.088
rosemount	18	518.0113
davis2	5	371.1138
swiftcurrent	40	370.7749
lafayette	3	354.9061

rodale	4	323.8998
ardec_manure	3	317.0886
sidney	2	264.6711
mead	2	262.3416
lexington	20	216.3887
hoytville	8	178.2824
ardec1	4	159.8528
wooster	32	126.5921
mead2	96	100.8503
pendleton2	27	82.50685
ithacaNE	14	55.31871
russellranch_LTRAS	20	34.84387
pendleton1	46	21.03628
lethbridge2	16	4.791931
scharleston	4	4.432256
mandan_crop	24	3.649181
otis	3	-6.39978
ithaca2	21	-72.8452
russellranch_WSREC	12	-98.5501
imperial	3	-146.929
dixonsprings	10	-188.818
saginaw	10	-206.556
lethbridgeABC	10	-335.572
kbs	2	-440.005
akron	1	-544.917
morrow	7	-648.464
elansing2	3	-656.732
broadbalk	18	-691.68
All fold 1 sites	114	57.9
All fold 2 sites	179	-124.8
All fold 3 sites	81	116.9
All fold 4 sites	53	265.5
All fold 5 sites	68	-174.2
Across all sites	**495**	**27.21**

Mean bias across all studies and PC x CFG combinations: 27 g C m^{-2}

Is bias smaller than PMU? **Yes!**

CROP x Corn x SOC

Pooled measurement uncertainty (PMU) = 335 g C m^{-2}

Table 18: Model bias per study (g C m^2 across entire time measured) 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 48 *Table 48 for study fold assignments.*

Study	**n observations**	**Bias**
hoytville	3	482.49
davis2	5	371.11
rodale	4	323.9
wooster	16	164.44
russellranch_LTRAS	6	45.95
otis	3	-6.4
mead	1	-84.55
mead2	54	-123.59
imperial	3	-146.93
saginaw	10	-206.56
morrow	4	-339.19
kbs	1	-911.73
Across all sites	**110**	**-35.9 +- 35.5**

Mean bias across all studies = -36 g C m^{-2}

Is bias smaller than PMU? **Yes!**

CROP x Soy x SOC

Pooled measurement uncertainty (PMU) = 335 g C m^{-2}

Table 19: Model bias per study (g C m^2 across entire time measured) 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 48 *Table 48 for study fold assignments.*

Study	n observations	Bias
hoytville	3	482.49
swiftcurrent	5	467.83
davis2	5	371.11
rodale	4	323.9
wooster	16	164.44
otis	3	-6.4
russellranch_LTRAS	8	-61.44
mead	1	-84.55
mead2	45	-143.19
imperial	3	-146.93
saginaw	8	-204.14
morrow	4	-339.19
kbs	1	-911.73
Across all sites	**106**	**-6.8 +- 37.2**

Mean bias across all studies = -7 g C m^{-2}

Is bias smaller than PMU? **Yes!**

CROP x Wheat x SOC

Pooled measurement uncertainty (PMU) = 290 g C m^{-2}

Table 20: Model bias per study (g C m^2 across entire time measured) 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 48

Table 48 for study fold assignments.

Study	n observations	Bias
hoytville	1	570.14
swiftcurrent	25	425.54
davis2	5	371.11
rodale	4	323.9
otis	3	-6.4
russellranch_LTRAS	11	-19.85
wooster	8	-56.99
russellranch_WSREC	6	-70.31
mead	1	-84.55
lethbridge2	12	-120.97
imperial	3	-146.93
mead2	18	-196.16
saginaw	10	-206.56
Across all sites	**120**	**-75.7 +- 35.7**

Mean bias across all studies = -36 g C m^{-2}

Is bias smaller than PMU? **Yes!**

DISTURB x Corn x SOC

Pooled measurement uncertainty (PMU) = 635 g C m^{-2}

Table 21: Model bias per study (g C m^2 across entire time measured) 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 48

Table 48 for study fold assignments.

Study	n observations	Bias
lafayette	3	354.91
lexington	8	342.38
rosemount	12	287.93
wooster	16	88.75
ithacaNE	14	55.32
kbs	1	31.72
scharleston	4	4.43
hoytville	5	-4.24
dixonsprings	10	-188.82
ithaca2	9	-203.69
Across all sites	**82**	**76.9 +- 22.0**

Mean bias across all studies = +77 g C m^{-2}

Is bias smaller than PMU? **Yes!**

DISTURB x Soy x SOC

Pooled measurement uncertainty (PMU) = 400 g C m^{-2}

Table 22: Model bias per study (g C m^2 across entire time measured) 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 48 *Table 48 for study fold assignments.*

Study	n observations	Bias
lafayette	3	354.91
kbs	1	31.72

wooster	10	-25.86
dixonsprings	10	-188.82
hoytville	3	-234.27
Across all sites	**27**	**-12.5 +- 44.9**

Mean bias across all studies = -13 g C m^{-2}

Is bias smaller than PMU? **Yes!**

DISTURB x Wheat x SOC

Pooled measurement uncertainty (PMU) = 282 g C m^{-2}

Table 23: Model bias per study (g C m^2 across entire time measured) 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 48

Table 48 for study fold assignments.

Study	n observations	Bias
sidney	2	264.67
pendleton2	12	113.76
mandan_crop	12	86.8
pendleton1	16	84.22
wooster	5	34.29
kbs	1	31.72
russellranch_WSREC	6	-126.79
akron	1	-544.92
hoytville	1	-687.83
Across all sites	**56**	**-82.7 +- 42.9**

Mean bias across all studies = -83 g C m^{-2}

Is bias smaller than PMU? **Yes!**

NFERT x Corn x SOC

Pooled measurement uncertainty (PMU) = 765 g C m^{-2}

Table 24: Model bias per study (g C m^2 across entire time measured) 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 48 *Table 48 for study fold assignments.*

Study	**n observations**	**Bias**
rosemount	6	978.18
mead	1	609.23
davis2	4	453.06
mead2	36	425.48
rodale	4	323.9
ardec_manure	2	237.82
ardec1	4	159.85
lexington	12	132.39
ithaca2	12	25.29
russellranch_LTRAS	3	-58.58
elansing2	3	-656.73
kbs	1	-911.73
morrow	3	-1060.83
Across all sites	**91**	**50.6 +- 62.6**

Mean bias across all studies = +51 g C m^{-2}

Is bias smaller than PMU? **Yes!**

NFERT x Wheat x SOC

Pooled measurement uncertainty (PMU) = 213 g C m^{-2}

Table 25: Model bias per study (g C m^2 across entire time measured) 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 48 *Table 48 for study fold assignments.*

Study	**n observations**	**Bias**
mead	1	609.23
davis2	4	453.06
mead2	12	388.05
rodale	4	323.9
swiftcurrent	15	279.49
russellranch_LTRAS	9	101.69
pendleton2	15	57.5
pendleton1	16	54.18
mandan_crop	12	-79.5
broadbalk	12	-409.05
morrow	1	-464.97
kbs	1	-911.73
Across all sites	**102**	**33.5 +- 43.1**

Mean bias across all studies = +34 g C m^{-2}

Is bias smaller than PMU? **Yes!**

ORG x Wheat x SOC

Pooled measurement uncertainty (PMU) = 376 g C m^{-2}

Table 26: Model bias per study (g C m^2 across entire time measured) 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 48 *Table 48 for study fold assignments.*

Study	n observations	Bias
lethbridge_manure	6	1492.96
davis2	2	704.05
mead	1	609.23
lethbridge2	4	382.09
rodale	2	-17.27
russellranch_LTRAS	3	-58.58
pendleton1	14	-89.05
broadbalk	12	-938.12
Across all sites	**44**	**260.7 +- 108.0**

Mean bias across all studies = +261 g C m^{-2}

Is bias smaller than PMU? **Yes!**

ORG x Corn x SOC

Reported to support ORG x all crops x SOC following the approach in Appendix D: Proposal for validating organic amendment applications

Pooled measurement uncertainty (PMU) = 530 g C m^{-2}

Table 27: Model bias per study (g C m^2 across entire time measured) 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 48 *Table 48 for study fold assignments.*

Study	n	Bias

	observations	
davis2	2	704.05
mead	1	609.23
ardec_manure	2	478.19
rodale	2	-17.27
russellranch_LTRAS	3	-58.58
elansing2	3	-656.73
Across all sites	**13**	**176.5± 143.8**

Mean bias across all studies = +177 g C m^{-2}

Is bias smaller than PMU? **Yes!**

ORG x Soy x SOC

Reported to support ORG x all crops x SOC following the approach in Appendix D: Proposal for validating organic amendment applications

Pooled measurement uncertainty (PMU) = 530 g C m^{-2}

Table 28: Model bias per study (g C m^2 across entire time measured) 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 48
Table 48 for study fold assignments.

Study	n observations	Bias
davis2	2	704.05
mead	1	609.23
rodale	2	-17.27
Across all sites	**5**	**432.0 ± 175.3**

Mean bias across all studies = +432 g C m^{-2}

Is bias smaller than PMU? **Yes!**

ORG x all crops x SOC

Reported following the approach in Appendix D: Proposal for validating organic amendment applications

Pooled measurement uncertainty (PMU) = 376 g C m^{-2}

Table 29: Model bias per study (g C m^2 across entire time measured) 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 48
Table 48 for study fold assignments.

Study	**n observations**	**Bias**
lethbridge_manure	6	1492.96
davis2	2	704.05
mead	1	609.23
ardec_manure	2	478.19
lethbridge2	4	382.09
rodale	2	-17.27
russellranch_LTRAS	3	-58.58
pendleton1	14	-89.05
elansing2	3	-656.73
broadbalk	12	-938.12
Across all sites	**44**	**260.7 +- 108.0**

Mean bias across all studies = +261 g C m^{-2}

Is bias smaller than PMU? **Yes!**

Model prediction error

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

Description of calculation method

Model uncertainty bounds on the SOC difference between the practice and the baseline were estimated using a Monte Carlo method as described in Gurung et al. (2020). In brief, the method provides a probabilistic framework in which samples from the posterior distribution of uncertain model parameters are first drawn using a Bayesian calibration method (see Model Calibration), and then the predictive distribution of the model outputs (e.g. SOC stocks and stock differences) is computed through Monte Carlo simulation. Prediction uncertainty is assigned by propagating the errors in the parameters and hyperparameters through the Monte Carlo simulation framework that accounts for three major sources of uncertainty in the model prediction: measurement error variance, model bias including site-level error, and model error variance similar to the method described by Kennedy and O'Hagan, (2001). After all simulations are complete, the 90% posterior prediction intervals are calculated by taking the 5th and the 95th percentile values from the Monte Carlo simulation, providing the central interval of the posterior prediction (Gelman et. al., 2014; page-33). The performance metric for acceptable model uncertainty is the average number of observations from out-of-sample validation data that fall within the 90% posterior prediction interval.

Because model uncertainty is estimated separately in each cross-validation fold and the fold assignment retains both independence between folds (including e.g. spatial correlation) and comparable predictor ranges between folds (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 uncertainty across folds is a valid estimate of uncertainty when the model is applied to new sites within the validated geographic, bioclimatic, and management domains (Roberts et al. 2017). The estimated uncertainty is likely to be somewhat conservative (overstated) relative to approaches such as leave-one-out, but the additional computational effort of using more folds was considered not worth the change in estimated uncertainty.

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): The posterior distribution of the model parameters is defined by the product of the data-likelihood function and the prior. 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 (Appendix E: Table 46; also see "Documentation of model parameter sets" for more details about the choice of prior values). 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.

Model prediction error across all categories

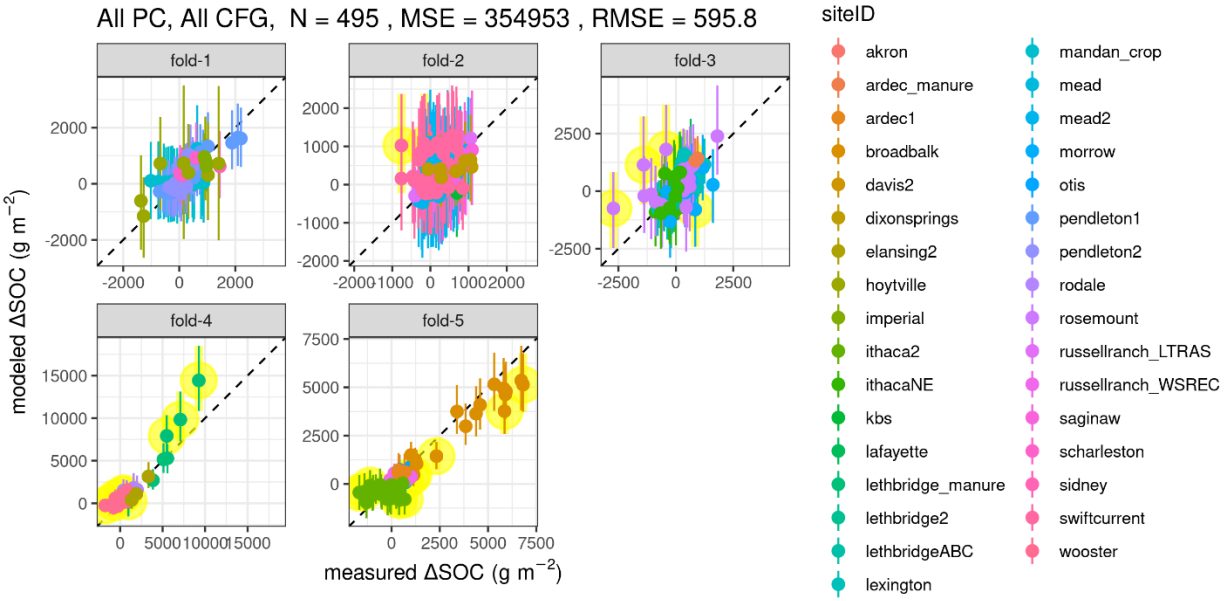

Figure 14: Model predictions versus measurements of SOC change in all practice changes and crop types. Error bars show 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

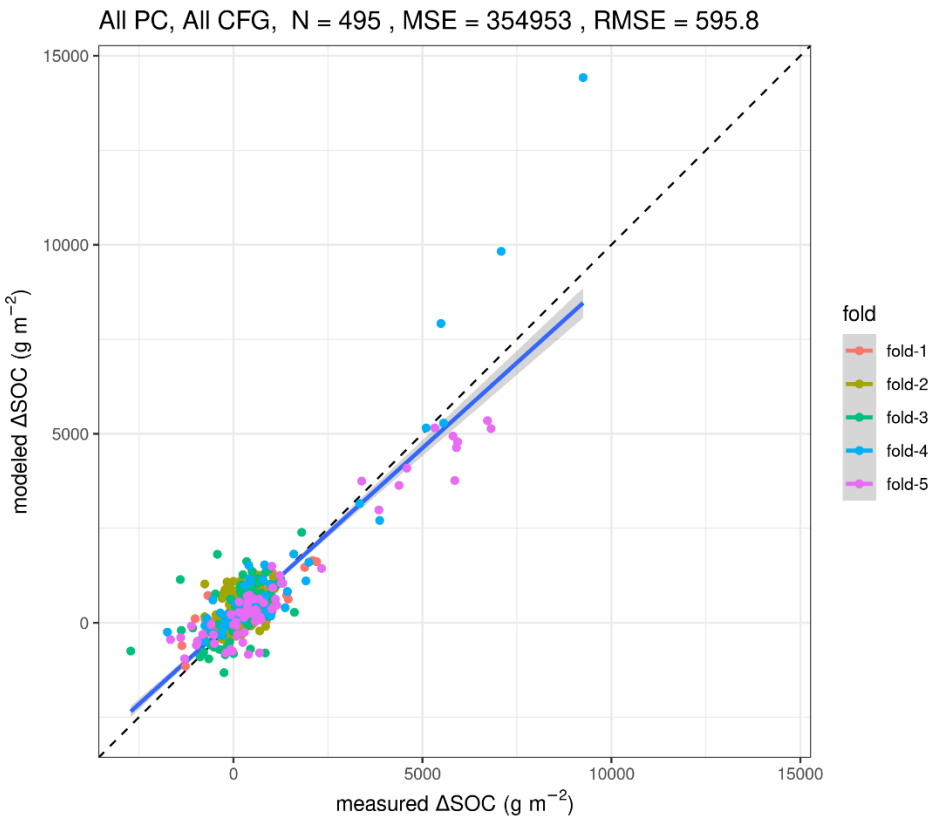

Figure 15: Scatterplot of model predictions versus measurements of SOC change in all practice changes and crop types. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

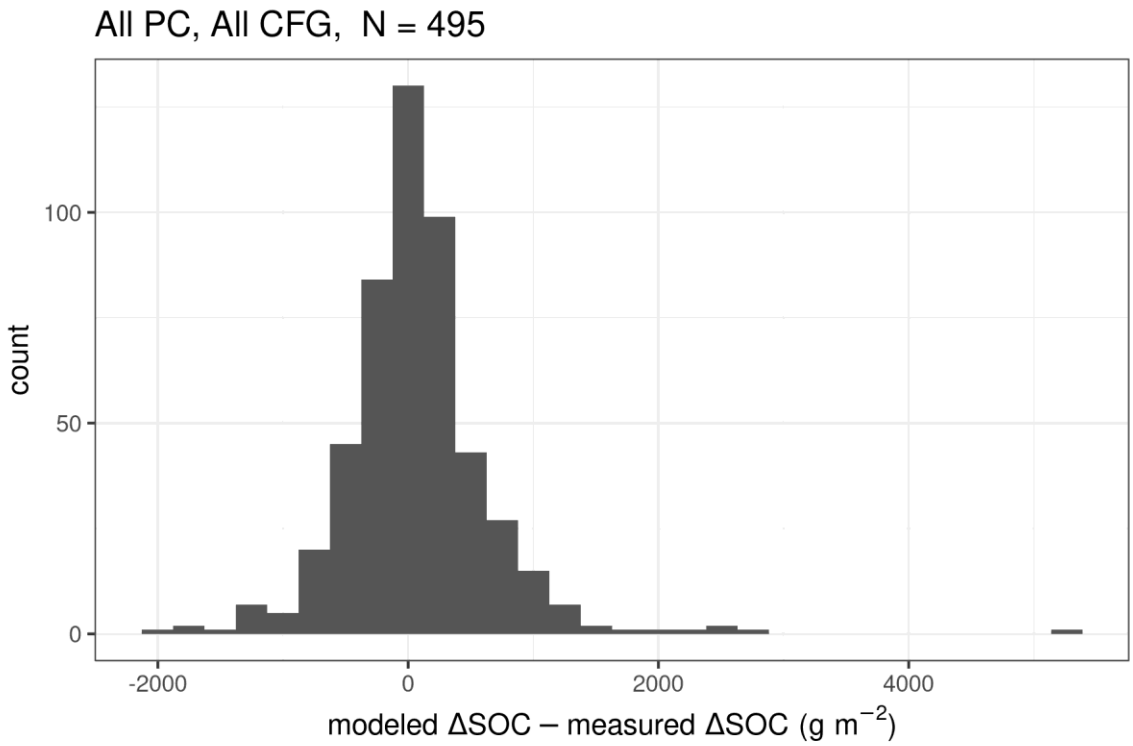

Figure 16: Histogram of model residuals (predicted - observed) for change in SOC in all studies used for model validation across all practices and crop types.

Table 30: : 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	114	114	0	100
2	179	177	2	99
3	81	75	6	93
4	53	44	9	83
5	68	55	13	81
all	**495**	**465**	**30**	**94**

Mean squared error: 354953 (g C m^{-2})^2; RMSE = 595.8 g C m^{-2}.

Do 90% of prediction intervals cover observed data? **Yes!**

CROP x Corn x SOC

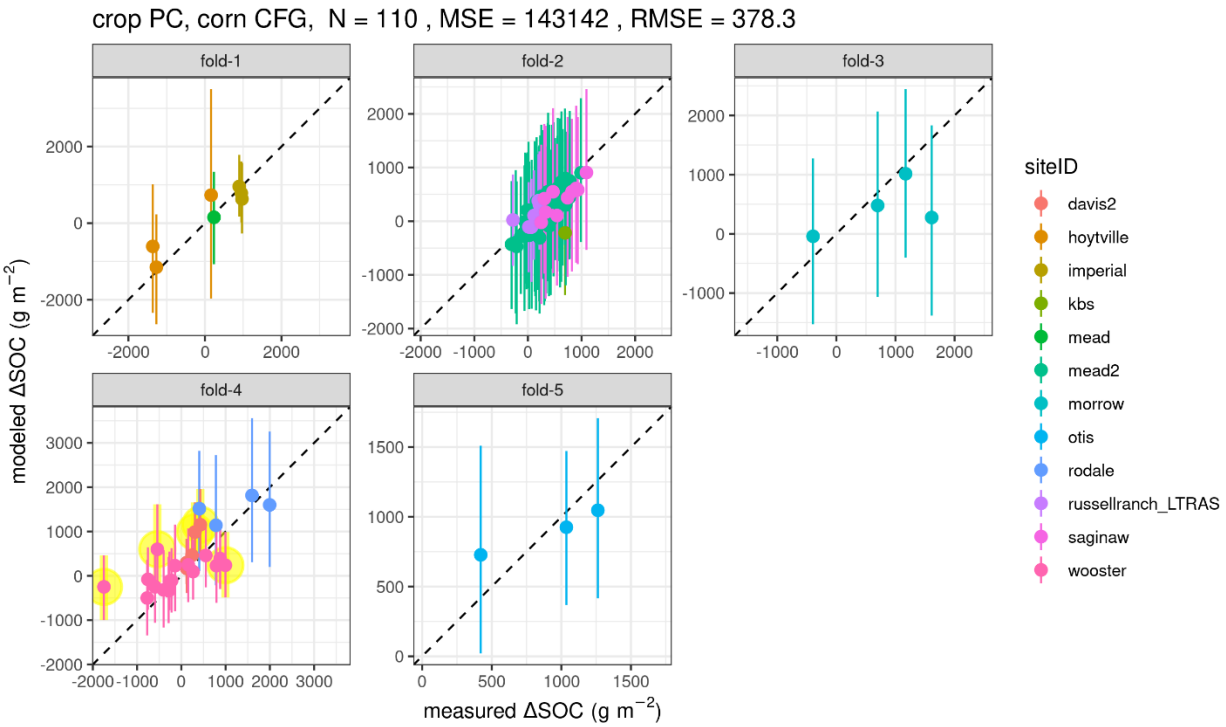

Figure 17: Model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the corn-type CFG. Error bars shows 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

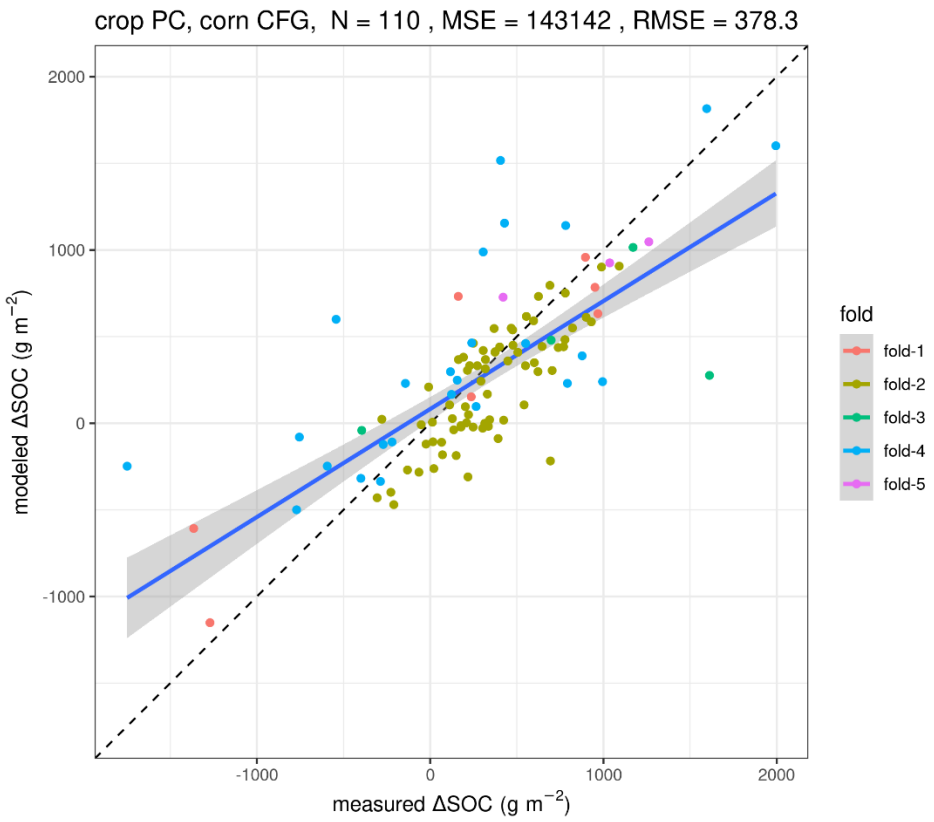

Figure 18: Scatterplot of model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the corn-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1 relationship.

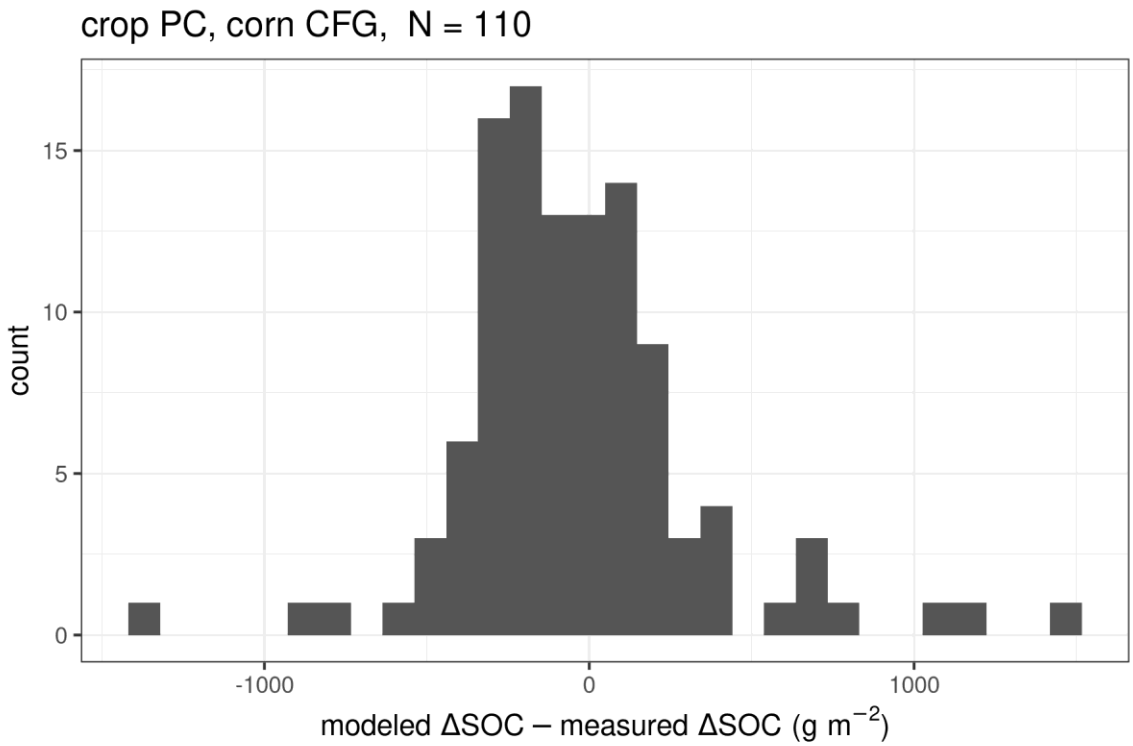

Figure 19: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed cropping practices involving crops from the corn-type CFG.

Table 31: Number of observed CROP x corn datapoints falling inside and outside of modeled 90% prediction intervals for each fold of the calibration/validation process

fold	n	n in	n out	% coverage
1	7	7	0	100
2	71	71	0	100
3	4	4	0	100
4	25	20	5	80
5	3	3	0	100
all	**110**	**105**	**5**	**95**

Mean squared error: 143142 (g C m^{-2})^2; RMSE = 378.3 g C m^{-2}.

Do 90% of prediction intervals cover observed data? **Yes!**

CROP x Soy x SOC

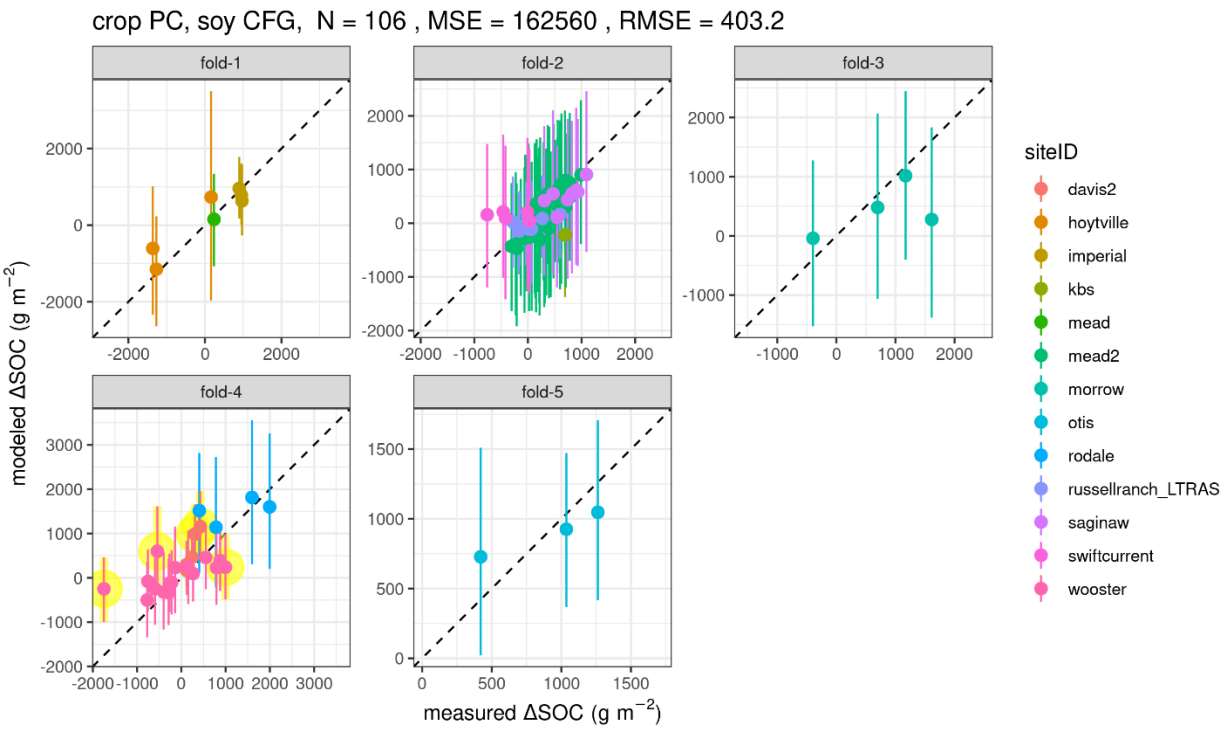

Figure 20: Model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the soy-type CFG. Error bars shows 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

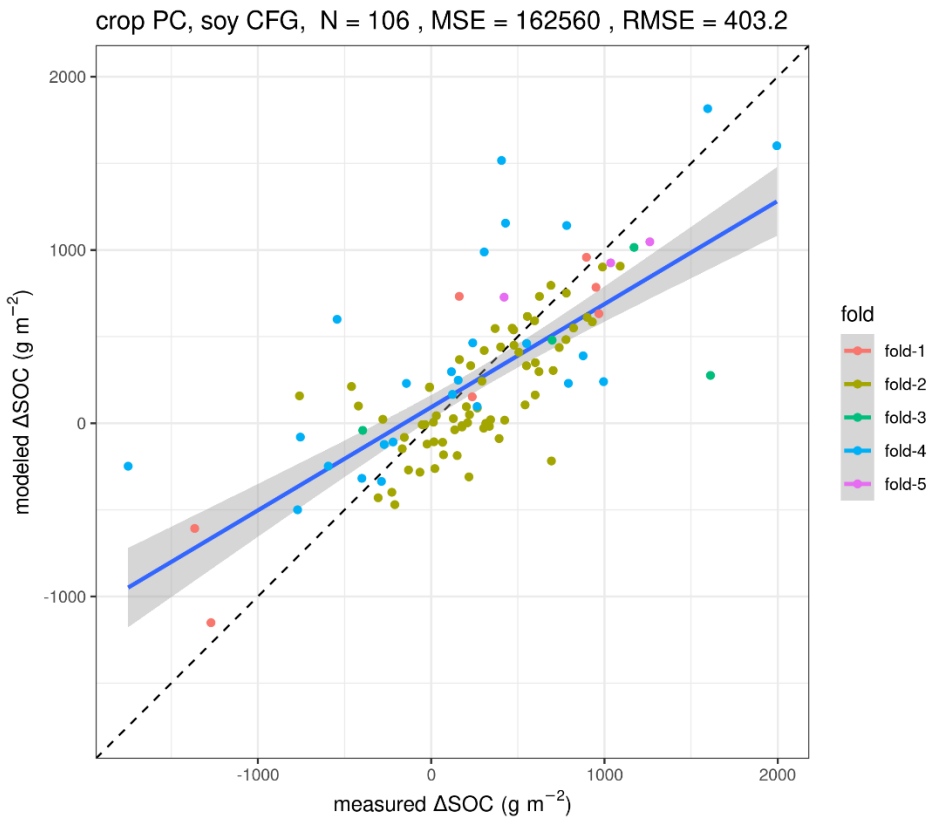

Figure 21: Scatterplot of model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the soy-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows 1:1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

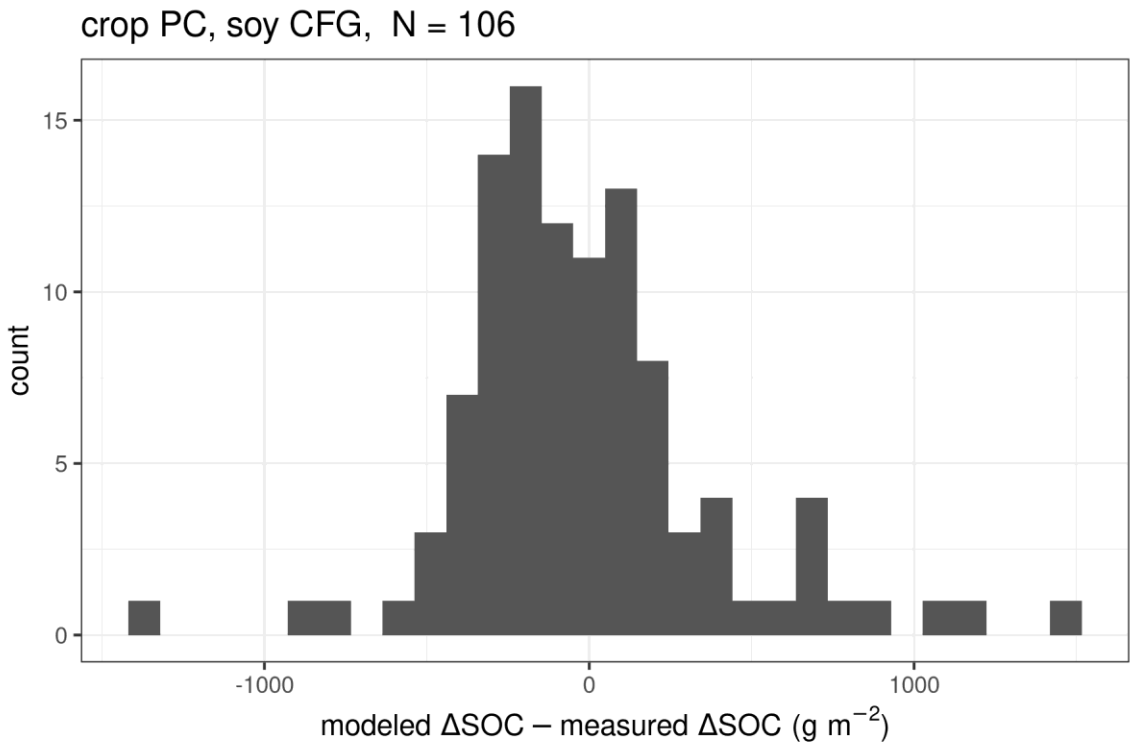

Figure 22: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed cropping practices involving crops from the soy-type CFG.

Table 32: Number of observed CROP x soy datapoints falling inside and outside of modeled 90% prediction intervals for each fold of the calibration/validation process

fold	n	n in	n out	% coverage
1	7	7	0	100
2	67	67	0	100
3	4	4	0	100
4	25	20	5	80
5	3	3	0	100
all	**106**	**101**	**5**	**95**

Mean squared error: 162560 (g C m^{-2})^2; RMSE = 403.2s g C m^{-2}.

Do 90% of prediction intervals cover observed data? **Yes!**

CROP x Wheat x SOC

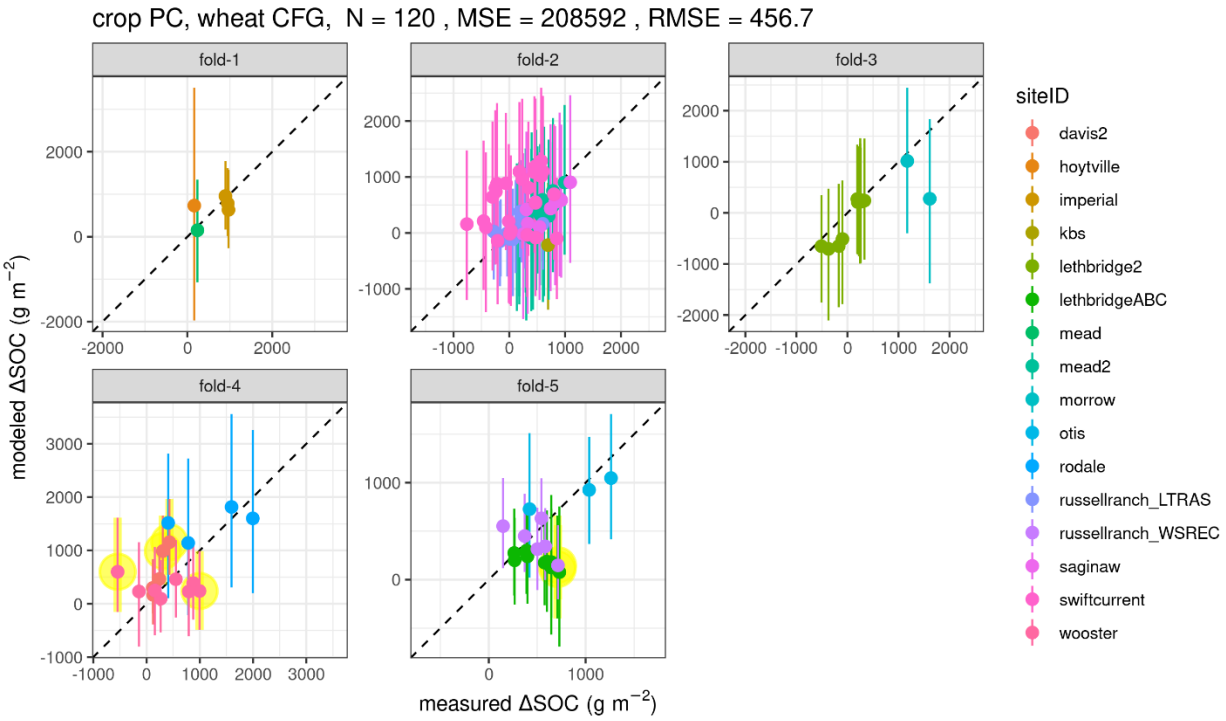

Figure 23: Model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the wheat-type CFG. Error bars shows 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

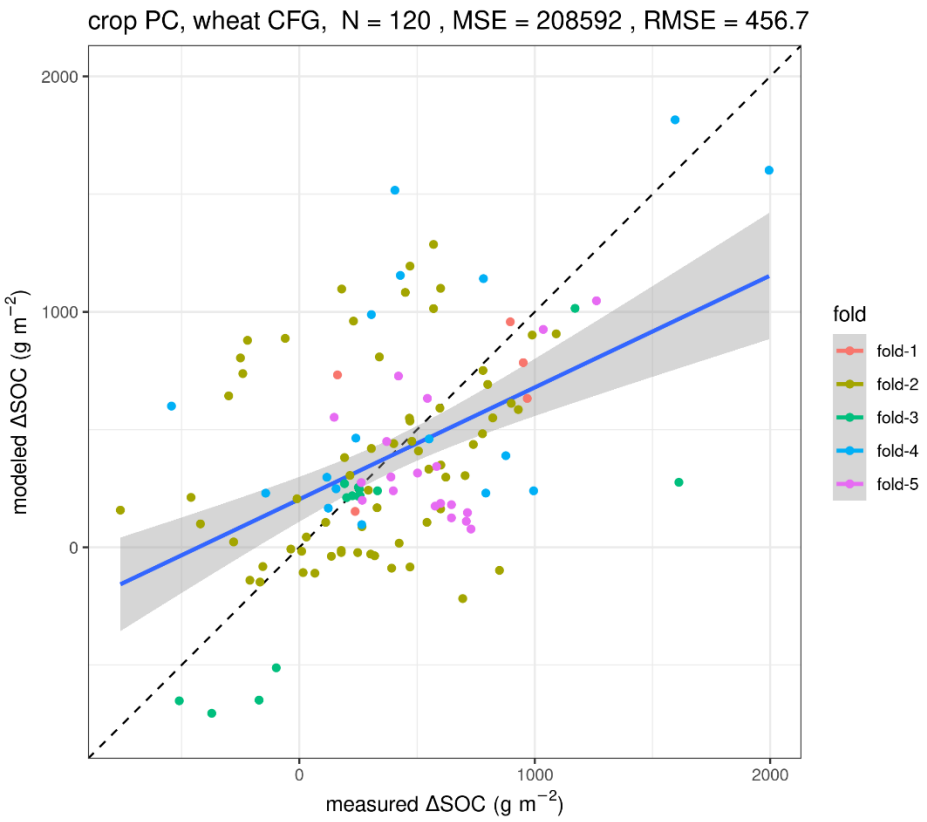

Figure 24: Scatterplot of model predictions versus measurements of SOC change in response to changed cropping practices involving crops from the wheat-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line shows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

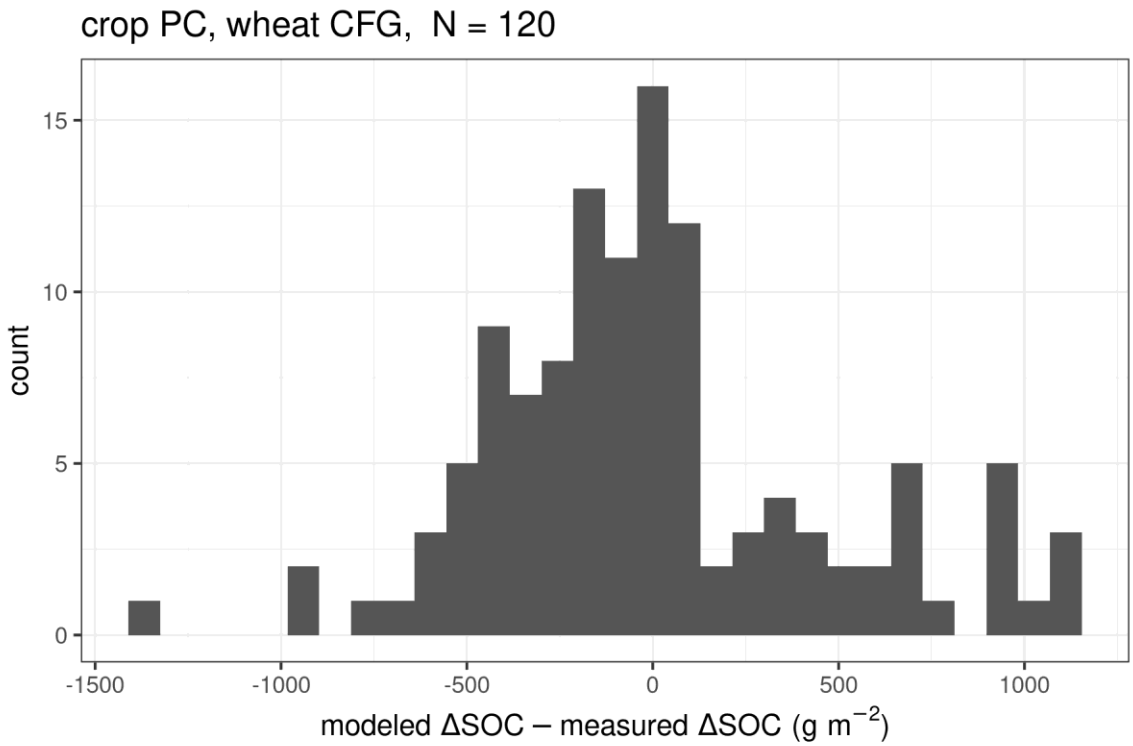

Figure 25: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed cropping practices involving crops from the wheat-type CFG.

Table 33: Number of observed CROP x cwheat datapoints falling inside and outside of modeled 90% prediction intervals for each fold of the calibration/validation process

fold	n	n in	n out	% coverage
1	5	5	0	100
2	65	65	0	100
3	14	14	0	100
4	17	13	4	76
5	19	17	2	89
all	**120**	**114**	**6**	**95**

Mean squared error: 162560 (g C m^{-2})^2; RMSE = 403.2s g C m^{-2}.

Do 90% of prediction intervals cover observed data? **Yes!**

DISTURB x Corn x SOC

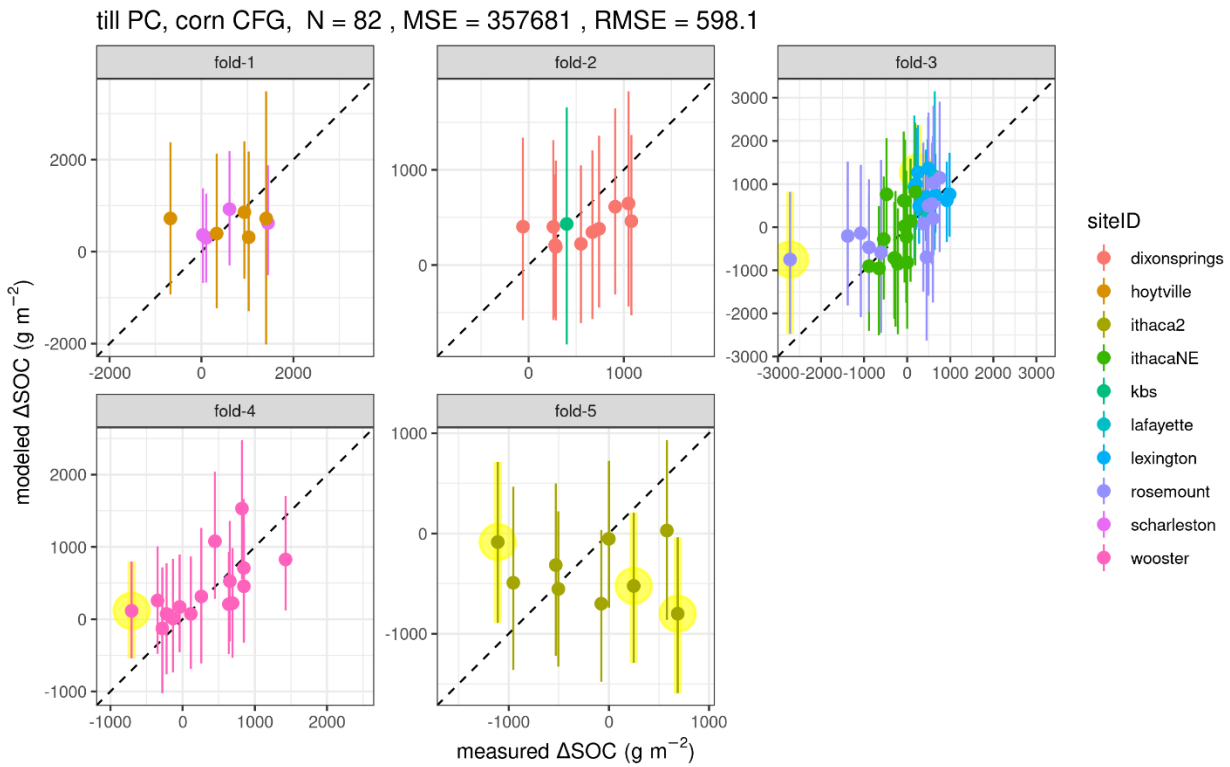

Figure 26: Model predictions versus measurements of SOC change in response to changed tillage or residue management practices involving crops from the corn-type CFG. Error bars shows 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

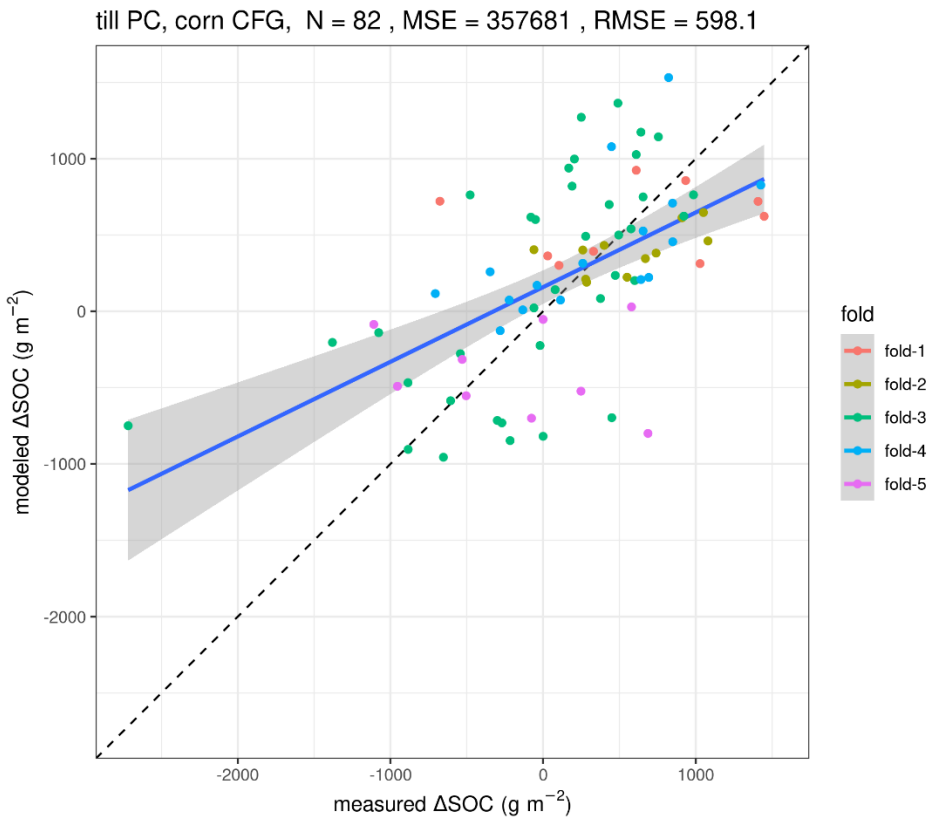

Figure 27: Scatterplot of model predictions versus measurements of SOC change in response to changed tillage or residue management practices involving crops from the corn-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error).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

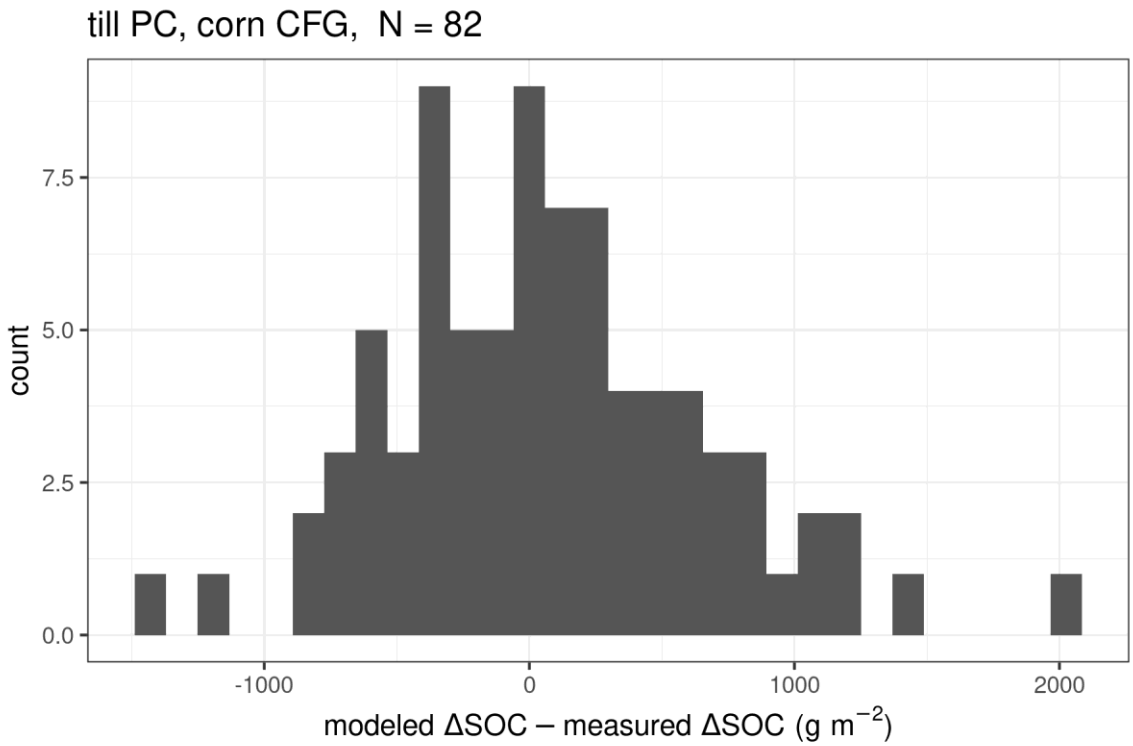

Figure 28: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed tillage or residue management practices involving crops from the corn-type CFG.

Table 34: Number of observed DISTURB x corn datapoints falling inside and outside of modeled 90% prediction intervals for each fold of the calibration/validation process

fold	n	n in	n out	% coverage
1	9	9	0	100
2	11	11	0	100
3	37	35	2	95
4	16	15	1	94
5	9	6	3	67
all	**82**	**76**	**6**	**93**

Mean squared error: 357681 (g C m^{-2})^2; RMSE = 598.1 g C m^{-2}.

Do 90% of prediction intervals cover observed data? **Yes!**

DISTURB x Soy x SOC

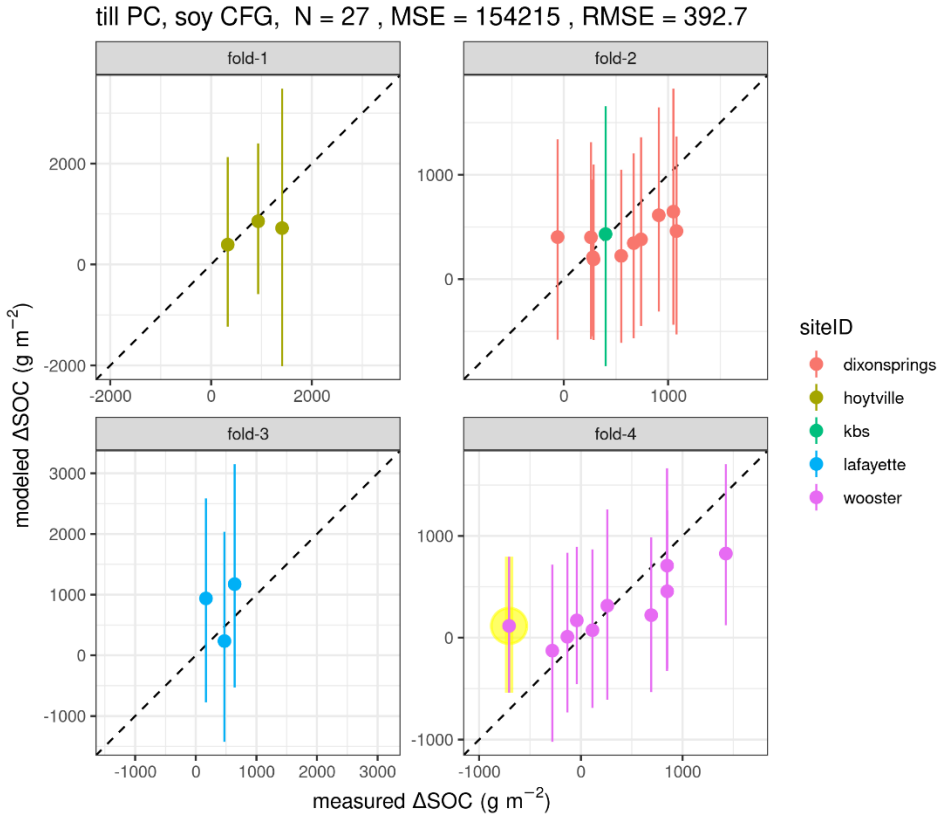

Figure 29: Model predictions versus measurements of SOC change in response to changed tillage or residue management practices involving crops from the soy-type CFG. Error bars shows 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

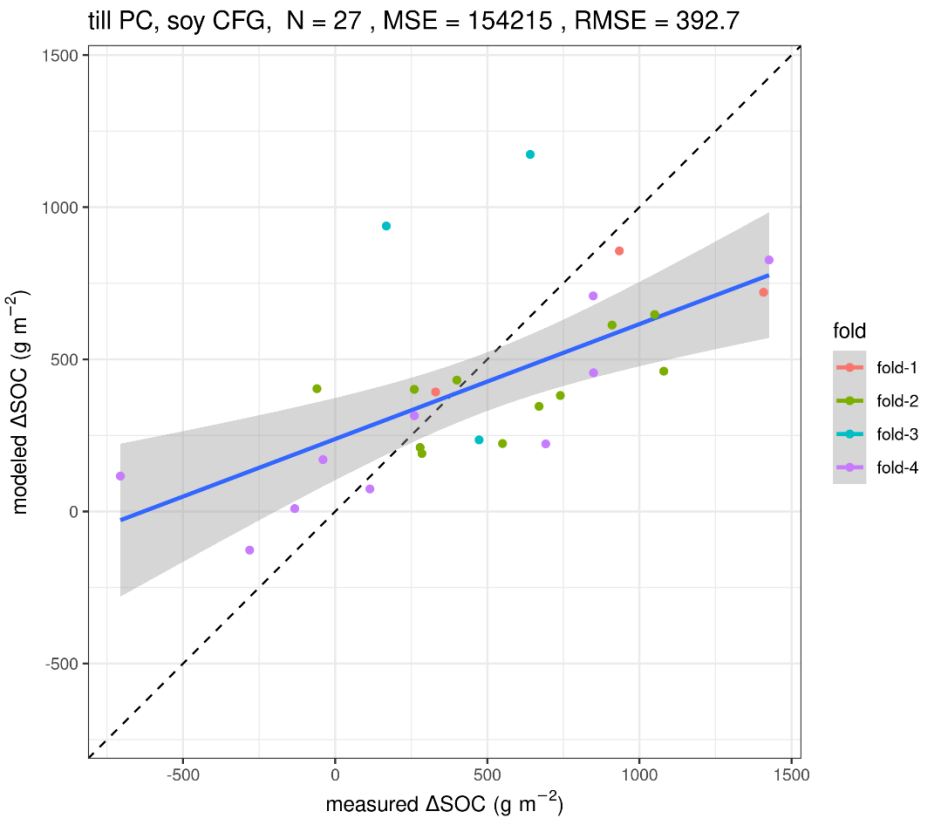

Figure 30: Scatterplot of model predictions versus measurements of SOC change in response to changed tillage or residue management practices involving crops from the soy-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error).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

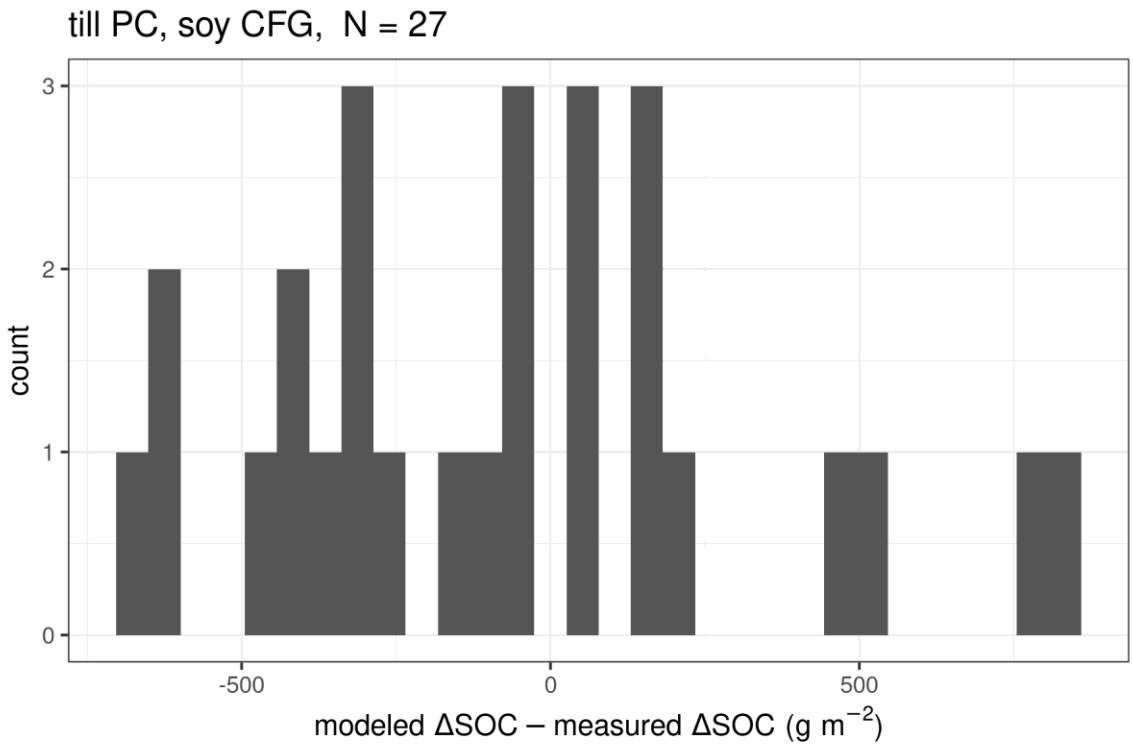

Figure 31: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed tillage or residue management practices involving crops from the soy-type CFG.

Table 35: Number of observed DISTURB x soy datapoints falling inside and outside of modeled 90% prediction intervals for each fold of the calibration/validation process

fold	n	n in	n out	% coverage
1	3	3	0	100
2	11	11	0	100
3	3	3	0	100
4	10	9	1	90
5	0	0	0	-
all	**27**	**26**	**1**	**96**

Mean squared error: 154215 (g C m^{-2})^2; RMSE = 392.7 g C m^{-2}.

Do 90% of prediction intervals cover observed data? **Yes!**

DISTURB x Wheat x SOC

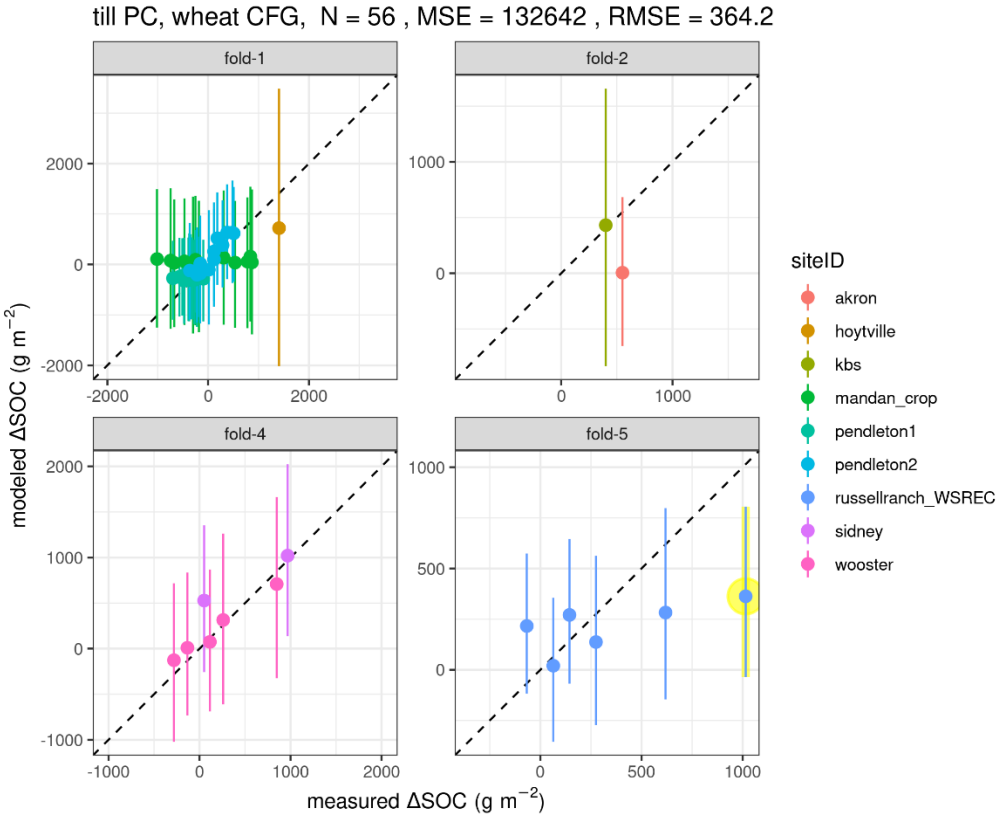

Figure 32: Model predictions versus measurements of SOC change in response to changed tillage or residue management practices involving crops from the wheat-type CFG. Error bars shows 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

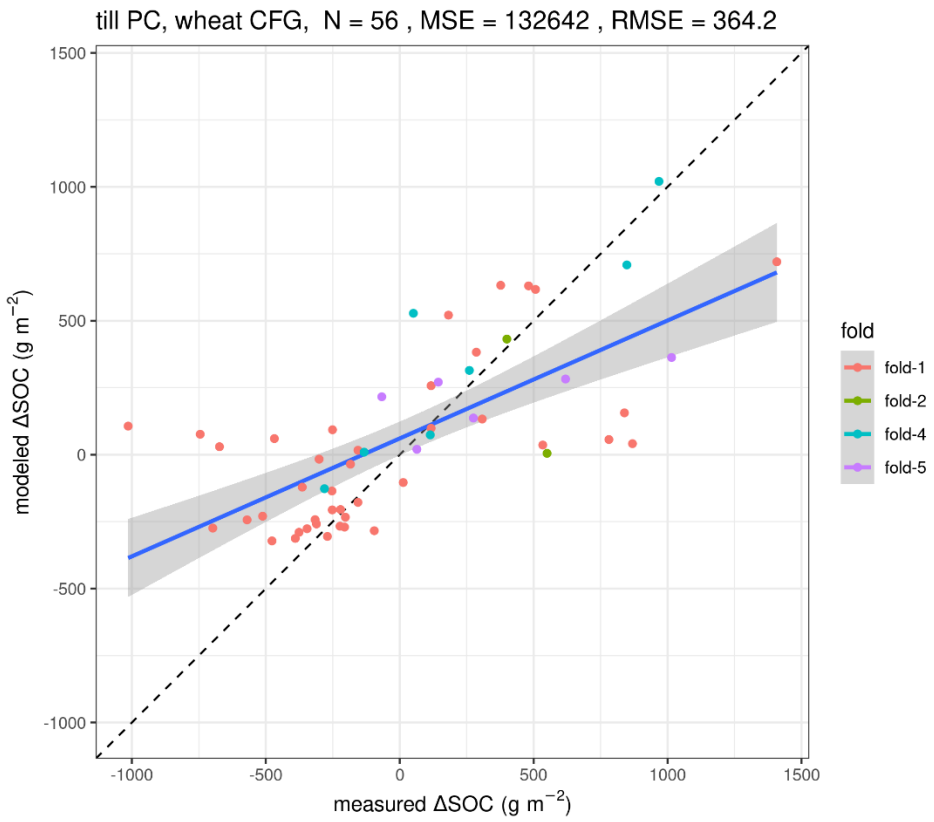

Figure 33: Scatterplot of model predictions versus measurements of SOC change in response to changed tillage or residue management practices involving crops from the wheat-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error).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

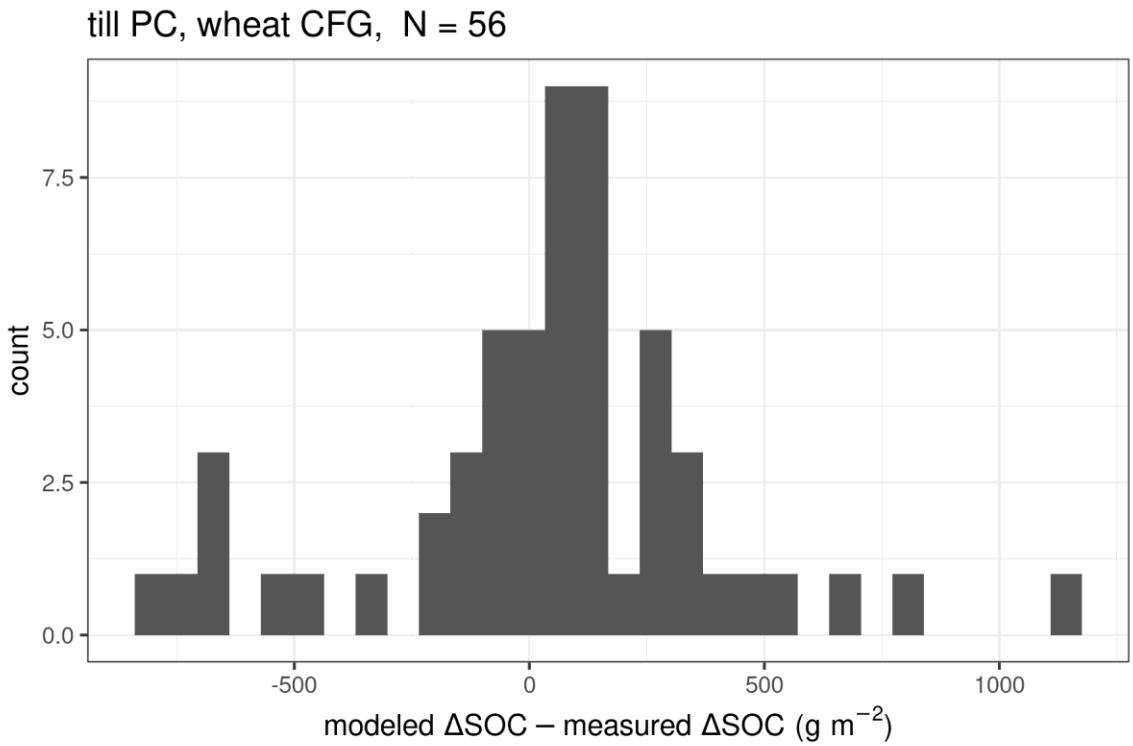

Figure 34: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed tillage or residue management practices involving crops from the wheat-type CFG.

Table 36: Number of observed DISTURB x wheat datapoints falling inside and outside of modeled 90% prediction intervals for each fold of the calibration/validation process

fold	n	n in	n out	% coverage
1	41	41	0	100
2	2	2	0	100
3	0	0	0	-
4	7	7	0	100
5	6	5	1	83
all	**56**	**55**	**1**	**98**

Mean squared error: 132642 (g C m^{-2})^2; RMSE = 364.2 g C m^{-2}.

Do 90% of prediction intervals cover observed data? **Yes!**

NFERT x Corn x SOC

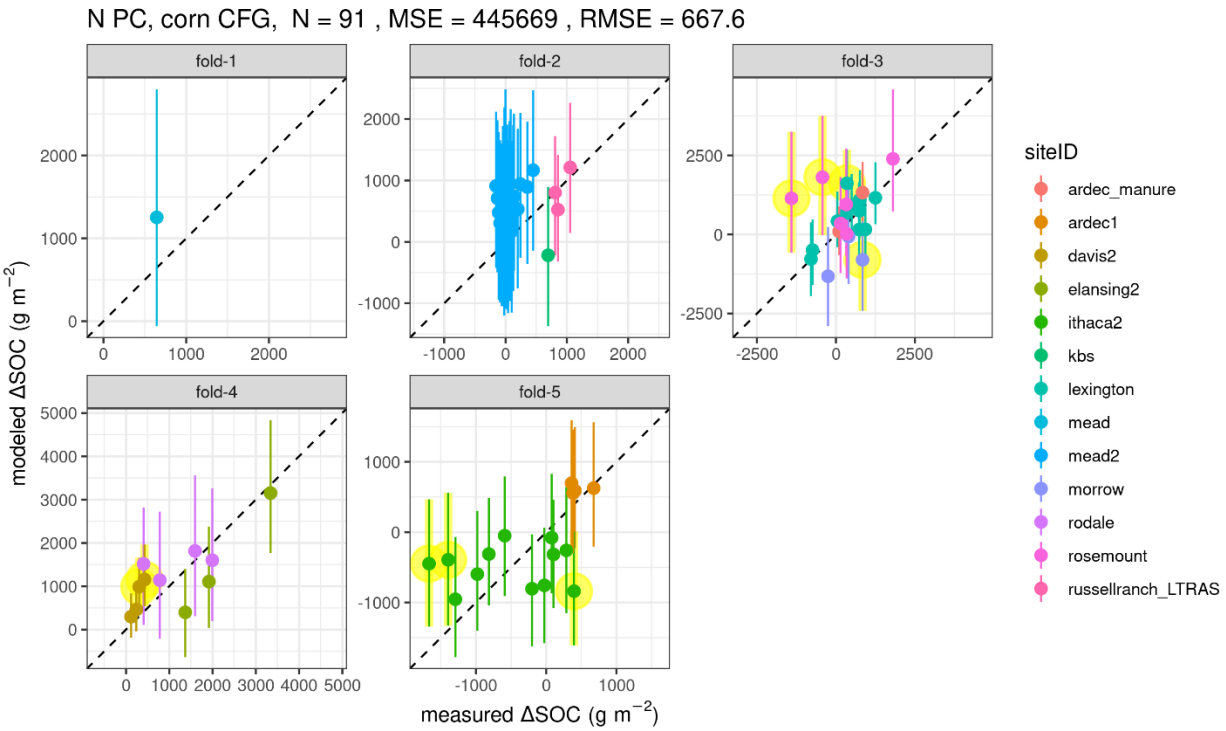

Figure 35: Model predictions versus measurements of SOC change in response to changed inorganic N fertilization practices involving crops from the corn-type CFG. Error bars shows 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

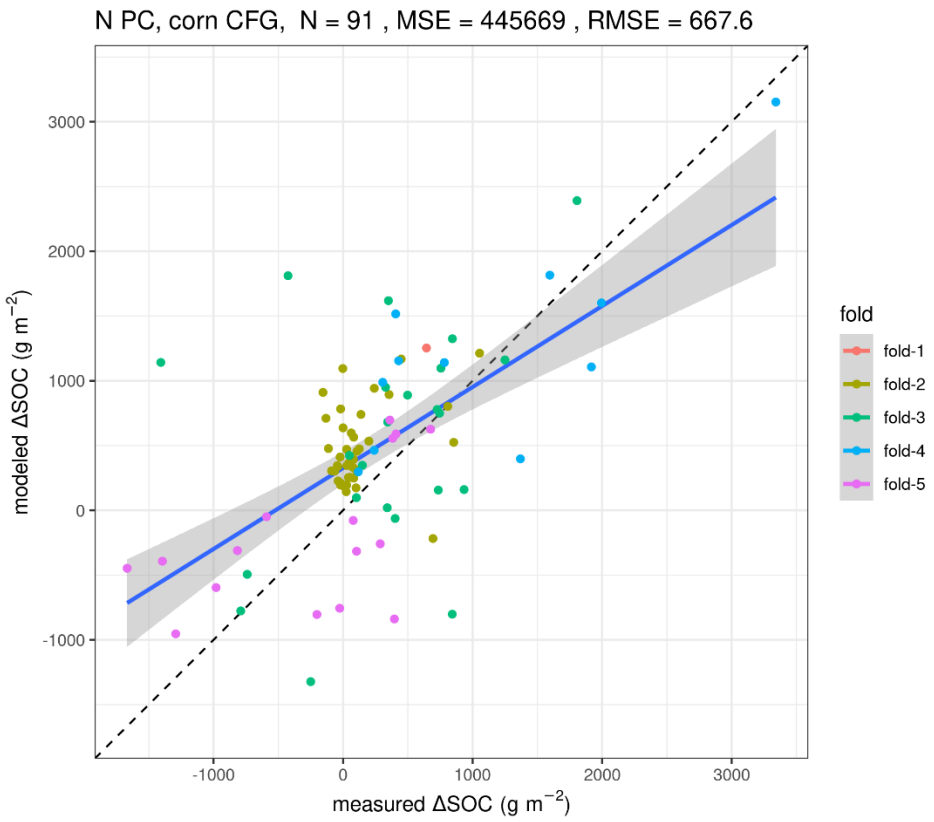

Figure 36: Scatterplot of model predictions versus measurements of SOC change in response to changed inorganic N fertilization practices involving crops from the corn-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed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

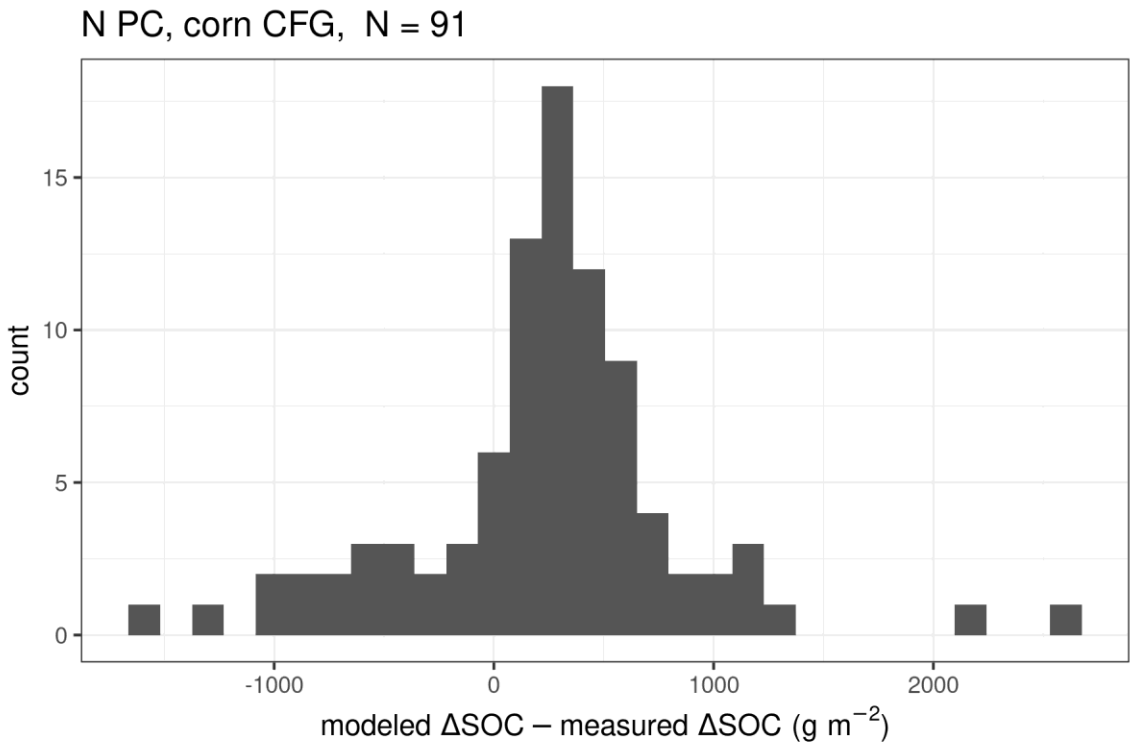

Figure 37: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed inorganic N fertilization practices involving crops from the corn-type CFG.

Table 37: Number of observed NFERT x corn datapoints falling inside and outside of modeled 90% prediction intervals for each fold of the calibration/validation process

fold	n	n in	n out	% coverage
1	1	1	0	100
2	40	40	0	100
3	23	19	4	83
4	11	9	2	82
5	16	13	3	81
all	**91**	**82**	**9**	**90**

Mean squared error: 445669 (g C m^{-2})^2; RMSE = 667.6 g C m^{-2}.

Do 90% of prediction intervals cover observed data? **Yes!**

NFERT x Wheat x SOC

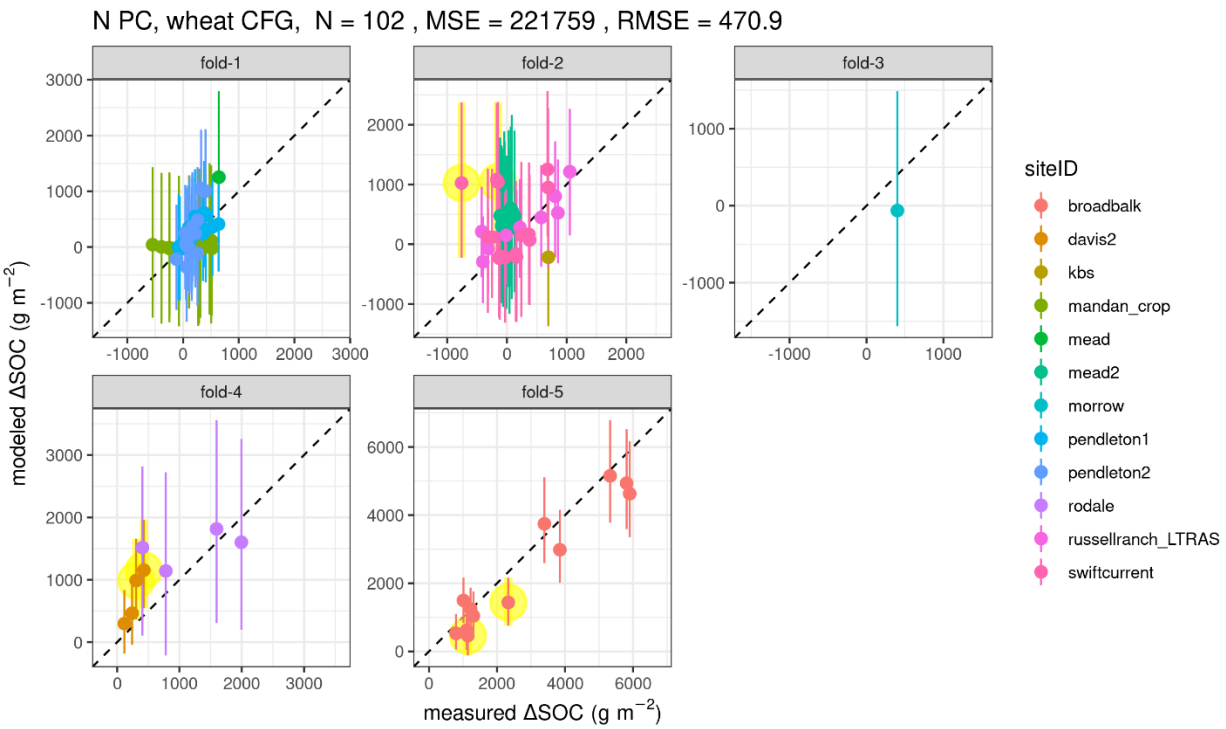

Figure 38: Model predictions versus measurements of SOC change in response to changed inorganic N fertilization practices involving crops from the wheat-type CFG. Error bars shows 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

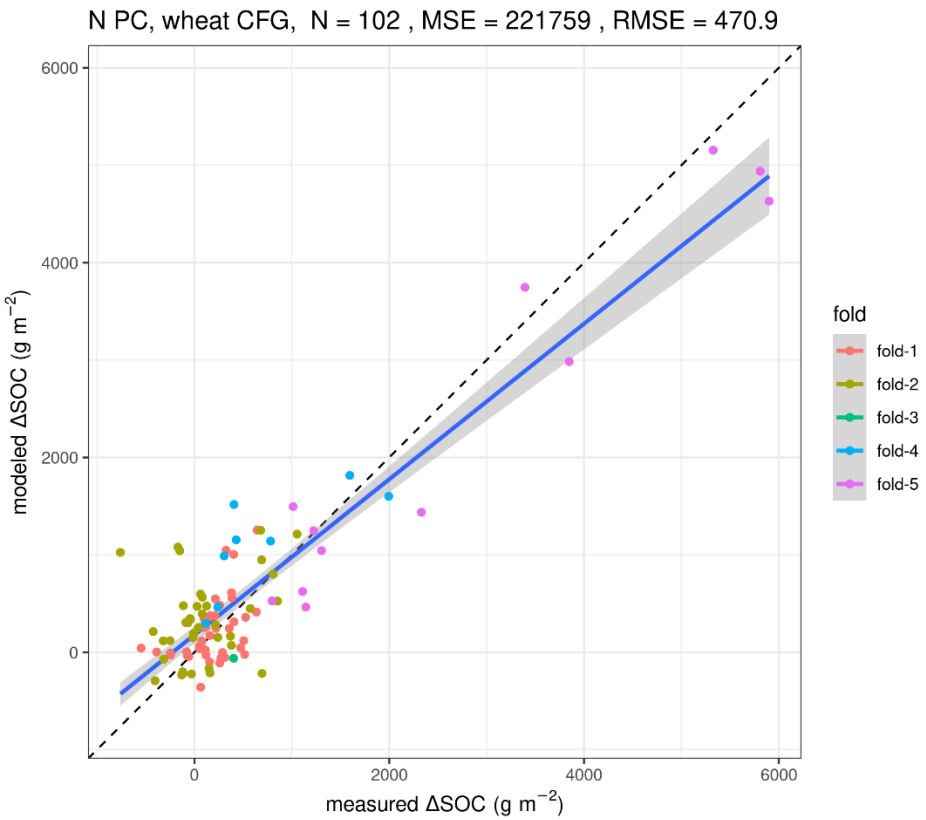

Figure 39: Scatterplot of model predictions versus measurements of SOC change in response to changed inorganic N fertilization practices involving crops from the wheat-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error).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

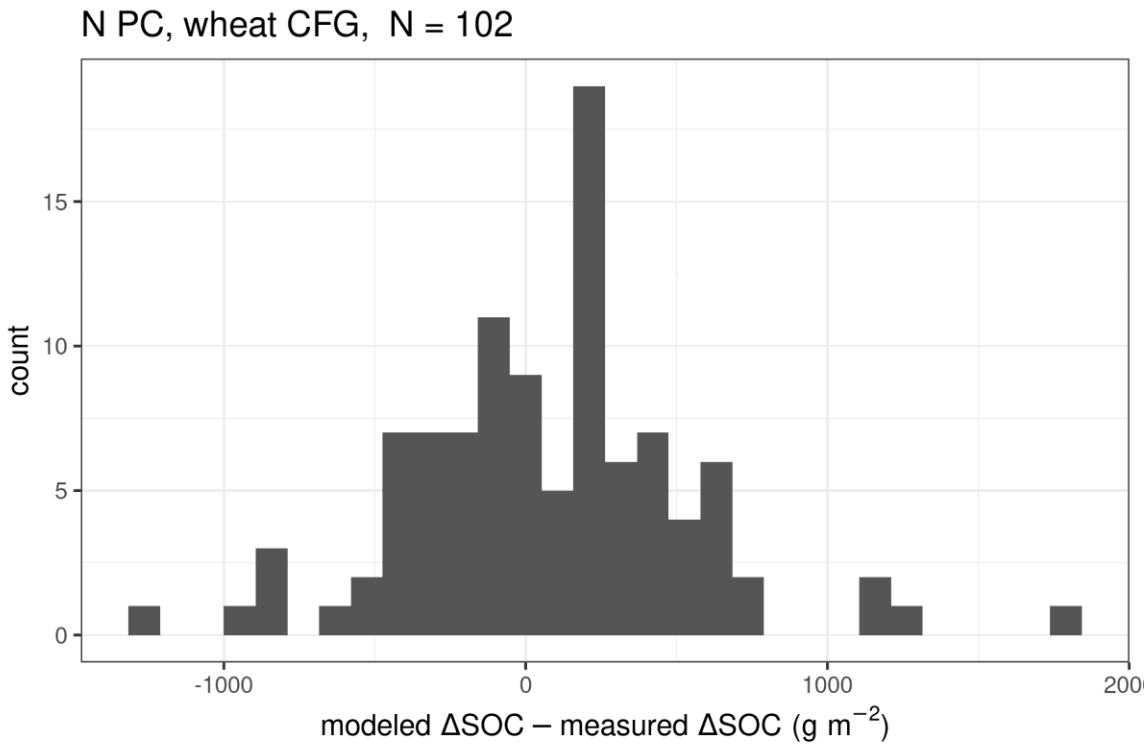

Figure 40: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed inorganic N fertilization practices involving crops from the wheat-type CFG.

Table 38: Number of observed NFERT x wheat datapoints falling inside and outside of modeled 90% prediction intervals for each fold of the calibration/validation process

fold	n	n in	n out	% coverage
1	44	44	0	100
2	37	35	2	95
3	1	1	0	100
4	8	6	2	75
5	12	10	2	83
all	**102**	**96**	**6**	**94**

Mean squared error: 221759 (g C m^{-2})^2; RMSE = 470.9 g C m^{-2}.

Do 90% of prediction intervals cover observed data? **Yes!**

ORG x Wheat x SOC

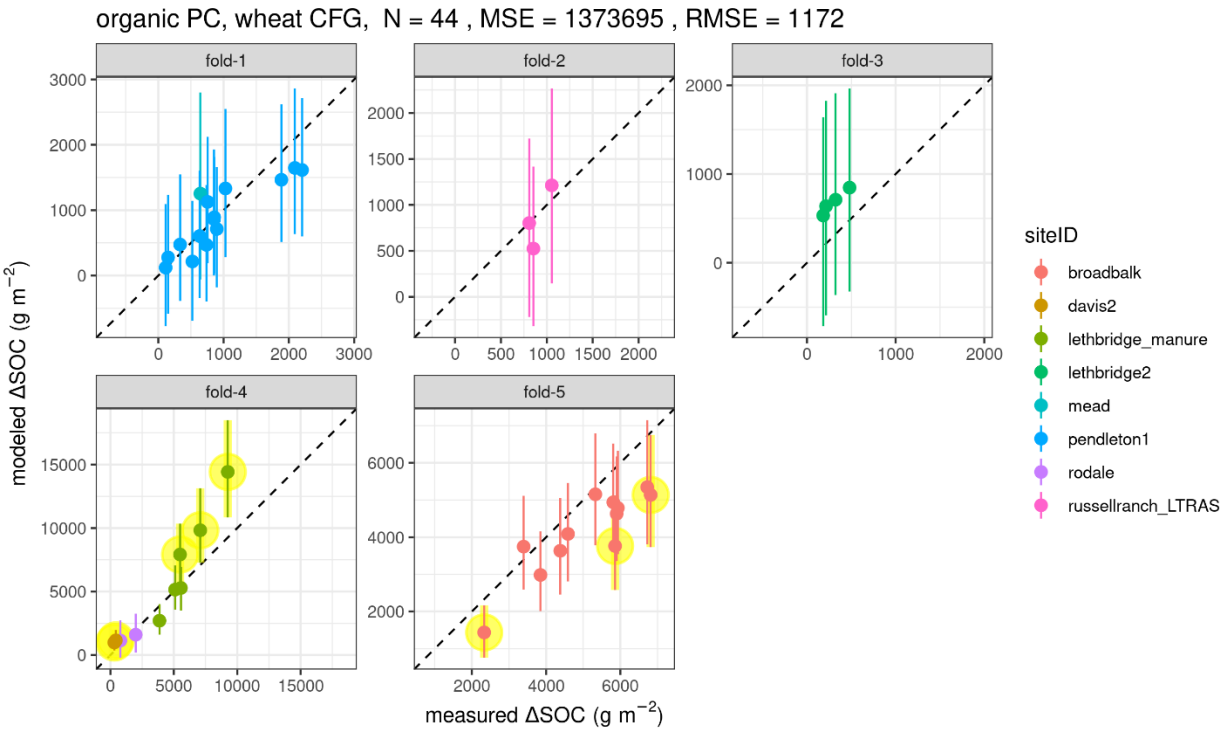

Figure 41: Model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the wheat-type CFG. Error bars shows 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

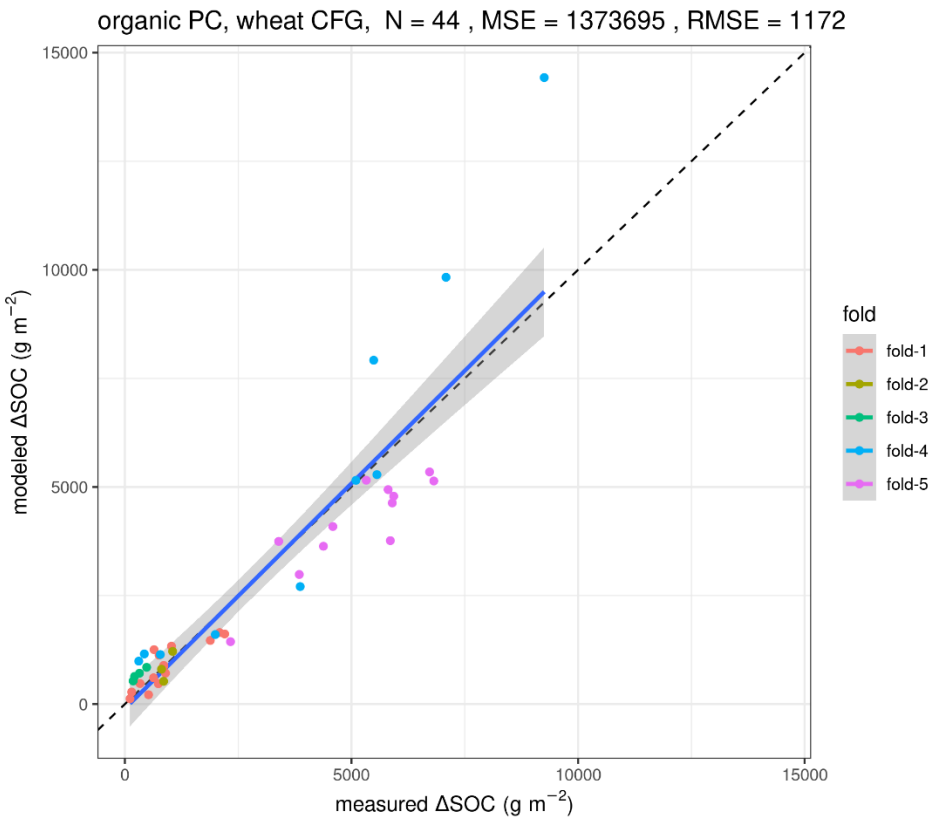

Figure 42: Scatterplot of model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the wheat-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line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

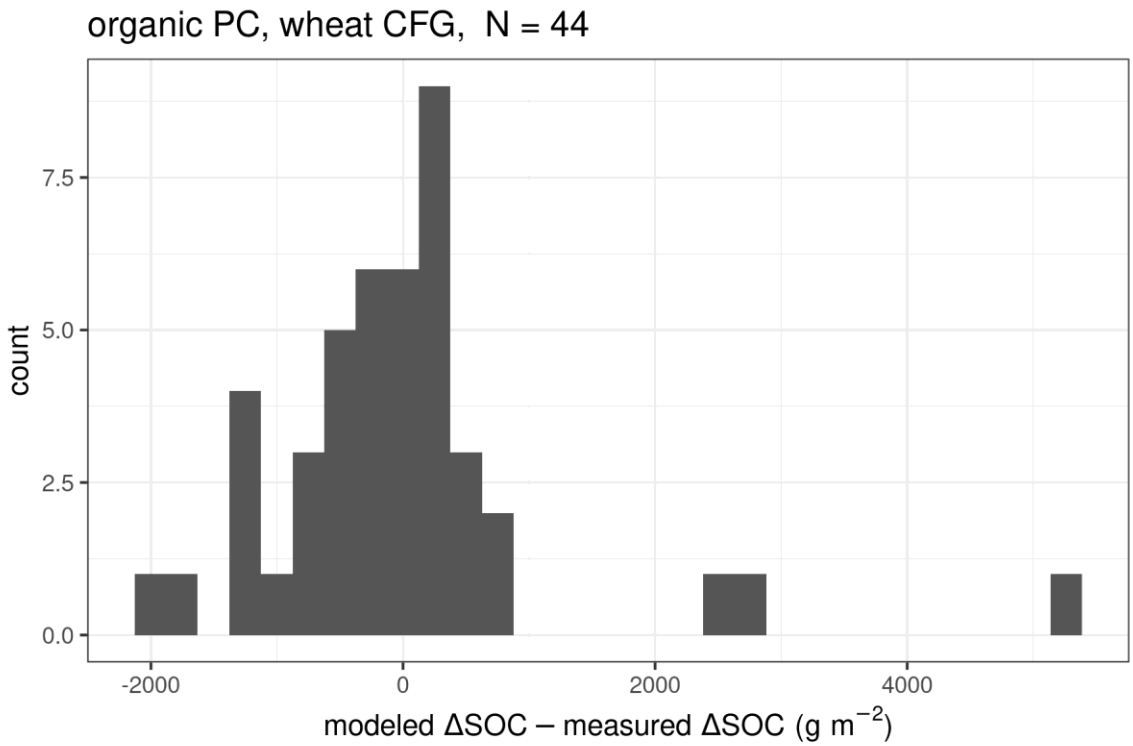

Figure 43: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed organic amendment practices involving crops from the wheat-type CFG.

Table 39: Number of observed ORG x wheat datapoints falling inside and outside of modeled 90% prediction intervals for each fold of the calibration/validation process

fold	n	n in	n out	% coverage
1	15	15	0	100
2	3	3	0	100
3	4	4	0	100
4	10	5	5	50
5	12	9	3	75
all	**44**	**36**	**8**	**82**

Mean squared error: 1373695 (g C m^{-2})^2; RMSE = 1172 g C m^{-2}.

Do 90% of prediction intervals cover observed data? **No, only 36 of 44 observations (82%) fall inside the model's 90% prediction intervals.**

The low interval coverage in this category appears to be driven primarily by the model underestimating variability for cases where the total change in SOC is very large (>5000 g C m^{-2}; Figure 41). Only two sites in the validation dataset (Lethbridge and Broadbalk) show changes this large, each for very different reasons. The Lethbridge manure site gained SOC rapidly (26 years) because of very large annual manure inputs (up to180 Mg manure ha^{-1} yr^{-1}, which is three times the maximum agronomic recommendation; Hao et al. 2003) , while the gains at Broadbalk occurred at slower rates over a longer time (more than 140 years) in response to more

agronomically typical manure rates (35 Mg manure ha^{-1} yr^{-1}; Jenkinson, 1991). Model performance under each of these scenarios is likely to be sensitive to different calibration parameters, so it is perhaps not surprising that models calibrated with information from only one of these sites will perform poorly on the other one, and reassuringly the model appears to perform better when informed by both: when considering all sites, average model bias is less than PMU (Figure 42;Table 26).

Because the predictive uncertainty appears to be adequate for all but dSOC far higher than will be observed in most crediting projects, and because we expect that these high-dSOC sites will be very useful for constraining the calibration of SOC turnover parameters in the presence of more moderate organic amendment rates, **we propose to consider the model *conditionally validated* for ORG x Wheat X SOC, with a requirement that it never be used for crediting in cases where the total dSOC may exceed 5000 g m^{-2}.** We discuss this proposal more below in the section **"**Restrictions on application of model**."**

29 of the 32 observations with observed dSOC below this threshold overlap the model's predictive intervals, for an inclusion rate of 91%, and the mean model bias from these is –20 compared to +291 when the high-dSOC observations are included.

ORG x Corn x SOC

Reported to support ORG x all crops x SOC following the approach in Appendix D: Proposal for validating organic amendment applications

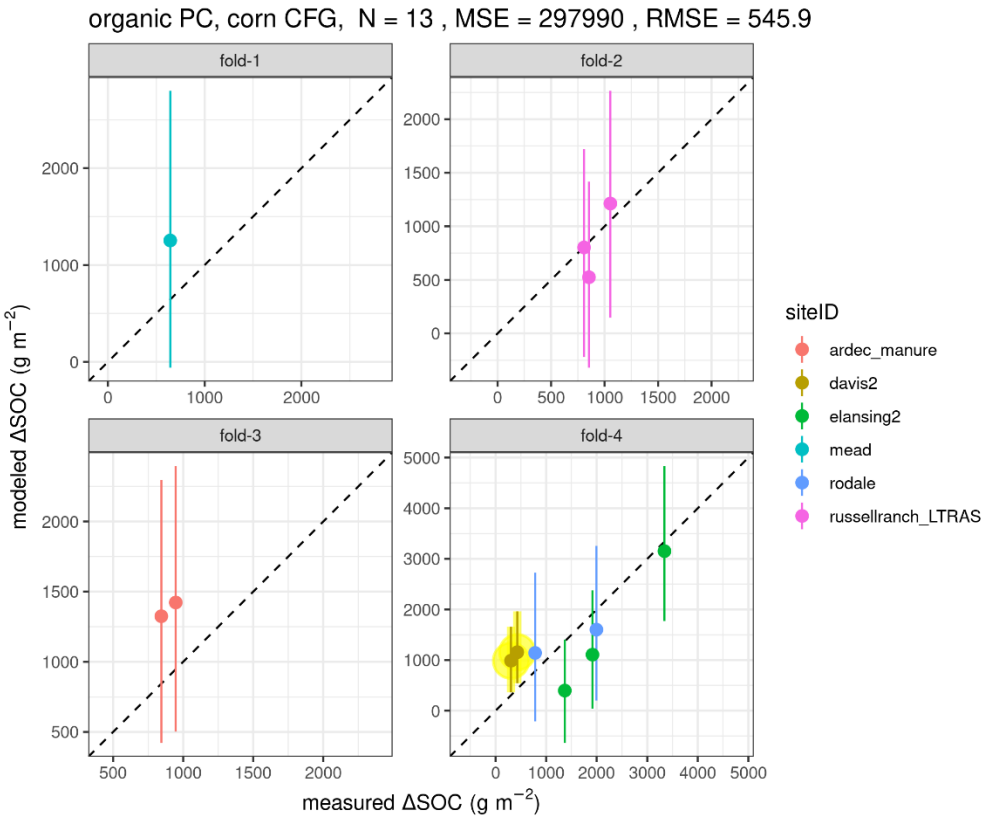

Figure 44: Model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the corn-type CFG. Error bars shows 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

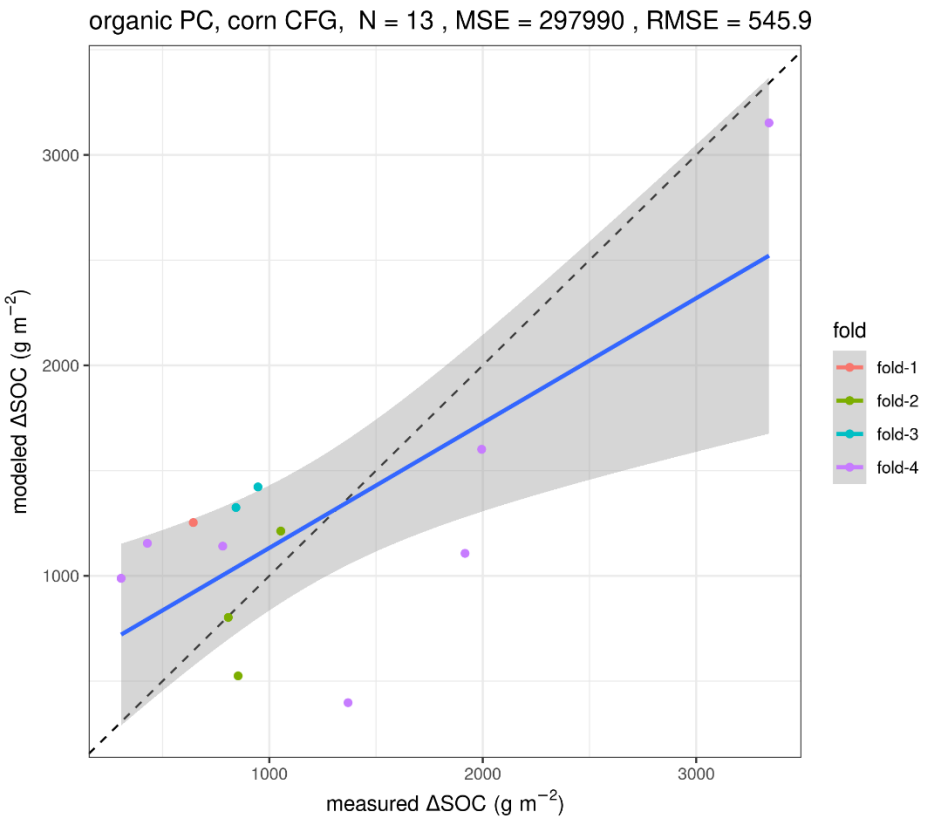

Figure 45: Scatterplot of model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the corn-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line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

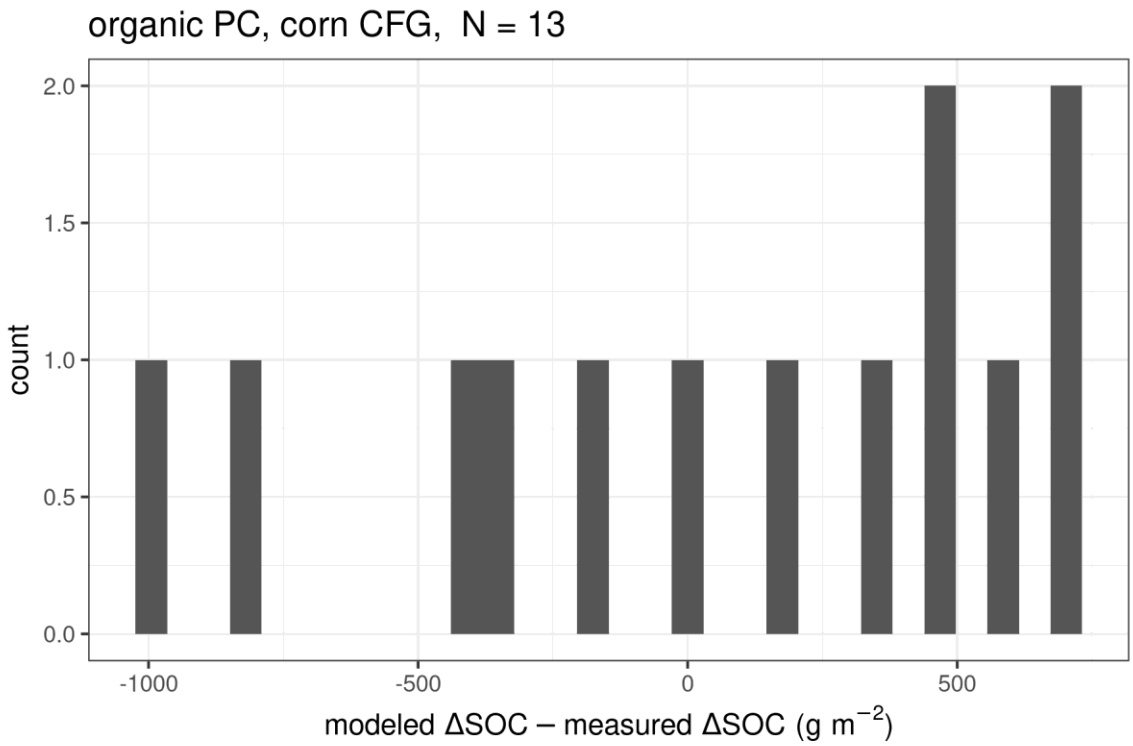

Figure 46: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed organic amendment practices involving crops from the corn-type CFG.

Table 40: Number of observed ORG x corn datapoints falling inside and outside of modeled 90% prediction intervals for each fold of the calibration/validation process

fold	n	n in	n out	% coverage
1	1	1	0	100
2	3	3	0	100
3	2	2	0	100
4	7	5	2	71
5	0	0	0	-
all	**13**	**11**	**2**	**85**

Mean squared error: 297990 (g C m^{-2})^2; RMSE = 546 g C m^{-2}.

Do 90% of prediction intervals cover observed data? **No, only 11 of 13 observations (85%) fall inside the model's 90% prediction intervals,** though given the small sample size this could be seen as binomial error: 11/13 is less than one observation short of 90% (i.e 12/13 = 92%), and if each point truly has a 90% chance of containing the observation then a set of 13 points would be expected to cover 11 or fewer observations about pbinom(q=11, size=13, prob=0.9) ~= 38% of the time.

ORG x Soy x SOC

Reported to support ORG x all crops x SOC following the approach in Appendix D: Proposal for validating organic amendment applications

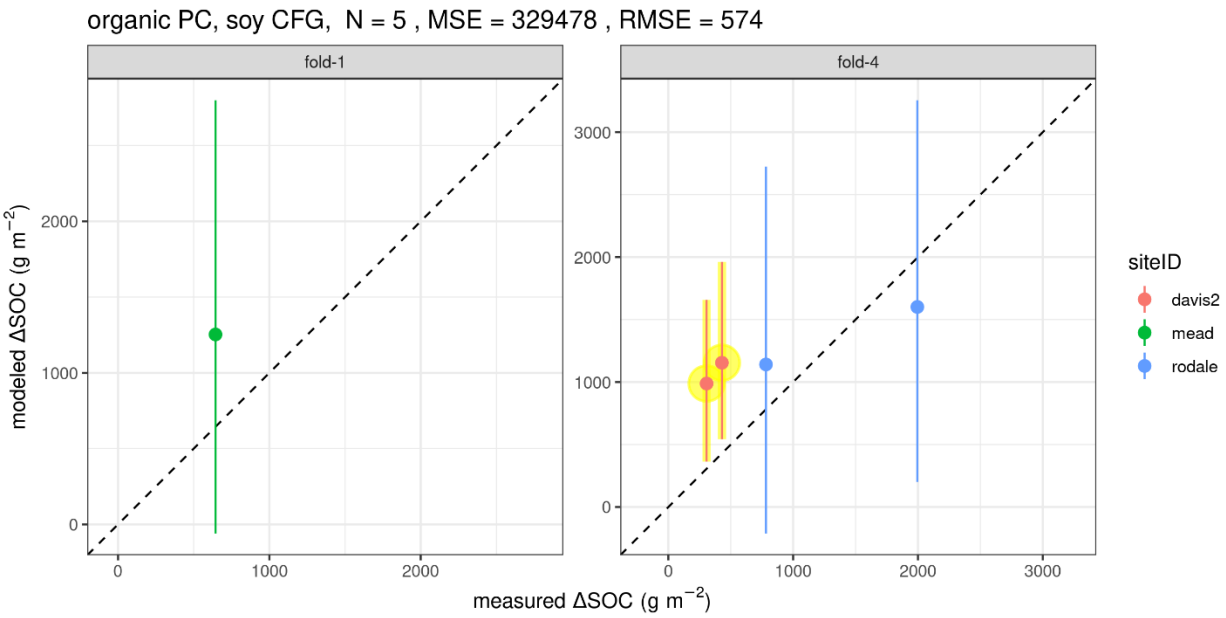

Figure 47: Model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the soy-type CFG. Error bars shows 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

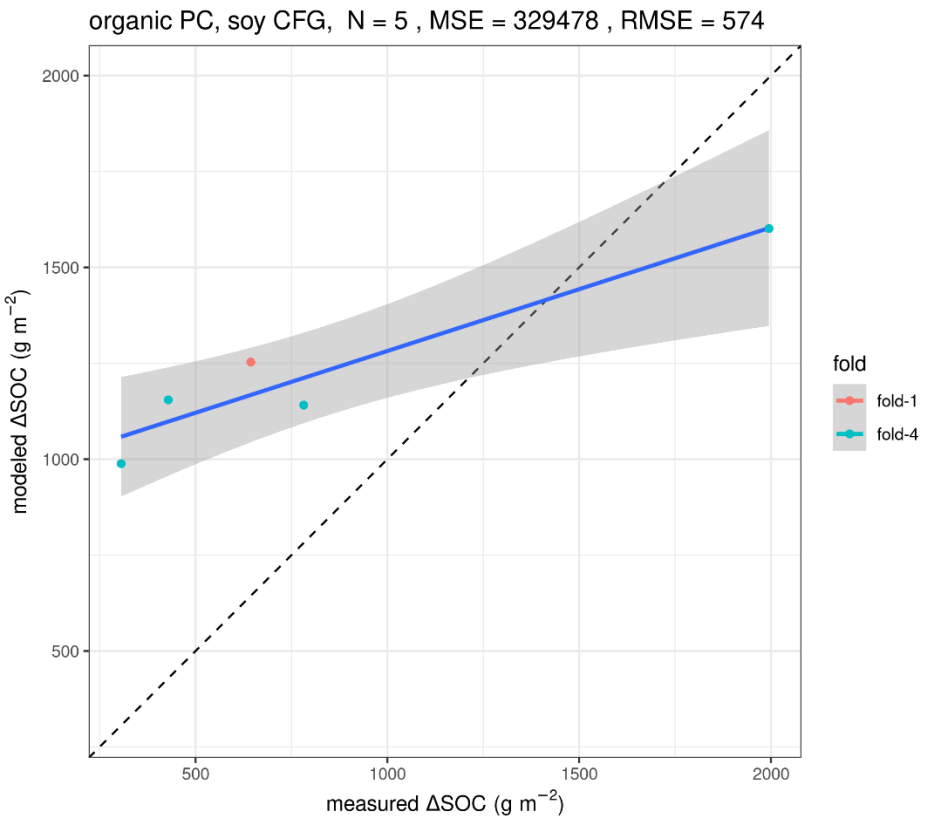

Figure 48: Scatterplot of model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the soy-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line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

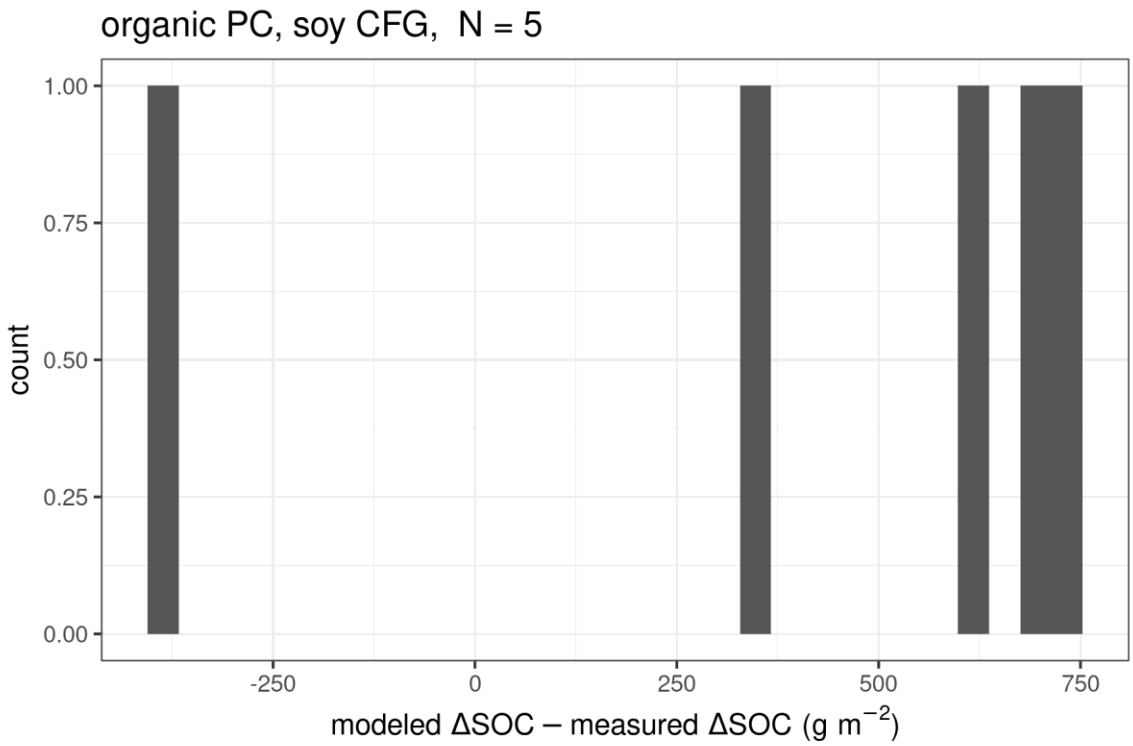

Figure 49: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed organic amendment practices involving crops from the soy-type CFG.

Table 41: Number of observed ORG x soy datapoints falling inside and outside of modeled 90% prediction intervals for each fold of the calibration/validation process

fold	n	n in	n out	% coverage
1	1	1	0	100
2	0	0	0	-
3	0	0	0	-
4	4	2	2	50
5	0	0	0	-
all	**5**	**3**	**2**	**60**

Mean squared error: 329478 (g C m^{-2})^2; RMSE = 574 g C m^{-2}.

Do 90% of prediction intervals cover observed data? **No, only 3 of 5 observations (60%) fall inside the model's 90% prediction intervals**, though clearly this is too small a dataset to draw any robust conclusions.

ORG x all crops x SOC

Reported following the approach in Appendix D: Proposal for validating organic amendment applications

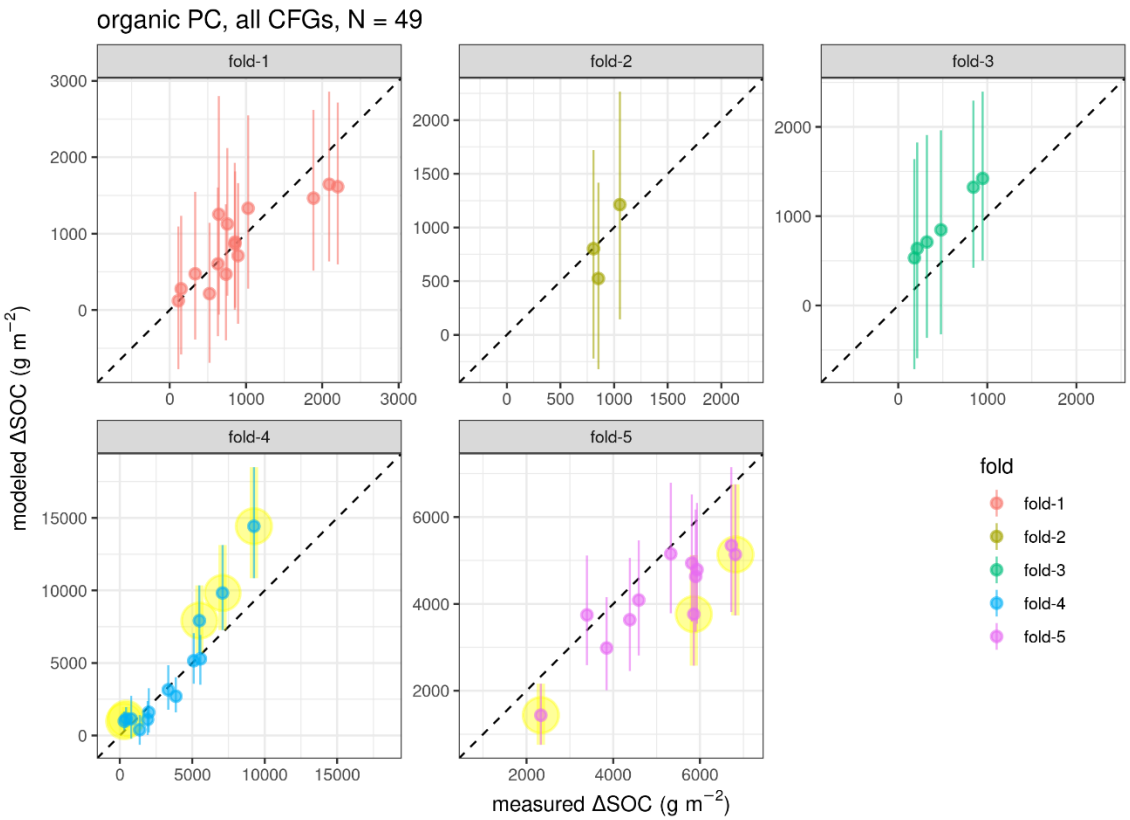

Figure 50: Model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from all CFGs. Error bars shows 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value.

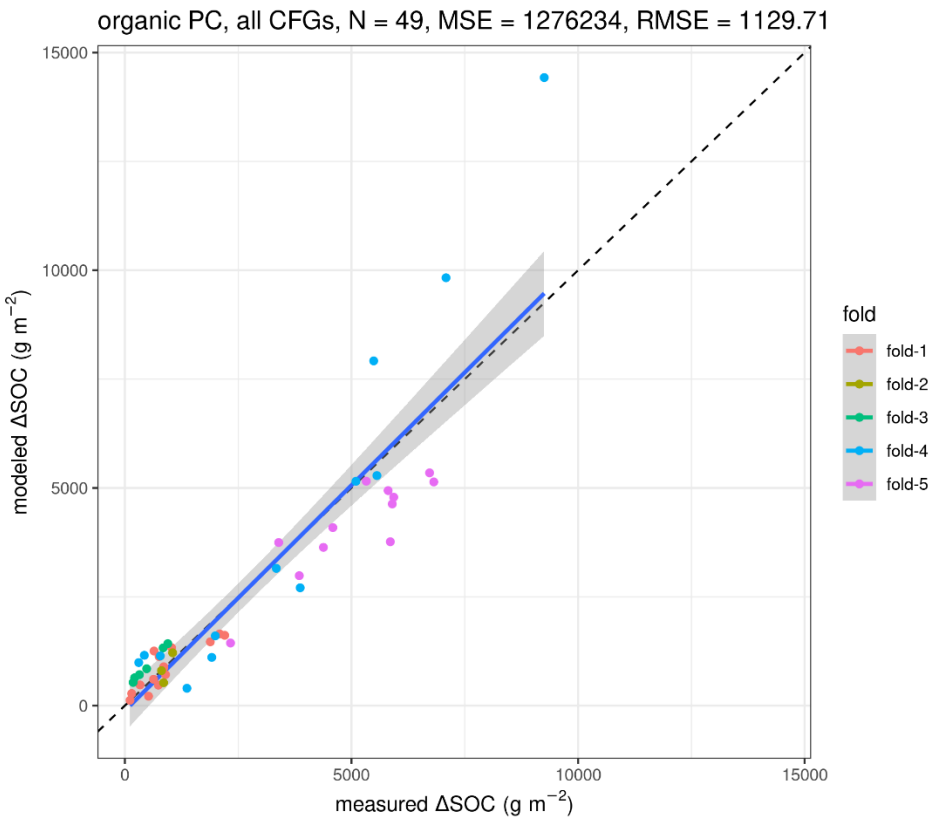

Figure 51: Scatterplot of model predictions versus measurements of SOC change in response to changed organic amendment practices involving crops from the soy-type CFG. Shaded area around solid line shows 95% CI around linear least-squares fit (does not consider model or measurement error). Dashed line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

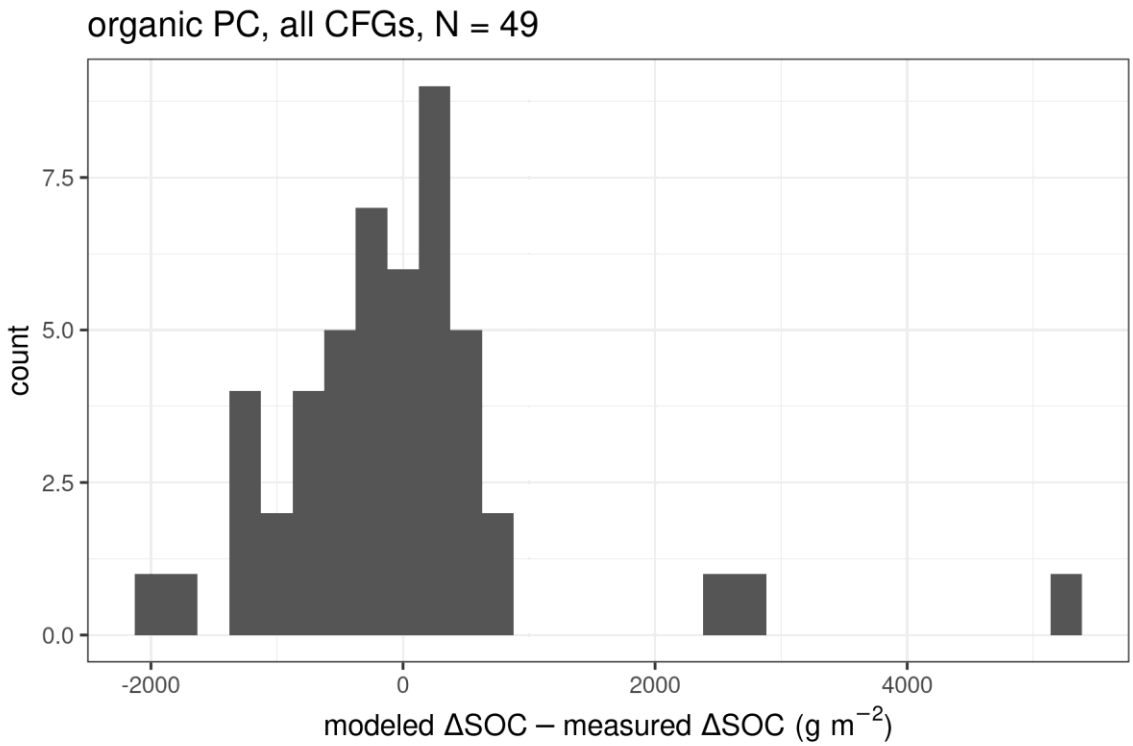

Figure 52: Histogram of model residuals (predicted - observed) for change in SOC in the studies used for model validation evaluating changed organic amendment practices involving any CFG.

Table 42: Number of observed ORG datapoints from any crop falling inside and outside of modeled 90% prediction intervals for each fold of the calibration/validation process

fold	n	n in	n out	% coverage
1	15	15	0	100
2	3	3	0	100
3	6	6	0	100
4	13	8	5	62
5	12	9	5	75
all	**49**	**41**	**8**	**84**

Mean squared error: 1276234 (g C m^{-2})^2; RMSE = 1130 g C m^{-2}.

Do 90% of prediction intervals cover observed data? **No, only 41 of 49 observations (84%) fall inside the model's 90% prediction intervals**. However when excluding datapoints where measured dSOC > 5000 g C m^{-2} as discussed in section "Restrictions on application of model", 34/37 observations (92%) fall inside the prediction intervals and additionally RMSE drops to 506 g C m^{-2}**.**

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. The resulting posterior distributions from this final step (Appendix E: Documentation of calibrated parameter sets) are very similar to the distributions obtained during cross-validation (Figure 55) 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.

Model bias across all PCs and CFGs:

During cross-validation: 27.2±16.9 g C m^{-2}

With final parameter set: 39.9±14.6 g C m^{-2}

MSE and RMSE across all PCs and CFGs:

During cross-validation: MSE 354953 (g C m^{-2})^2, RMSE 596 g C m^{-2}

With final parameter set: MSE 316529 (g C m^{-2})^2, RMSE 563 g C m^{-2}

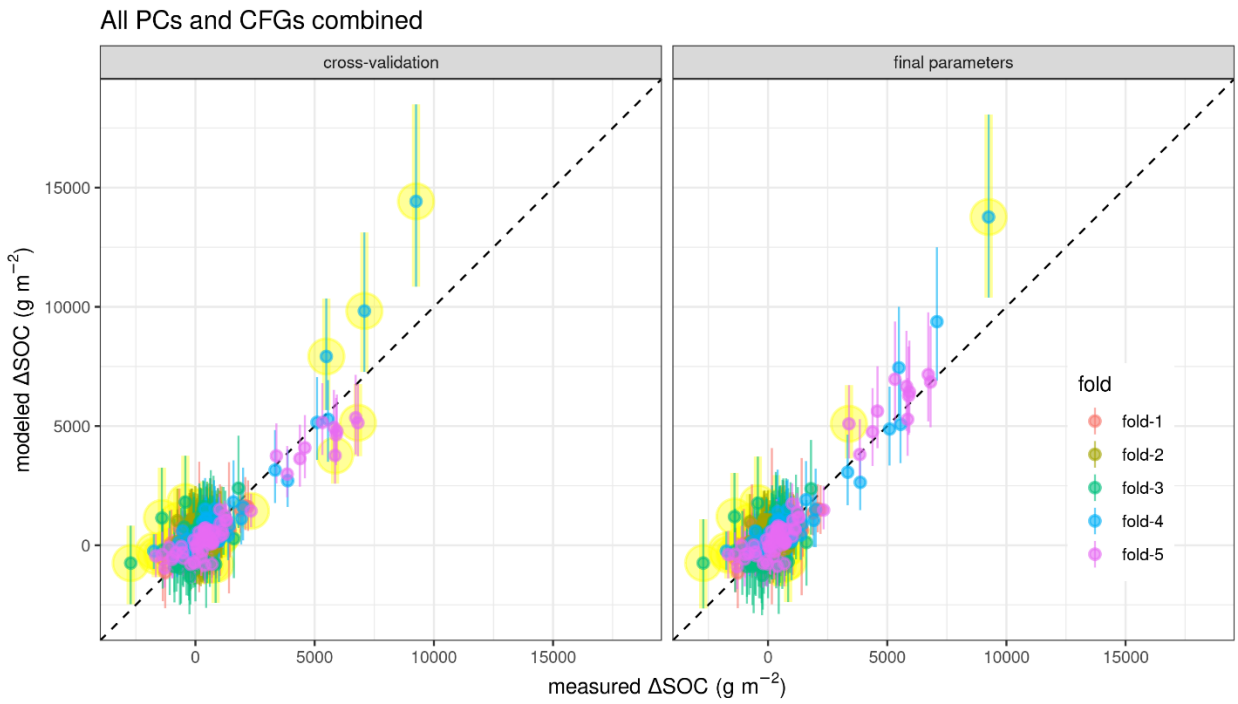

Figure 53: Model predictions versus measurements of SOC change, shown for out-of-sample sites during cross-validation (left) and when fitting the same data using the final parameter set (right). Error bars shows 90% prediction intervals; intervals highlighted in yellow do not overlap the observed value. Note that fold identity was not used during fitting with the final parameter set, so colors in the right panel are shown only for comparison between panels.

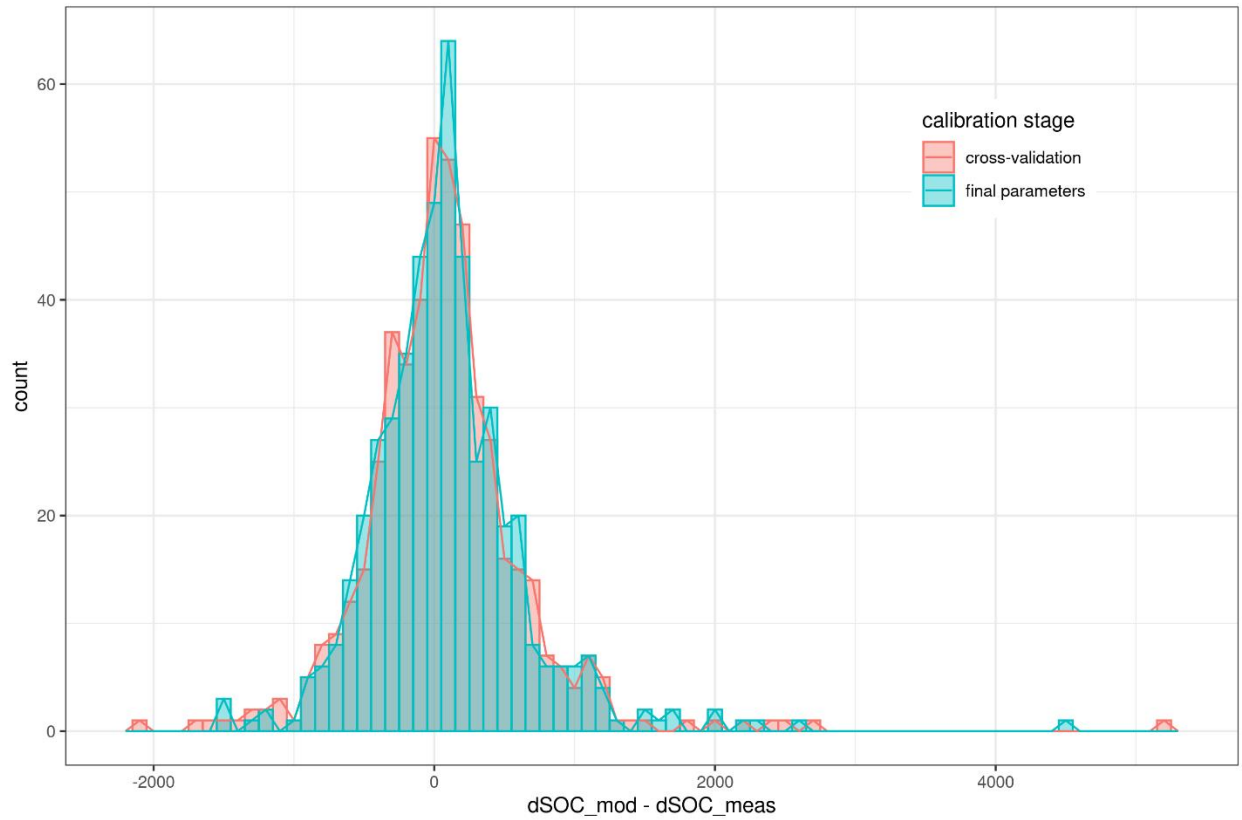

Figure 54: Histograms of model residuals obtained during cross-validation (red) and when fitting the validation data using the final parameter set (blue).

Table 43: Comparison of model bias in each PC x CFG category during cross-validation and with final parameter set. Numbers in parentheses are standard errors.

PC	CFG	PMU	n obs	n sites	Bias k-fold	Bias final params	Final bias smaller?
organic	wheat	376.19	44	8	260.66 (107.96)	367.77 (69.22)	FALSE
till	corn	635.34	82	10	76.87 (21.97)	32.64 (15.27)	TRUE
N	corn	764.67	91	13	50.56 (62.59)	56.15 (61.05)	FALSE
N	wheat	212.83	102	12	33.49 (43.61)	95.81 (38.68)	FALSE
crop	soy	335.04	106	13	-6.75 (37.22)	-11.38 (37.49)	FALSE
till	soy	400.48	27	5	-12.46 (44.87)	-54.71 (30.38)	FALSE
crop	corn	335.04	110	12	-35.92 (35.47)	-37.37 (36.05)	FALSE
crop	wheat	289.8	120	16	-75.71 (35.67)	-74.07 (36.93)	TRUE
till	wheat	282.2	56	9	-82.67 (42.9)	-75.63 (40.86)	TRUE

Table 44: Comparison of MSE and RMSE in each PC x CFG category during cross-validation and with final parameter set.

PC	CFG	MSE k-fold	RMSE k-fold	MSE final params	RMSE final params	Final MSE smaller?
organic	wheat	1373695	1172	1009394	1004.7	TRUE
N	corn	445669	667.6	433407	658.3	TRUE
till	corn	357681	598.1	324831	569.9	TRUE

N	wheat	221759	470.9	241296	491.2	FALSE
crop	wheat	208592	456.7	214057	462.7	FALSE
crop	soy	162560	403.2	168687	410.7	FALSE
till	soy	154215	392.7	128177	358	TRUE
crop	corn	143142	378.3	149917	387.2	FALSE
till	wheat	132642	364.2	130616	361.4	TRUE

Table 45: Comparison of posterior 90% prediction interval coverage in each PC x CFG category during cross-validation and with final parameter set.

			k-fold			**Final params**		
PC	**CFG**	**n**	**n in**	**n out**	**Percent coverage**	**n in**	**n out**	**Percent coverage**
till	wheat	56	55	1	98	55	1	98
till	soy	27	26	1	96	26	1	96
crop	corn	110	105	5	95	106	4	96
crop	soy	106	101	5	95	102	4	96
crop	wheat	120	114	6	95	115	5	96
till	corn	82	76	6	93	78	4	95
N	wheat	102	96	6	94	96	6	94
organic	wheat	44	36	8	82	41	3	93
N	corn	91	82	9	90	82	9	90
All categories combined		**495**	**465**	**30**	**94**	**472**	**23**	**95**

Model validation outputs for use in SEP uncertainty calculations

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 CAR 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.

Original SEP

Follows Model Requirements, Section 3.5 (p18)

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. We report an RMSE of 596 g C m^{-2} in "Model prediction error across all categories," giving a value of 354953 (g C m^{-2})^2 for this MSE.

Revised SEP

Follows Model Requirements, Section 3.5 (p18), as amended in Appendix H

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:

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 will proceed as described for "Original SEP", above.

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 hyperparameters capturing

model structural uncertainty (both summarized in Table 47, Appendix E). 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 the "Model prediction error" section of this report. When running the model for crediting, the ensemble of simulations for a given datapoint will consist of 71 DayCent-CR simulations, one using each unique combination of parameter states in the stored posterior, which are then duplicated according to their frequency in the posterior and then combined with independent draws from the uncertainty hyperparameters to obtain 220 Monte Carlo predictions. These are then summarized to quantify model prediction uncertainty for that datapoint. Because DayCent-CR is a deterministic model, this produces identical results to what would have been obtained by running the model for every one of the 220 posterior samples.

Restrictions on application of model

As identified in the uncertainty evaluation for ORG x Wheat x SOC, DayCent-CR 1.0 showed evidence of underestimating the uncertainty of its predictions for extremely large changes in SOC (total dSOC > 5000 g m^{-2} over any time interval), but has adequate uncertainty coverage and sufficient lack of bias for all SOC changes up to 5000. ORG X Wheat x SOC was the only PC x CFG combination for which changes of this magnitude were observed in the available validation data, so we cannot determine whether this limitation is specific to organic amendments or whether it applies to large SOC changes more generally. The model otherwise demonstrated consistent performance at and below 5000 g m^{-2} across all categories.

5000 g m^{-2} is much larger than the dSOC we expect to observe in the project, and is probably larger than is biologically achievable within a single project remeasurement interval of 5 years. However, changes this large would be highly influential on the project estimates if they were to occur, and understating their uncertainty would be anticonservative. Therefore, 5000 g m^{-2} sets the limit for model predictions across the entire project domain i.e. all crops and practices, for any consecutive time interval between soil measurement true-ups. Any model result that predicts a **change >= 5000 g C m^{-2} would be restricted from use in crediting.**

Citations

Cited in report text

Barré, P., Eglin, T., Christensen, B. T., Ciais, P., Houot, S., Kätterer, T., Van Oort, F., Peylin, P., Poulton, P. R., Romanenkov, V., & Chenu, C. (2010). Quantifying and isolating stable soil organic carbon using long-term bare fallow experiments. *Biogeosciences*, 7(11), 3839–3850. https://doi.org/10.5194/bg-7-3839-2010

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*. https://doi.org/10.1006/jmps.1999.1279

Del Grosso, S. J., Parton, W. J., Mosier, A. R., Walsh, M. K., Ojima, D. S., & Thornton, P. E. (2006). DAYCENT national-scale simulations of nitrous oxide emissions from cropped soils in the United States. *Journal of Environmental Quality*, 35(4), 1451-1460

Del Grosso, S. J., Parton, W. J., Adler, P. R., Davis, S. C., Keough, C., & Marx, E. (2012). DayCent model simulations for estimating soil carbon dynamics and greenhouse gas fluxes from agricultural production systems. *Managing agricultural greenhouse gases: coordinated agricultural research through GRACEnet to address our changing climate*, 241-250

Eve, M., D. Pape, M. Flugge, R. Steele, D. Man, M. Riley-Gilbert and S. Biggar (Eds). 2014. Quantifying Greenhouse Gas Fluxes in Agriculture and Forestry: Methods for Entity-Scale Inventory. USDA Office of the Chief Economist, Climate Change Program Office. *USDA Technical Bulletin* 1939. Viewed at http://www.usda.gov/oce/climate_change/estimation.htm

Gurung RB, Ogle SM, Breidt FJ, et al (2020) Bayesian calibration of the DayCent ecosystem model to simulate soil organic carbon dynamics and reduce model uncertainty. *Geoderma* 376:114529. https://doi.org/https://doi.org/10.1016/j.geoderma.2020.114529

Hao, X., Chang C., Travis, G.R., and Zhang, F. 2003. Soil carbon and nitrogen response to 25 annual cattle manure applications. J. *Plant Nutr. Soil Sci*. 166, 239-245

Jenkinson, DS. 1991. The Rothamstead long-term experiments: Are they still of use? *Agron J*. 83, 2-10. https://doi.org/10.2134/agronj1991.00021962008300010008x

Joseph Guillaume and Felix Andrews (2012). dream: DiffeRential Evolution Adaptive Metropolis. R package version 0.4-2. URL http://CRAN.R-project.org/package=dream

Kennedy MC, O'Hagan A, (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B-Statistical Methodol*. 63, 425–450. https://doi.org/https://doi.org/10.1111/1467-9868.00294

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

PRISM Climate Group: PRISM Climate Data, Oregon State University. http://prism.oregonstate.edu

R Core Team, 2018. R: A language and environment for statistical computing.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., & Dormann, C. F. (2017). Cross-Validation Strategies for Data with Temporal, Spatial, Hierarchical, or Phylogenetic Structure. *Ecography* 40, 913–929. https://doi.org/10.1111/ecog.02881Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Math Model. Comput. Exp.*, 1, 407–414.

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. United States Department of Agriculture, Natural Resources Conservation Service. Available online at https://gdg.sc.egov.usda.gov/.

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

USEPA (2020). Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2018

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling and Software*. https://doi.org/10.1016/j.envsoft.2015.08.013

Vrugt, J. A., & Ter Braak, C. J. F. (2011). DREAM(D): An adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-15-3701-2011

Vrugt, Jasper A., Ter Braak, C. J. F., Diks, C. G. H., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*. https://doi.org/10.1515/IJNSNS.2009.10.3.273

Vrugt, J. A., & Ter Braak, C. J. F. (2011). DREAM(D): An adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-15-3701-2011

Zhang, Y., A. Suyker, and K. Paustian. 2018. Improved crop canopy and water balance dynamics for agroecosystem modeling using DayCent. *Agronomy Journal* 110:511-524

Zhang Y, Arabi M, Paustian K, (2020). Analysis of parameter uncertainty in model simulations of irrigated and rainfed agroecosystems. *Environmental Modelling and Software*

Sources for validation data

Bista et al. 2016: Prakriti Bista,Stephen Machado, Rajan Ghimire, Stephen J. Del Grosso, and Melissa Reyes-Fox. 2016. Simulating Soil Organic Carbon in a Wheat–Fallow System Using the Daycent Model *Agron. J.* 108, 2554-2565. doi:10.2134/agronj2016.04.0202

Blevins et al. 1983: R.L. BLEVINS, G.W. THOMAS, M.S. SMITH, W.W. FRYE and P.L. CORNELIUS. 1983. CHANGES IN SOIL PROPERTIES AFTER 10 YEARS CONTINUOUS NON-TILLED AND CONVENTIONALLY TILLED CORN. *Soil & Tillage Research* 3, 135-146

Campbell and Zentner 1997: Campbell C.A., and Zentner R.P. 1997. Crop production and soil organic matter in long-term crop rotations in the semi-arid Northern Great Plains of Canada. Chapter 23 in: *Soil Organic Matter in Temperate Agroecosystems*. E.A. Paul. K. Paustian, E.T. Elliott, and C.V. Cole (eds.). CRC Press, Inc.

Campbell et al. 2007: C. A. Campbell, A. J. VandenBygaart, R. P. Zentner, B. G. McConkey, W. Smith, R. Lemke, B. Grant, and P. G. Jefferson. 2007. Quantifying carbon sequestration in a minimum tillage crop rotation study in semiarid southwestern Saskatchewan. *Can. J. Soil Sci.* 87, 235-250

Christenson 1997: Christenson, D. R. 1997. Soil organic matter in sugar beet and dry bean cropping systems in Michigan. Chapter 10 In: *Soil Organic Matter in Temperate Agroecosystems*. E.A. Paul, K. Paustian, E.T. Elliott, and V. Cole (eds.). CRC Press, Inc.

Clapp et al. 2000: C.E. Clapp, R.R. Allmaras, M.F. Layese, D.R. Linden and R.H. Dowdy 2000. Soil organic carbon and 13C abundance as related to tillage, crop residue, and nitrogen fertilization under continuous corn management in Minnesota. *Soil & Tillage Research* 55, 127-142. doi:10.1016/j.still.2005.07.015

Clark et al. 1998: Clark, M.S., Horwath, W.R., Shennan C., and Scow, K.M. 1998. Changes in soil chemical properties resulting from organic and low-input farming practices. *Agron. J.* 90, 662-671

Collins et al. 1999: Collins, H.P., Blevins, R.L., Bundy, L.G., Christenson, D.R., Dick, W.A., Huggins, D.R., and Paul, E.A. 1999. Soil carbon dynamics in corn-based agroecosystems: Results from Carbon-13 Natural Abundance. *Soil Sci. Soc. Am. J.* 63, 584-591

Collins et al. 2000: Collins, H.P., E.T. Elliott, K. Paustian, L.G. Bundy, W.A. Dick, D.R. Huggins, A.J.M. Smucker and E.A. Paul. 2000. Soil carbon pools and fluxes in long-term corn belt agroecosystems. *Soil Biol. Biochem.* 32, 157-168. doi:10.1016/S0038-0717(99)00136-4

Denef et al. 2008: Denef, K., Stewart, C.E., Brenner J., and Paustian, K. 2008. Does long-term center-pivot irrigation increase soil carbon stocks in semi-arid agro-ecosystems? *Geoderma* 145, 121-129. doi:10.1016/j.geoderma.2008.03.002

Dick et al. 1997: Dick, W. A., Edwards, W.M., and McCoy, E.L. 1997. Continuous application of no-tillage to Ohio soils: Changes in crop yields and organic matter-related soil properties. Chapter 12 in *Soil Organic Matter in Temperate Agroecosystems*. E.A. Paul. K. Paustian, E.T. Elliott, and C.V. Cole (eds.). CRC Press, Inc, 171-182

Dick et al. 2013: Dick, W.A., Triplett, G.B., and Van Doren, D.M. 2013. Triplett-Van Doren Long-term Tillage and Crop Rotation Data. Wooster, Ohio: Ohio Agricultural Research and Development Center. Ohio State University Knowledge Bank. *Online at:* https://kb.osu.edu/handle/1811/55716

Dolan et al. 2006: M.S. Dolan, C.E. Clapp, R.R. Allmaras, J.M Baker, J.A.E. Molina 2006. Soil organic carbon and nitrogen in a Minnesota soil as related to tillage, residue and nitrogen management. *Soil & Tillage Research* 89, 221-231

Elliott et al. 1994: Elliott, E.T. I.C. Burke, C.A. Monz, S.D. Frey, K. H. Paustian, H.P. Collins, E.A. Paul, C.V. Cole, R.L. Blevins, W.W. Frye, D.J. Lyon, , A.D. Halvorson, D.R.

Huggins, R.F. Turco, and M.V. Hickman. 1994. Terrestrial carbon pools in grasslands and agricultural soils: Preliminary data from the Corn Belt and Great Plains Regions. Chapter 12 in: *SSSA Special Publication 35: Defining Soil Quality for a Sustainable Environment* 35, 179-192. doi:10.2136/sssaspecpub35.c12

Ghimire et al. 2015: Rajan Ghimire,Stephen Machado, and Karl Rhinhart 2015. Long-Term Crop Residue and Nitrogen Management Effects on Soil Profile Carbon and Nitrogen in Wheat–Fallow Systems. *Agron. J.* 107, 2230-2240. doi:10.2134/agronj14.0601

Ghimire et al. 2017: Rajan Ghimire,Stephen Machado, and Prakriti Bista. 2017. Soil pH, Soil Organic Matter, and Crop Yields in Winter Wheat–Summer Fallow Systems *Agron. J.* 109 706-717. doi:10.2134/agronj2016.08.0462

Halvorsen et al. 2002: Ardell D. Halvorson, Brian J. Wienhold, and Alfred L. Black. 2002. Tillage, Nitrogen, and Cropping System Effects on Soil Carbon Sequestration. *Soil Sci. Soc. Am. J.* 66, 906-912

Halvorson and Jantalia 2011: Halvorson, A.D, and Jantalia, C.P. 2011. Nitrogen fertilization effects on irrigated no-till corn production and soil carbon and nitrogen. *Agron. J.* 103, 1423-1431 doi:10.2134/agronj2011.0102

Halvorson et al. 2016: Halvorson, A.D., Stewart, C.E., and Del Grosso, S.J. 2016. Manure and inorganic nitrogen affect irrigated corn yields and soil properties. *Agron. J.* 108, 519-531. doi:10.2134/agronj2015.0402

Hao et al. 2003: Hao, X., Chang C., Travis, G.R., and Zhang, F. 2003. Soil carbon and nitrogen response to 25 annual cattle manure applications. *J. Plant Nutr. Soil Sci.* 166, 239-245

Ismail et al. 1994: Ismail, I., Blevins, R.L., and Frye, W.W. 1994. Long-term no-tillage effects on soil properties and continuous corn yields. *Soil Sci. Soc. Am. J.* 58, 193-198

Janzen et al. 1997: Janzen, H. H., Johnston, A.M., Carefoot, J.M., and Lindwall, C.W. 1997. Soil organic matter dynamics in long-term experiments in southern Alberta. Chapter 21 in: *Soil Organic Matter in Temperate Agroecosystems*. E.A. Paul. K. Paustian, E.T. Elliott, and C.V. Cole (eds.). CRC Press, Inc.

Jarecki and Lal 2005: Jarecki, M.K., and Lal, R. 2005. SOIL ORGANIC CARBON SEQUESTRATION RATES IN TWO LONG-TERM NO-TILL EXPERIMENTS IN OHIO. *Soil Sci.* 170, 280-291. doi:10.1097/01.ss.0000162286.95137.70

Jin and Varvel. 2018a: Jin, V., and Varvel, G. 2018. All; 2001-2016. ver. GPSR_NATRES. Fort Collins, CO: USDA-ARS Natural Resources Database. Project: NEMEIRR. URL: https://gpsr.ars.usda.gov/NEMEIRR. File Downloaded: 2018-Sep-27;[10:54:29]. PID: 50929d1d-7921-441c-b814-f0f1dda9c84e

Jin and Varvel. 2018b: Jin, V., and Varvel, G. 2018. All; 1998-2016. ver. GPSR_NATRES. Fort Collins, CO: USDA-ARS Natural Resources Database. Project: NEMERREM. URL: https://gpsr.ars.usda.gov/NEMERREM. File Downloaded: 2018-Sep-27;[10:54:33]. PID: d235196c-2341-4164-a77d-d7d329b3fd15

Jin et al. 2015: Jin, V.L., Schmer, M.R., Wienhold, B.J., Stewart, C.E, Varvel, G.E., Sindelar, A.J., Follett, R.F., Mitchell, R.B., and Vogel, K.P. 2015. Twelve Years of Stover Removal Increases Soil Erosion Potential without Impacting Yield Soil. *Sci. Soc. Am. J.* 79, 1169-1178. doi:10.2136/sssaj2015.02.0053

Khan et al. 2007: Khan, S.A., Mulvaney, R.L., Ellsworth, R., and Boast, C.W. 2007. The Myth of Nitrogen Fertilization for Soil Carbon Sequestration. *J. Environ. Quality* 36, 1821-1832

Kong et al. 2005: Angela Y. Y. Kong, Johan Six, Dennis C. Bryant, R. Ford Denison, and Chris van Kessel. 2005. The Relationship between Carbon Input, Aggregation, and Soil Organic Carbon Stabilization in Sustainable Cropping Systems *Soil Sci. Soc. Am. J.* 69, 1078-1085. doi:10.2136/sssaj2004.0215.

Mestelan 2008: Mestelan, S.A. 2008. IMPACT OF LONG-TERM NO TILL AND PLOW TILL ON SOIL PROPERTIES AND SOIL NUTRIENT CYCLING. Dissertation, Ohio State University

Mikha et al. 2013: Mikha, M.M., Vigil, M.F., and Benjamin, J.G. 2013. Long-term tillage impacts on soil aggregation and carbon dynamics under wheat-fallow in the Central Great Plains. *Soil Sci. Soc. Am. J.* 77, 594-605. doi:10.2136/sssaj2012.0125

Mitchell et al. 2015: Jeffrey P. Mitchell, Anil Shrestha, William R. Horwath, Randal J. Southard, Nicholaus Madden, Jessica Veenstra, and Daniel S. Munk 2015. Tillage and Cover Cropping Affect Crop Yields and Soil Carbon in the San Joaquin Valley, California. *Agron. J.* 107, 588-596. doi:10.2134/agronj14.0415

Mitchell et al. 2017: Jeffrey P. Mitchell, Anil Shresth, Konrad Mathesius, Kate M. Scow, Randal J. Southard, Richard L. Haney, Radomir Schmidt, Daniel S. Munk, William R. Horwath 2017. Cover cropping and no-tillage improve soil health in an arid irrigated cropping system in California's San Joaquin Valley, USA. *Soil & Tillage Research* 165, 325-335. doi:10.1016/j.still.2016.09.001

Monreal and Janzen 1993: Monreal, C. M. and Janzen, H. H. 1993. Soil organic-carbon dynamics after 8O years of cropping a Dark Brown Chernozem. *Can. J. Soil Sci.* 73, 133-136

Olson et al, 2010: Olson, K.R., Ebelhar, S.A., and Lang, J.M. 2010. Cover crop effects on crop yields and soil organic carbon content. *Soil Sci.* 175, 89-98. doi:10.1097/SS.0b013e3181cf795

Paul et al. 2004: Paul, E.A., H.P. Collins, K. Paustian, E.T. Elliott, S. Frey, N. Juma, H.H. Janzen, C.A. Campbell, R.P. Zentner, G.P. Lafond and A.P. Moulin. 2004. Management effects on the dynamics and storage rates of organic matter in long-term crop rotations. *Can. J Soil Sci.* 84, 49-61. doi:10.4141/S03-022

Pimental et al. 2005: DAVID PIMENTEL, PAUL HEPPERLY, JAMES HANSON, DAVID DOUDS, AND RITA SEIDEL. 2005. Environmental, Energetic, and Economic Comparisons of Organic and Conventional Farming Systems. *BioScience* 55, 573-582

Powlson et al. 2012: Powlson, D. S., Bhogal, A., Chambers, B. J., Coleman, K., Macdonald, A. J., Goulding, K. W. T. and Whitmore, A. P. 2012. The potential to increase soil carbon

stocks through reduced tillage or organic material additions in England and Wales: a case study. *Agriculture, Ecosystems & Environment* 146, 23-33. doi:10.1016/j.agee.2011.10.004

Rasmussen and Smiley 1997: Rasmussen, P. E., and R. W. Smiley 1997. Soil carbon and nitrogen change in long-term agricultural experiments at Pendleton, Oregon. Chapter 26 in: *Soil Organic Matter in Temperate Agroecosystems*. E.A. Paul. K. Paustian, E.T. Elliott, and C.V. Cole (eds.). CRC Press, Inc.

Rothamsted Research (2014): 2014. Broadbalk soil organic carbon content. Electronic Rothamsted Archive. https://doi.org/10.23637/KeyRefOABKsoc. doi:10.23637/KeyRefOABKsoc. *Note:* this is the electronic repository for data described in Powlson et al. 2012.

Schmer et al. 2014: Schmer, M.R., Jin, V.L., Weinhold, B.J., Varvel, G.E., and Follett, R.F. 2014. Tillage and Residue Management Effects on Soil Carbon and Nitrogen Under Irrigated Continuous Corn. *Soil Sci. Soc. Am. J.* 78, 1987-1996. doi:10.2136/sssaj2014.04.0166

Stewart et al. 2015: Stewart, C.E., Follett, R.F., Preussner, E.G., Varvel, G.E., Vogel, K.P.. And Mitchell, R.B. 2015. Nitrogen and harvest effects on soil properties under rainfed switchgrass and no-till corn over 9 years: implications for soil quality. *GCB Bioenergy* 7, 288-301. doi:10.1111/gcbb.12142

Varvel 2006: Varvel, G.E. 2006. Soil Organic Carbon Changes in Diversifified Rotations of the Western Corn Belt. *Soil Sci. Soc. Am. J.* 70, 426-433. doi:10.2136/sssaj2005.0100

Veenstra et al. 2006: Jessica J. Veenstra, William R. Horwath, Jeffrey P. Mitchell and Daniel S. Munk 2006. Conservation tillage and cover cropping influence soil properties in San Joaquin Valley cotton-tomato crop. *CALIFORNIA AGRICULTURE* 60, 146-153

Vitosh et al. 1973: Vitosh, M.L, Davis, J.F., and Knezek. B.D. 1973. Long-term effects of manure, fertilizer, and plow depth on chemical p roperties of soils and nutrient movement in a monoculture corn system. *J. Environ. Quality* 2, 296-299