



OPEN A multi model ensemble reveals net climate benefits from regenerative practices in US Midwest croplands

Bruno Basso^{1,2,3}✉, Tommaso Tadiello¹✉, Neville Millar¹, G. Philip Robertson^{2,3,4}, Keith Paustian^{5,6}, Fidel S. Maureira¹, Susana Albarenque¹, Brian Baer¹, Lydia Price¹, Prateek Sharma¹, Chris Villalobos¹, Ames Fowler^{1,5}, Mathieu Delandmeter⁷, Marco Acutis⁸, Sotirios Archontoulis⁹, Kristofer R. Covey¹⁰, Luca Doro¹¹, Benjamin Dumont⁷, Peter R. Grace¹², Gerrit Hoogenboom¹³, James W. Jones¹³, Alessia Perego⁸, Alexander Ruane¹⁴, Claudio O. Stöckle¹⁵ & Yao Zhang⁶

Process-based cropping systems models (CSMs) are key components of measurement, monitoring, reporting, and verification frameworks of carbon markets, but model-specific differences limit their applicability across diverse pedo-climatic conditions and agronomic practices. Multi-model ensemble (MME) provides an opportunity to better estimate changes in soil organic carbon (SOC) and nitrous oxide (N₂O) emissions from agronomic practices at scale. We used an MME across 46 million hectares of US Midwest cropland at a resolution of 4-km² to assess the aggregate ability of different regenerative practices to sequester SOC and N₂O emissions compared to their counterfactual dynamic baselines. MME was validated against long-term trials and compared to its constituent CSMs, showing greater accuracy and lower uncertainty. The results show that adopting no-till combined with cover crops increased SOC stocks by 0.36 ± 0.12 Mg ha⁻¹ yr⁻¹, corresponding to a net regional SOC gain of 16.4 Tg C yr⁻¹ compared to business-as-usual baselines. These benefits are halved when each management is practiced individually, and the SOC gains are only fully realized with low initial carbon stock. By including N₂O emissions, we can assess the overall climate mitigation potential, specifically, the extent to which carbon sequestration can offset direct N₂O emissions. The magnitude of this potential varies depending on management practices and geographic location with net climate benefits on average ranging from 0 to 3 Mg CO₂-eq ha⁻¹ yr⁻¹. High-resolution MME results allow for robust estimates of climate mitigation, reducing barriers to carbon market participation and supporting regenerative agriculture initiatives at scale.

Keywords Regenerative agriculture, soil carbon, Multi-model ensemble, Carbon credits, N₂O emissions, Dynamic baselines

Soil organic carbon (SOC) storage and nitrous oxide reduction emissions are key strategies for agricultural climate change mitigation but implementing regenerative practices on croplands at scale is challenging due to inherent field variability, numerous quantification approaches, practical and financial barriers, and vested

¹Department of Earth and Environmental Sciences, Michigan State University, East Lansing 48824, USA. ²W.K. Kellogg Biological Station, Michigan State University, Hickory Corners 49060, USA. ³DOE Great Lakes Bioenergy Research Center, Michigan State University, East Lansing 48824, USA. ⁴Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI 48824, USA. ⁵Department of Soil and Crop Sciences, Colorado State University, Fort Collins 80521, USA. ⁶Natural Resource Ecology Laboratory, Colorado State University, Fort Collins 80523, USA. ⁷Plant Sciences Axis - Crop Science Lab, Gembloux Agro-Bio Tech, TERRA Teaching and Research Centre, University of Liege, 5030 Gembloux, Belgium. ⁸Department of Agricultural and Environmental Sciences - Production, Landscape, Agroenergy, University of Milan, 20133 Milan, Italy. ⁹Department of Agronomy, Iowa State University, Ames 50011, USA. ¹⁰Environmental Studies and Sciences Program, Skidmore College, New York 12866, USA. ¹¹Blackland Research Center, Texas A&M AgriLife Research, Temple 76502, USA. ¹²Centre for Agriculture and the Bioeconomy, Queensland University of Technology, Brisbane 4000, Australia. ¹³Department of Agricultural and Biological Engineering, University of Florida, Gainesville 32611, USA. ¹⁴Goddard Institute for Space Studies, National Aeronautics and Space Administration, New York 10025, USA. ¹⁵Biological Systems Engineering, Washington State University, Pullman 99164, USA. ✉email: basso@msu.edu; tadiell1@msu.edu

interests limit data accuracy, reproducibility, and scalability. Uncoordinated, inconsistent carbon market initiatives reflect these challenges¹, failing to attract significant producer participation and hindering confidence in environmental and financial outcomes. In short, the identification and quantification of climate benefits at a meaningful scale has been stymied by broad ranges and large uncertainties².

Individual process-based ecosystem models developed for agriculture, more precisely called cropping system models (CSMs)³, have been used to investigate the effects of management and climate change on SOC and GHG emissions^{4–6} at regional to global scales (e.g., Refs.^{7,8}). CSMs incorporate varying theoretical assumptions, often using limited empirical observations for model calibration and validation^{9,10}. Owing to the different crops, management practices, and geographic regions for which they have been developed, different CSMs can provide divergent predictions for SOC change, particularly for those models applied to environments and scenarios, present and future, for which they have not been calibrated or developed¹¹.

Research over the last several decades has shown that there is no silver-bullet crop and biogeochemical model¹², although some have been more widely tested than others, and some specific combinations of models can perform better¹³. So, even if CSMs provide an attractive alternative to large-scale SOC and N₂O measurement and monitoring programs, no one model can satisfactorily quantify its change across all combinations of soils, climates, crops, and agronomic practices at varying scales, despite the well-validated performance of individual models in specific contexts. For all these reasons, models practitioners and certificate providers still face the challenge of selecting the right model, appropriate to the specific conditions in which they are working¹⁴.

Climate modelers have addressed this problem with multi-model ensemble (MME) approaches (e.g., Ref.¹⁵), which have also worked well for predicting climate-induced crop productivity change^{4,16}. MMEs capture differences in model structure as well as broader sets of environmental conditions and management practices than those for which individual models have been developed and tested. Consequently, they provide a substantial advantage for evaluating SOC change and GHG emissions in response to changes in management practices and climate, which could be particularly attractive for addressing measurement, monitoring, reporting, and verification (MMRV) challenges of current carbon credit frameworks^{1,17}. At small scales, previous MMEs have been used to predict SOC change more accurately than individual models in croplands (e.g. Refs.^{4,13,18}), grasslands (e.g., Ref.¹⁹), and bare soil (e.g., Ref.²⁰), but have yet to be applied to large, landscape-scale questions despite the scientific and policy importance of mitigation interventions in agriculture²¹.

Soil sampling programs typically cannot meet MMRV challenges because of the cost of sampling and analysis, the need to adequately capture spatial variability, and the decadal or longer time frame required to detect credible SOC change^{22–25}. Consequently, on-farm observations, unless conducted as part of funded scientific trials with scientific expert supervision, typically cannot reliably capture the magnitude of carbon dynamics, particularly the rate of SOC change²⁶, leading to large uncertainties in SOC datasets²⁷. These are major problems for calculating cropland SOC budgets and credits²⁸, with important implications for crediting SOC gains¹⁷.

Ultimately, the global challenge of climate mitigation demands a new paradigm for evaluating SOC change. It should not be assessed in isolation, but in terms of its ability to offset major agricultural emissions—most notably, N₂O²⁹. A single practice that increases SOC does not necessarily represent the most effective management strategy if associated N₂O emissions are not also considered^{30,31}.

This study investigates the large-scale effects of regenerative management practices on SOC and N₂O dynamics using a multi-model ensemble (MME) approach. The ensemble includes widely used CSMs—APSIM, ARMOSA, CropSyst, DayCent, DSSAT, EPIC, SALUS, and STICS—and is applied across 40,000 unique locations across 934 counties in 12 US Midwest States. Estimates of SOC stock change will be used to (1) quantify the potential SOC accrual from the adoption of regenerative practices relative to a dynamic business-as-usual baselines, and (2) assess the climate mitigation potential of these practices by evaluating their capacity to offset N₂O emissions at scale.

We hypothesize that SOC gains will continue to offset N₂O emissions—with the extent of offset depending on site-specific interactions between pedo-climatic conditions and the practices adopted—ultimately resulting in a net positive climate mitigation effect from implementing regenerative agricultural practices.

Material and methods

Process-based cropping systems models (CSMs)

Eight different process-based CSMs were used to develop the MME (Table S1): Apsim³², Armosa³³, Cropsyst³⁴, Daycent³⁵, Dssat³⁶, Epic³⁷, Salus³⁸ and Stics³⁹. Models were selected primarily based on their ability to reproduce interactions between soil, plant, management and the atmosphere to simulate crop yields and ecosystem processes, including SOC change and N₂O emissions. Each model uniquely simulates and mathematically represents these processes (Table S2) and has been shown to appropriately represent biogeochemical cycles under well-validated field conditions representative of cropping systems and soils of the US Midwest cropland. All models operate with daily weather inputs and can simulate carbon and nitrogen cycling, SOC change, nitrous oxide emissions, water balance, and plant growth in response to different crop management practices including fertilization, tillage, and irrigation. A brief description of each model is provided in the supplemental material (Sect. 1).

Scaling the multi-model ensemble

The MME was scaled to cover the US Midwest cropland over a total of ~46 million hectares that includes large areas in the 12 states of Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. Simulations were performed over 42 years (1980–2022) using the gridMET weather observations data (see ‘Data inputs’ section), with 1980 to 1990 used for setting initial conditions through a spin-up, and the remainder for data analysis.

Management scenarios and dynamic baselines

Eight management scenarios were simulated to evaluate SOC dynamics, four (1–4) with full nitrogen (FN, Table 1) fertilization and four (5–8) with reduced N fertilization (RN, Table S3). These scenarios constitute business-as-usual (BAU) and regenerative agricultural practices and crop rotations across the US Midwest, reflecting those most frequently studied in the literature at test site locations (see 'Experimental sites with observed data' section). Crop rotations include maize and soybean with or without a cover crop (Table S4) and with either conventional tillage or continuous no-till (Table S5). These scenarios allow comparisons of SOC change and N₂O emissions due to different management practices over the long-term at specific locations and at multiple scales using a dynamic baseline, a time and space dependent reference that accounts for natural and management-driven variations in SOC levels.

Data inputs

Simulations were conducted at 1/24th degree spatial resolution as determined by gridMET weather observations⁴⁰. Each grid (approximately 4 km²) represents an individual simulation-based unit (with a geolocated centroid) labeled with a unique identifier (UID). Grids were extracted for the 12 Midwest states involved in the analysis. The cropland area (maize or soybean grown for at least one year between 2008 to 2022) within each grid was calculated, as provided (30 m resolution) by the USDA National Agricultural Statistics Service Cropland Data Layer⁴¹. The percentage of cropland area per grid was calculated by dividing the cropland area by the total cropland area summed across the 12 Midwest states. Then with grids ordered from smallest to largest percent cropland, the percentages were summed to calculate the cumulative percent of cropland area. We selected UIDs that when combined covered 90% of the cropland area (Fig. S1). Selected UIDs were then categorized into discrete latitude bins to enable more straightforward identification and homogenous extraction of the final number of UIDs to be run. Within each bin, the main soil textures were identified, and their total area (km²) determined. Forty thousand total UIDs (balancing the computational power required and area representability) were then chosen across the region using a weighted random selection process, where the number of grids selected within each latitude and bin-texture combination was proportional to the area of cropland within the bin and the area covered by each texture. This selection process gave a higher probability for grids with a larger crop area to be selected.

For each UID, daily variables of maximum temperature, minimum temperature, precipitation accumulation, downward surface shortwave radiation, wind velocity, maximum relative humidity, and minimum relative humidity were downloaded from gridMET for 1980 to 2022.

Soil data for model inputs were taken from the USDA Natural Resources Conservation Service Gridded Soil Survey Geographic (gSSURGO)⁴² at 30 m spatial resolution. Variables extracted included bulk density, sand/silt/clay content, stone content, organic matter, and pH for each layer in the soil profile, using the most representative component percentage from the gSSURGO database. From these data, the following variables were calculated: organic carbon, total nitrogen, drained upper limit and lower limit⁴³, and saturated hydraulic conductivity⁴⁴. The dominant soil type within each weather grid was selected to represent that specific weather grid, excluding missing or shallower than 30 cm soil profiles. For each CSM, the SOC pools were initialized to match the mass of the initial SOC stock, with each CSM default value adopted to set their percentage distribution following the⁴⁵ procedure.

Annual planting and harvesting dates for maize and soybean for 1980–2022 were estimated from USDA NASS state-level weekly progress reports, spatially interpolated with county-level annual crop production area. Daily progress of planting or harvesting was linearly interpolated between the weekly progress reports to find when 50% of the state was planted or harvested. State centroids were computed, weighted by county-level crop production. The planting or harvesting dates were spatially interpolated between the weighted state centroids using a generalized additive model (GAM). The model is defined as:

$$PD = s(Lat, Lon) + ti(Lat * year) + ti(Lon * year) \quad (1)$$

where PD is the predicted date, $s(Lat, Lon)$ is the two-dimensional smoothing function with latitude and longitude as covariates, and $ti(Lat * year)$ and $ti(Lon * year)$ are tensor functions of the random effect of year and latitude and longitude, respectively. The model was applied to the weather grid centroids for each year, predicting planting and harvesting dates for each weather grid. Missing annual county-level production data was filled using the average across all available years. When all data was missing (e.g., soybean progress reports for 1979), the long-term average centroid and progress values were used in the model.

Nitrogen (N) fertilizer application amounts for maize were based on the USDA ARMS survey as reported in⁴⁶. The average N rate was used to represent each state, except for Wisconsin where the maximum reported N rate was used to account for the application of manure not included in the reported range of N rates, as indicated by the relatively low range. For Kansas and Nebraska, where data were not available, the same rate as a nearby state (Missouri and Iowa, respectively) was applied (Table S6).

Crop maturity regions identification and cultivars parametrization

Crop maturity regions (MR) were identified to better represent crop cultivar extent across the US Midwest. These regions were based on six latitude bins each for maize and soybean, with each bin determined by the crop growing degree days (GDD) accumulated at maturity. Initially, the GDD data for maize for each UID were derived from⁴⁷ and for soybean, computed from weather data (according to NASS planting and harvest dates). Then, the distinctions between each MR for each crop were computed by calculating the quantile-based breaks for the desired number of bins (6). These breaks were then used to categorize all UIDs into a distinct maturity region group (Fig. S2). The final GDD value for each MR was computed as the median of each group (Table S7).

For each CSM, crop cultivars were then parametrized using the GDD value according to the MR. In addition, to ensure consistency across CSM, harvest index (HI) and crop base temperature were set to standard values. Defaults for HI were initially set to 0.5 and 0.4 for maize and soybean, respectively (some CSMs then recalculated these HI values throughout the simulation period), and for crop base temperature were 8 °C and 10 °C for maize and soybean, respectively. Thus, this rationale was intended to set common main crop parameters across models to simulate crop yield (Figs. S3 and S4).

Multi-model ensemble computation

Daily simulation using the eight CSMs over 42 years across 40,000 UIDs in the US Midwest cropland generated a large dataset of more than 4 billion data values per output metric. For each model and UID, the output data was aggregated first by year and then grouped into a unique dataset. Then, for each management scenario, the yearly data were grouped by county. Outliers were detected using the interquartile range (IQR) method. This function calculates the range between the first (25th percentile) and third (75th percentile) quartile and defines outliers as data points that fall below the first quartile minus 1.5 times the IQR or above the third quartile plus 1.5 times the IQR. Once outliers were removed, the MME median (e-median) was used as the estimator of the ensemble model simulations.

When aggregating data by county, a weighted mean was used to account for the varying cropland area extents associated with each UID within the county.

$$\text{Weighted mean} = \frac{\sum_{i=1}^n w_i * x_i}{\sum_{i=1}^n w_i} \quad (2)$$

where x_i is the value of the i^{th} UID value and w_i is the weight associated with the i^{th} observation. Each weight is assigned based on the cropland area of each UID (i.e., grid). Equal weight was adopted across models. Except for Figs. 1 and 5, where single UID means are used, scenario outputs are reported throughout the manuscript as mean and standard deviation at the county level. The delta SOC change for the MME was calculated as the mean of each annual SOC change.

The global warming potential (GWP), reported as climate mitigation was quantified following the equation:

$$\text{Climate mitigation} [Mg CO_2 - eq ha^{-1} yr^{-1}] = GHG_{N_2O} - GHG_{\Delta SOC} \quad (3)$$

where GHG_{N_2O} represents the C_2O equivalent coming from N_2O emission⁴⁸, and $GHG_{\Delta SOC}$ the C_2O equivalent coming from the carbon sequestered or lost in the soil.

R software was used for statistical analysis and graphical visualization.

Multi-model ensemble validation

To test the MME performance in reproducing SOC change, we evaluated its accuracy and uncertainty in comparison with every individual CSMs. The MME and each CSM were evaluated comparing the simulated SOC stock and annual SOC change values for different management practices (i.e., experimental treatments) with observed (measured) values from 17 experimental sites (see 'Experimental sites with observed data' section). We tested MME and CSMs without performing a specific calibration for the soil-related parameters. Our goal was to demonstrate that the MME is suitable for upscaling over large areas where site specific soil parameter calibration is not feasible. Therefore, we rely primarily on the crop yield parametrization described in the section 'Crop maturity regions identification and cultivars parametrization' for the upscaling, and on the following section for the experimental sites. Lastly, MME uncertainties were also evaluated on the upscaled simulations across the US Midwest.

Experimental sites with observed data

To ensure consistent, high quality observed data sets were used for MME validation, experimental sites were selected from the literature when five data conditions were present: [1] initial soil conditions (prior to experimental treatments), including SOC stock, [2] multiple years of crop yield data, [3] a SOC time series (with either SOC stock values or BD and C% available for SOC stock calculation), [4] detailed management descriptions and ancillary information, and [5] the presence of maize or soybean, grown either continuously, or in a two-year rotation, or within an extended rotation with other crops (e.g., wheat and alfalfa). Sites were also intentionally selected to cover a wide range soil types and management, including conventional tillage, reduced tillage, no-till, different fertilization rates, and the presence of cover crops. Sampling depth up to 30 cm was preferred, but in 10 locations the depth of sampling was less (i.e., between 10 and 25 cm). In total, 17 experimental locations, with 52 site treatments, were located in North America (16 in the USA and 1 in Canada) (Fig. S5).

To simulate these sites, weather data was extracted from gridMET for US sites and NASA Power⁴⁹ for Canadian sites. Soil data were preferentially and primarily sourced from the publications, but if unavailable, gSSURGO data at the closest site were used. Water hydraulic soil properties were recalculated using pedotransfer functions as reported above. If not reported, initial soil water content was set equal to the drained upper limit (DUL) and initial inorganic nitrogen was set to 40 kg N ha⁻¹ for the entire soil profile, with ammonium and nitrate ratios assigned based on each model's default. Table S8 reports a full description of all sites and treatments.

To ensure appropriate biomass feedback from the crop residues and roots, each CSM was calibrated at each experimental site only against the observed crop yield data without involving any soil-related parameters. All the selected CSMs are designed to accurately predict crop yields, root and residue biomass (which includes

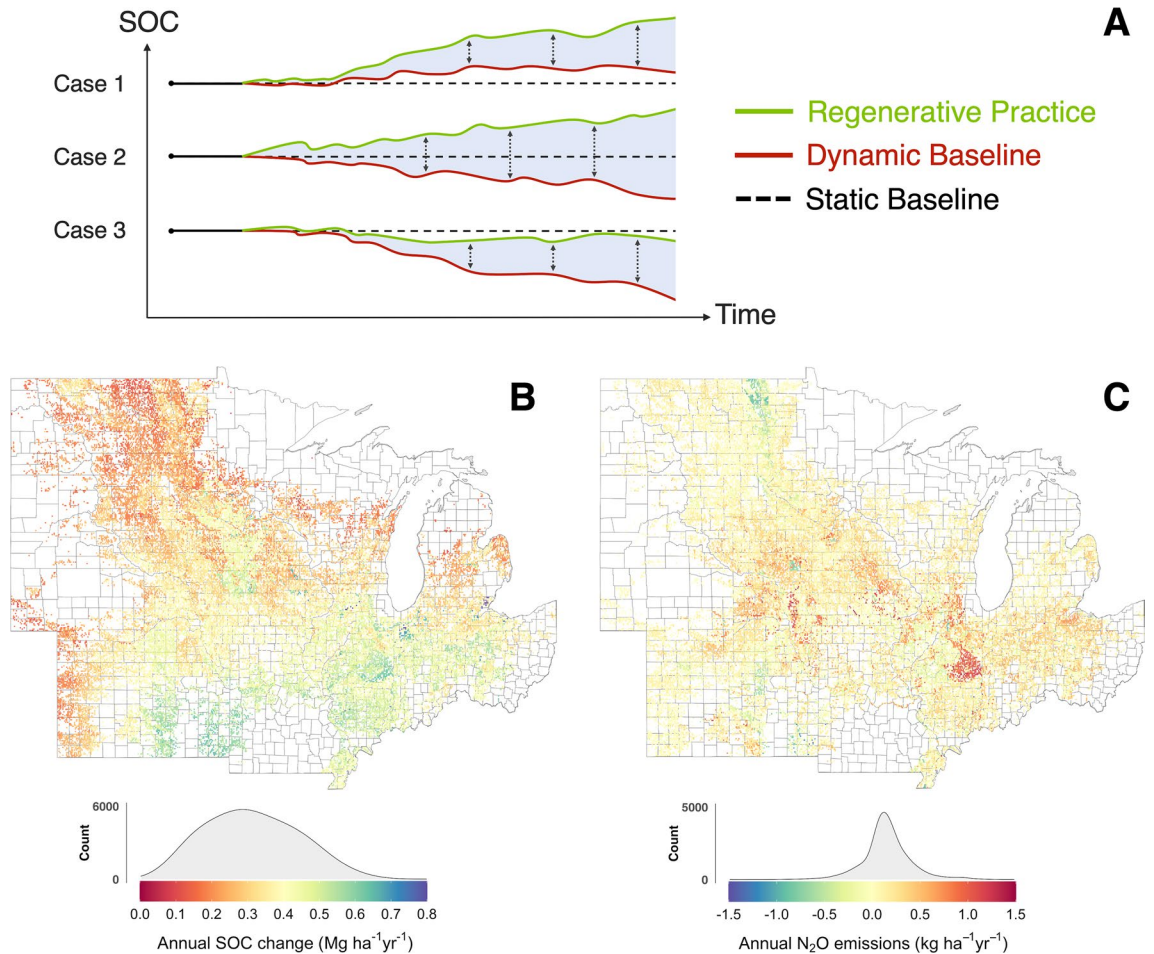


Fig. 1. Representation of the dynamic baseline concept and the mean MME annual SOC change and annual N_2O emissions difference between scenarios. **(A)** The counterfactual scenario dynamic baselines (red), as well as the regenerative practice lines (green), account for yearly climate variability and site-specific soil properties, while the static baseline reflects the assumption of steady-state conditions. Three different generic cases are provided as examples of potential errors if changes are compared to static baseline and not the dynamic baseline. **(B)** MME annual soil organic carbon change ($Mg\ ha^{-1}\ yr^{-1}$, 0–30 cm) and **(C)** annual N_2O emissions ($kg\ ha^{-1}\ yr^{-1}$) as the difference between Scenario 4 (NT FN CC, no-till with full nitrogen fertilization and cover crops) and Scenario 1 (CT FN, conventional till with full nitrogen fertilization). Each UID included in the study ($n = 40,000$) is reported as a single-colored dot. The gray density histograms represent the distribution of UIDs values across the US Midwest cropland Maps were created using RStudio v2025.05.0.496 <https://www.r-project.org>.

total above-ground biomass minus the grain yield) because of their critical importance for modeling long-term SOC dynamics. The RRMSE (Relative Root Mean Square Error, Eq. (4)) was used to evaluate individual CSMs calibration performance (Table S9). Most of the time, total biomass of cover crop was not reported, and default models' values were used to set its productivity. Data_S1, in the supplemental material, reports for each CSM and location, with the full set of parameters used, their default values, and the final values obtained from the calibration process.

$$RRMSE (\%) = 100 * \frac{\sqrt{\sum_{i=1}^n (P_i - O_i)^2}}{\bar{O}} \quad (4)$$

where P_i and O_i are predicted and observed data respectively, n the number of P/O pairs and \bar{O} the mean of observed data.

Multi-model ensemble and individual model accuracy and uncertainty analysis

Before evaluating the MME and each CSM's performance in terms of accuracy and uncertainty, we compared the measured delta SOC change values from the 52 experimental site treatments with ranges from a recent meta-analysis on regenerative practices⁵⁰ to identify values that were inconsistent with the literature range, i.e., for no-

till (-0.5 – 1.22 Mg ha⁻¹ yr⁻¹), conventional tillage (-0.47 – 0.5 Mg ha⁻¹ yr⁻¹), and reduced tillage (-0.97 – 0.65 Mg ha⁻¹ yr⁻¹). To do so, we calculated the weighted mean delta SOC change for the measured data following the same mathematical form of Eq. (2) (where x_i is the i th measured delta SOC change and w_i its weight based on time elapsed since the previous SOC observation), and noted for each site*treatment whether it was consistent with the data range (Table S10). For the accuracy and uncertainty analyses described below, individual paired comparisons between simulated and observed data were included only when observed weighted mean annual delta SOC change was consistent with the literature ranges reported above. For the MME and individual CSMs instead, the weighted mean annual delta SOC change was calculated as the average of the annual SOC changes over the simulated time period, with weights applied to account for shorter durations in the initial or final years of the simulation.

To evaluate MME and CSM accuracy, a composite score was developed by combining two metrics: the Savage score⁵¹, and the weighted standard deviation (s) of model ranks. The absolute value of the prediction errors (PE, Eq. (5)), computed as the difference between predicted (i.e., simulated) and measured delta SOC change was used to rank every j th model accuracy within each i th site*treatment combination.

$$PE = \frac{CF}{0.5} * (P_{j(i)} - O_i) \quad (5)$$

where $P_{j(i)}$ and O_i are the predicted and measured values of the j th model and i th site*treatments combination respectively, and CF is a correction factor that considers measurement uncertainties^{13,52}.

Each site*treatment combination ranking was then transformed into a Savage score. The total Savage score (S_j) for each model, calculated as the sum of its scores across all site*treatment combinations, was used to rank model accuracy, with greater weight assigned to models demonstrating consistently high performance. Rank variability, represented by the weighted standard deviation of ranks ($s_{rank(j)}$), quantified the stability of model rankings. The weight ($1/\text{rank}$) of the standard deviation was included to favor models that consistently fluctuate among top-ranking positions, rather than those that exhibit variability in lower-ranking positions.

Both metrics were normalized to a 0–1 scale and, for each model, the composite score (CS) was computed as:

$$CS = (0.5 * S_j) + 0.5 * (1 - s_{rank(j)}) \quad (6)$$

where S_j is the normalized sum of the Savage-scores assigned to the j th model and $s_{rank(j)}$ the weighted standard deviation of the j th model rank, while 0.5 ensures the same weight to both metrics. This approach ensured that the highest-ranked models demonstrated both superior accuracy and low variability.

To assess uncertainty, we used the mean squared error of prediction (MSEP) and its decomposition in prediction squared bias ($bias^2$) and prediction variance (var) as reported by Maiorano et al.⁵³ on the raw SOC stock variable. This metric was used to quantify MME uncertainty in comparison with individual CSMs. Since no calibration was performed against SOC data, measured data from experimental sites were used as the prediction dataset.

$$bias^2 = \frac{1}{N} \sum_{i=1}^N \left(Y_i - \frac{1}{M} \sum_{m=1}^M \hat{Y}_{m,i} \right)^2 \quad (7)$$

where N is the number of treatments, Y_i is the measured variable for the treatment i , $\hat{Y}_{m,i}$ is the simulated variable for the treatment i by the m CSM (or MME). The predicted variance is the variance of the values simulated by the population of m models (1 when individual CSMs are tested) averaged across treatments:

$$var = \frac{1}{N} \sum_{i=1}^N var_M(\hat{Y}_i) \quad (8)$$

where N is the number of treatments, \hat{Y}_i is the simulated variable for the treatment i . The estimate of MSEP can then be obtained as:

$$MSEP = bias^2 + var \quad (9)$$

We further evaluated how the uncertainty estimates changed with the number of models involved using the coefficient of variation of the multi-model ensemble e-median over the upscaled dataset in the US Midwest. To do so, we performed a bootstrap calculation (i.e., random sampling with replacement) for each value of M (number of models in the ensemble) from 1 to 8. For each ensemble size M we extracted 10,000 bootstrap samples of M models with replacement. This bootstrap sampling was repeated N times equal to the number of scenarios (1–4), and C times equal to the number of counties involved (934), considering single UID 30-year overall C gain/loss values within each county. The total number of samplings with replacement exceeded 35 million. The variation in the e-median across the bootstrap samples due to random model selection was computed with the coefficient of variation (CV):

$$CV(\hat{y}_{e\text{-median},M}) = \frac{1}{N} \frac{1}{C} \sum_{i=1}^N \sum_{j=1}^C \left(100 * \frac{sd_B(\hat{y}_{e\text{-median},i}^M)}{mean_B(\hat{y}_{e\text{-median},i}^M)} \right) \quad (10)$$

where $CV(\hat{y}_{e-median, Mj})$ is the coefficient of variation of e-median for the model ensemble of size M , sd_B and $mean_B$ are the standard deviation and the mean of the B (number of bootstrap sample) e-medians of the model ensemble of size M for the i th scenario (from 1 to 4) and the j th county.

Results

Regenerative practices impact on SOC change and N₂O emissions

The evaluation across ~46.2 million hectares of US Midwest cropland using MME results grouped by counties shows that in comparison with a dynamic baseline of conventional till without cover crops (CT), combined adoption of no-till and cover crops (NT CC; Scenario 4; Table 1) increases SOC accrual annually by 16.4 Tg C (on average 0.36 ± 0.12 Mg C ha⁻¹ yr⁻¹, Fig. 1B), while increasing the N₂O emissions by on average 0.16 kg ha⁻¹ yr⁻¹ (Fig. 1C). This rate of SOC stock increase is approximately twice that of when either no-till (NT; Scenario 3) or cover crops (CT CC; Scenario 2) are practiced independently on conventional till (i.e., 8.9 Tg C or 0.19 ± 0.13 Mg C ha⁻¹ yr⁻¹ for NT and 8.5 Tg C or 0.18 ± 0.14 Mg C ha⁻¹ yr⁻¹ for CC, Fig. 2). Similarly, under no-till, the introduction of a cover crop (NT CC) increases SOC accrual by 7.5 Tg C (on average 0.16 ± 0.10 Mg C ha⁻¹ yr⁻¹, Fig. 2C). See figures S6 and S7 for a full set of management comparisons. When N fertilizer is reduced to 75% of the BAU rate (Scenarios 5 vs. 1, 6 vs. 2, 7 vs. 3, and 8 vs. 4), SOC stock changes are negligible, averaging -0.02 Mg C ha⁻¹ yr⁻¹, while averaged across the RN scenarios, N₂O emissions decrease by 15% (-0.22 ± 0.45 kg ha⁻¹ yr⁻¹), and with cover crop adoption, increase by 11% (0.15 ± 0.51 kg ha⁻¹ yr⁻¹) and 12% (0.16 ± 0.45 kg ha⁻¹ yr⁻¹) under conventional and no-till management (Fig. S8), respectively. See figures S9 to S11 for a full set of county-level annual delta SOC and N₂O emissions per scenario.

Pedo-climatic variables and SOC change

Using MME analysis, locations with higher initial carbon stocks (see Fig. S12) typically show SOC loss (up to -0.43 Mg C ha⁻¹ yr⁻¹ in sandy soils under CT; Fig. 3). Conversely, soils with low initial carbon stocks typically show SOC increases under both CT and NT, particularly with cover crops (CT CC and NT CC). Larger, consistent SOC accrual rates were apparent across soil textures and initial carbon stock when no-till and cover crops (NT CC) were present together, with similar trends emerging for RN scenarios (Fig. S13). When individual CSMs were evaluated, changes in pedo-climatic variable inputs (initial SOC stock, clay content, and temperature and precipitation; Fig. 4) show no consistent impact on annual SOC change. As initial SOC stock increases, a decrease in the soil's capacity to store more C is observed (CSMs 1, 3, 4, 5, and 6), but some show variable (CSMs 7 and 8) and no (CSM 2) response to initial SOC stocks. Increasing clay content beyond ~30% shows an increase (CSMs 1 and 8), no change (CSMs 2, 4, and 7), or a decrease (CSM 3) in SOC stock. Soils at higher latitude show lower annual SOC change in CT scenarios but were inconsistent under NT management.

Climate mitigation potential

The climate mitigation potential shown in Fig. 5 reflects the extent to which SOC changes can offset N₂O emissions, based on the comparison between scenario 4 and scenario 1. In this comparison, no-till combined with cover crop (scenario 4) resulted in an average climate mitigation potential of 1.2 Mg CO₂-eq ha⁻¹ yr⁻¹ across the Midwest. This regenerative practice emits up to 2.5 Mg CO₂-eq ha⁻¹ yr⁻¹ less than its conventional tillage counterpart, although the magnitude of this benefit varies by geographic location (Fig. 5).

When each scenario is evaluated individually for its ability to offset N₂O emissions (Fig. S14), none consistently achieve full offset. Instead, the climate mitigation effect is highly dependent on site-specific pedo-climatic conditions. The mitigation potential tends to increase along a management gradient—from conventional tillage without cover crops to no-till with cover crops—highlighting the cumulative benefits of stacking regenerative practices.

Discussion

Standardizing SOC quantification

Adopting regenerative management practices can reduce the carbon footprint of row-crop agriculture, but to what degree and whether it is meaningful as part of climate mitigation efforts is debated⁵⁴. A major issue in






Scenario ID	Tillage	Crop rotation	N amount (%)	Cover crop	Acronym
1	Conventional				CT
2		maize - soybean	100% 	Rye 	CT CC
3	No-till				
4				Rye 	NT CC

Table 1. Scenarios simulated using a multi-model ensemble (MME) approach, representing different combinations of management practices involving tillage and cover cropping.

The full nitrogen fertilization (100%) refers to the Midwest state based “business as usual” average amount and was applied only for maize. A two-year crop rotation, with maize followed by soybean, was implemented in all scenarios. Scenario acronyms: CT (conventional till), NT (no-till), FN (full nitrogen), CC (cover crop). Graphical icons illustrate the different practices and are adopted throughout the figures.

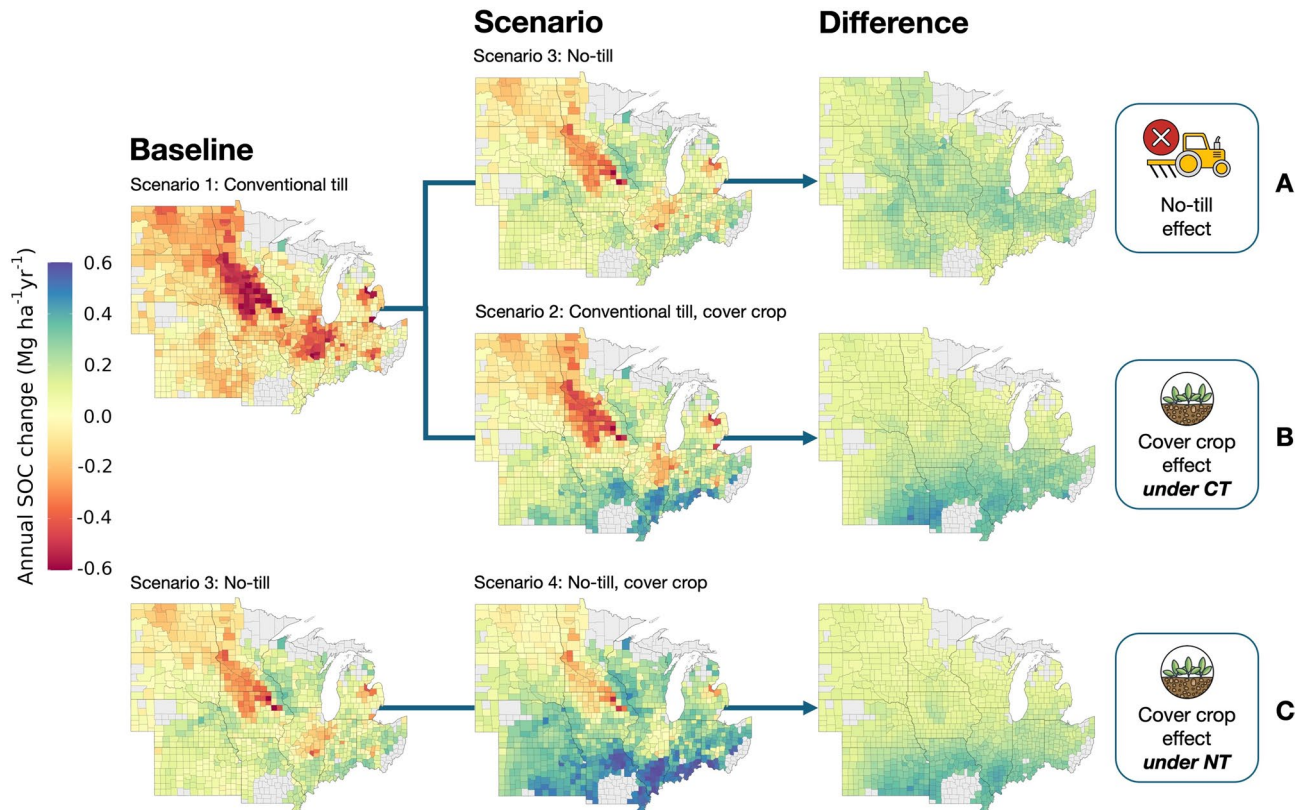


Fig. 2. Effect of different regenerative practices on county-based MME mean annual SOC change. (A) Effect of no-till on county annual soil organic carbon (SOC) change ($\text{Mg ha}^{-1} \text{ yr}^{-1}$, 0–30 cm) shown as the difference between Scenario 3 (NT FN, no-till with full nitrogen fertilization) and the baseline Scenario 1 (CT FN, conventional till and full nitrogen fertilization); (B) effect of cover crop on county annual soil organic carbon (SOC) change ($\text{Mg ha}^{-1} \text{ yr}^{-1}$, 0–30 cm) shown as the difference between Scenario 2 (CT FN CC, conventional till with full nitrogen fertilization and cover crop) and the baseline Scenario 1 (CT FN, conventional till and full nitrogen fertilization); (C) effect of cover crop under no-till on county annual soil organic carbon (SOC) change ($\text{Mg ha}^{-1} \text{ yr}^{-1}$, 0–30 cm) shown as the difference between Scenario 4 (NT FN CC, no-till, full nitrogen fertilization and cover crop) and the baseline Scenario 3 (NT FN, no-till with full nitrogen fertilization). Baselines serve as counterfactual benchmarks for comparison with different scenarios. Individual counties are represented by their mean weighted values, with weights accounting for the agricultural area percentage of each county's UID.

determining impact is the lack of standardization in quantifying SOC stock change⁵⁵; estimates for individual fields based on re-inventorying, even after extended periods of time and at high sampling densities, can be variable and inaccurate, with gains or losses in SOC stock often determined even when no change has occurred²⁶. These challenges are consistent with our findings from comparisons of SOC changes from individual long-term experimental treatments (Table S8) with meta-analysis ranges⁵⁰, where 38% (20 out of 52) of the peer-reviewed data fell outside the range, some showing 'unrealistic' rates of change (e.g., $8 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$; Table S10). The likelihood for spurious values is increased when SOC change is determined using consecutive short-term sampling (e.g., 4–5 years apart), typical with current C market approaches for paying producers, where contract lengths of only one year are common¹. Short-term contracts that eschew consideration of long-term SOC trajectories, risk exposure to measurements that are strongly impacted by seasonal trends (e.g., 'saw-tooth' dynamics). While acknowledging high quality SOC field measurements are needed to better understand SOC dynamics and help to initialize models, their use in determining SOC stock change is problematic for short-duration payment requirements, and effectively unfeasible for larger-scale projects.

Upscaling regenerative practices

By using high spatial resolution data aggregated to the US Midwest cropland, MME SOC change values are applicable and consistent with observations at multiple scales. As a management option, combining cover crops and no-till shows the greatest promise ($\sim 16.4 \text{ Tg C yr}^{-1}$) averaged over the total area investigated. With cover crop introduction, the MME SOC increase of $0.17 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$ ($\sim 8.0 \text{ Tg C yr}^{-1}$) across tillage types, is consistent with the meta-analyses of Peng et al.⁵⁶ and McClelland et al.⁵⁷ who found increases of $0.24 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$ and $0.21 \text{ Mg C ha}^{-1} \text{ yr}^{-1}$, respectively. Over similar total areas of US cropland, potential SOC accrual suggests that MME provides a reasonable balanced estimate between non-conservative values that use IPCC

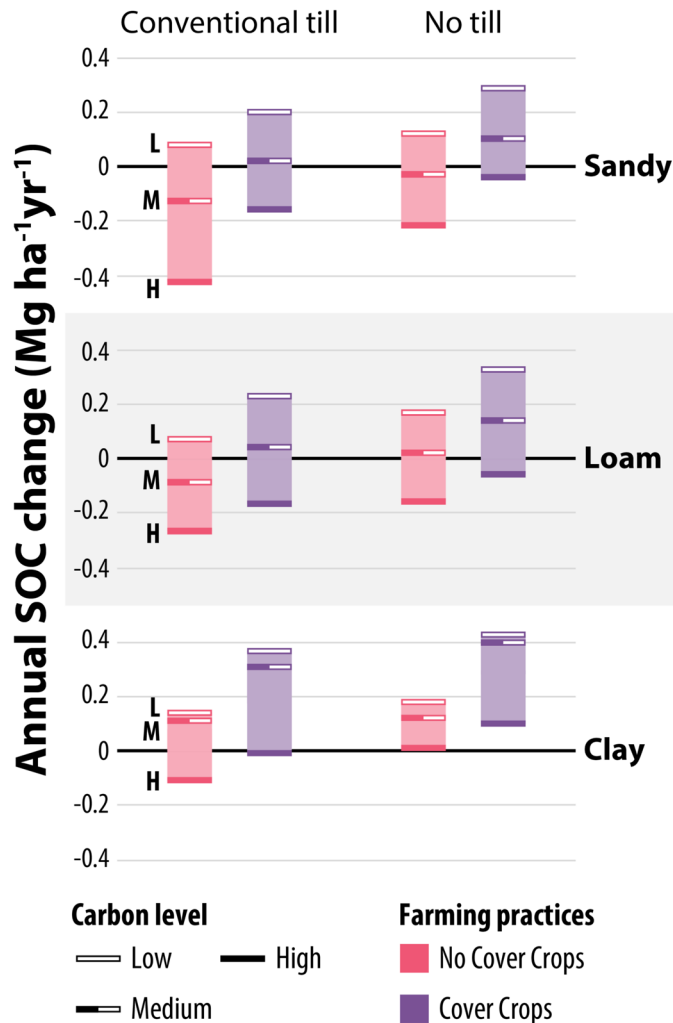


Fig. 3. MME mean annual SOC change rates across scenarios based on different soil textures and initial SOC stock levels across 46 M ha. Annual soil organic carbon (SOC) change ($\text{Mg ha}^{-1} \text{yr}^{-1}$, 0–30 cm) across scenarios (1 to 4, from left to right), soil texture (Sandy, Loam and Clay), and initial SOC stock level (Low, $< 40 \text{ Mg C ha}^{-1}$; Medium, $40\text{--}80 \text{ Mg C ha}^{-1}$; and High, $> 80 \text{ Mg C ha}^{-1}$). Annual SOC change values derived from all UIDs mean across the US Midwest cropland. All scenarios use maize-soybean crop rotation.

factors⁵⁸, 13.3 and $17.7 \text{ Tg C yr}^{-1}$ for no-till and cover crop adoption, respectively), and conservative evaluation that uses satellite imagery and longitudinal surveys⁵⁹, 5.3 Tg C yr^{-1} for cover crop adoption).

Even without site-specific validation, cumulative annual N_2O emissions estimated by the MME align with values reported in the literature or with IPCC-based estimates (see Table S11). For example, cover crop adoption increases N_2O emissions by on average $\sim 0.15 \text{ kg ha}^{-1} \text{yr}^{-1}$ across tillage managements (Fig. S8), an increase consistent with the variable results reported by Basche et al.⁶⁰. Prior multi-model studies have found significant uncertainty and divergence of N_2O emissions across models and have adopted different approaches that omit agricultural management and crop growth⁶¹. Although we make no claim, an MME approach that includes these factors may be well suited to ‘filter’ out data outliers, often reported in single-model studies due to the challenges of accurate simulation, especially on daily timescales.

Estimating N_2O emissions and changes in SOC is essential to fully evaluate the global warming potential (GWP, expressed as $\text{Mg CO}_2\text{-eq ha}^{-1} \text{yr}^{-1}$) of individual agricultural practices³¹. Balancing SOC sequestration against N_2O emissions provides a more accurate assessment of the climate mitigation potential, specifically by determining whether increases in SOC stock, when they occur, are sufficient to offset associated N_2O emissions completely¹².

When combined with a dynamic baseline, this approach further clarifies the actual climate mitigation potential of a practice by explicitly comparing it to a realistic counterfactual scenario. Figure 5 illustrates the climate mitigation potential of Scenario 4 (NT FN CC: no-till with full nitrogen fertilization and cover crops) relative to Scenario 1 (CT FN: conventional tillage with full nitrogen fertilization). Even when accounting for N_2O emissions, adopting no-till plus cover cropping results in an average net saving of $1.2 \text{ Mg CO}_2\text{-eq ha}^{-1} \text{yr}^{-1}$ compared to the conventional tillage baseline. Although direct comparison with the baseline remains essential, a separate assessment of Scenario 4 itself (Fig. S14) reveals that SOC accrual alone is not always sufficient to fully

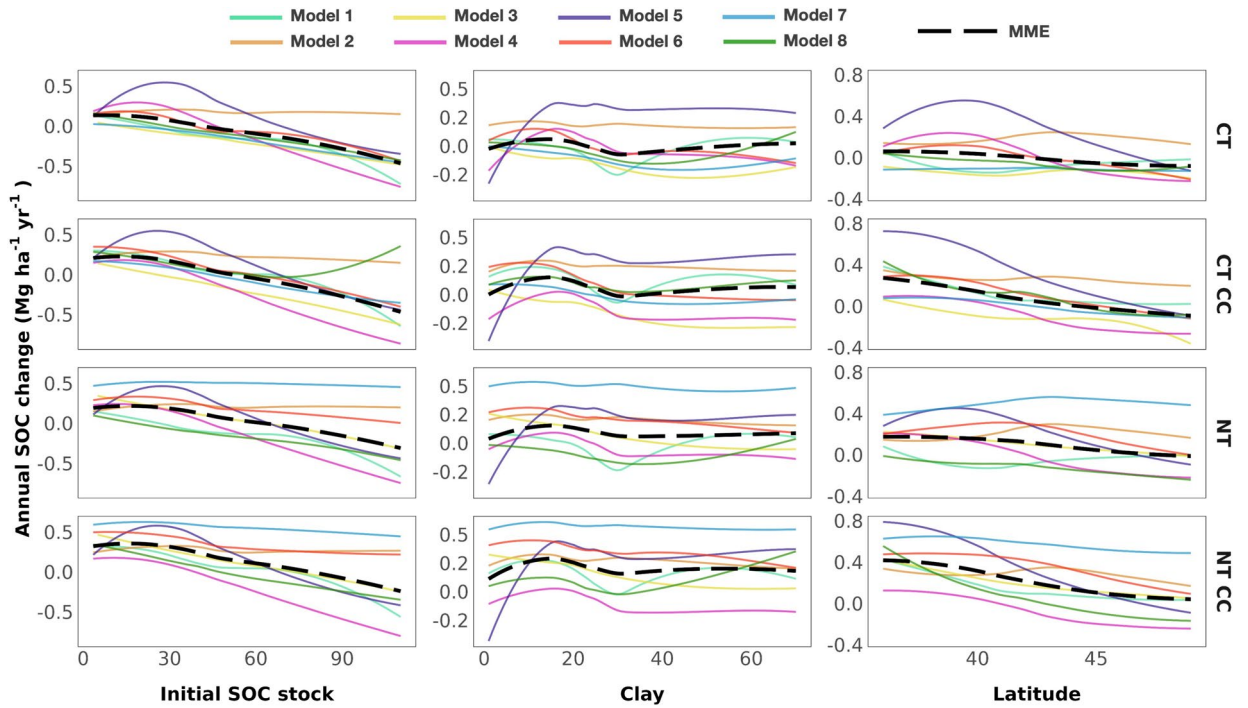


Fig. 4. MME mean annual SOC change response to variations in initial SOC stock, clay content, and latitude across 46 M ha. Each line, represented by a unique color, corresponds to an individual model, with the black dotted line representing the MME median. They represent individual model annual SOC change ($\text{Mg ha}^{-1} \text{yr}^{-1}$, 0–15 cm) response to variations in initial SOC stock (Mg ha^{-1}), clay content (%), and latitude ($^{\circ}$). The different scenarios (1 to 4 from top to bottom) are displayed at the right end of each row. Scenario acronyms: CT (conventional till), NT (no-till), FN (full nitrogen), CC (cover crop).

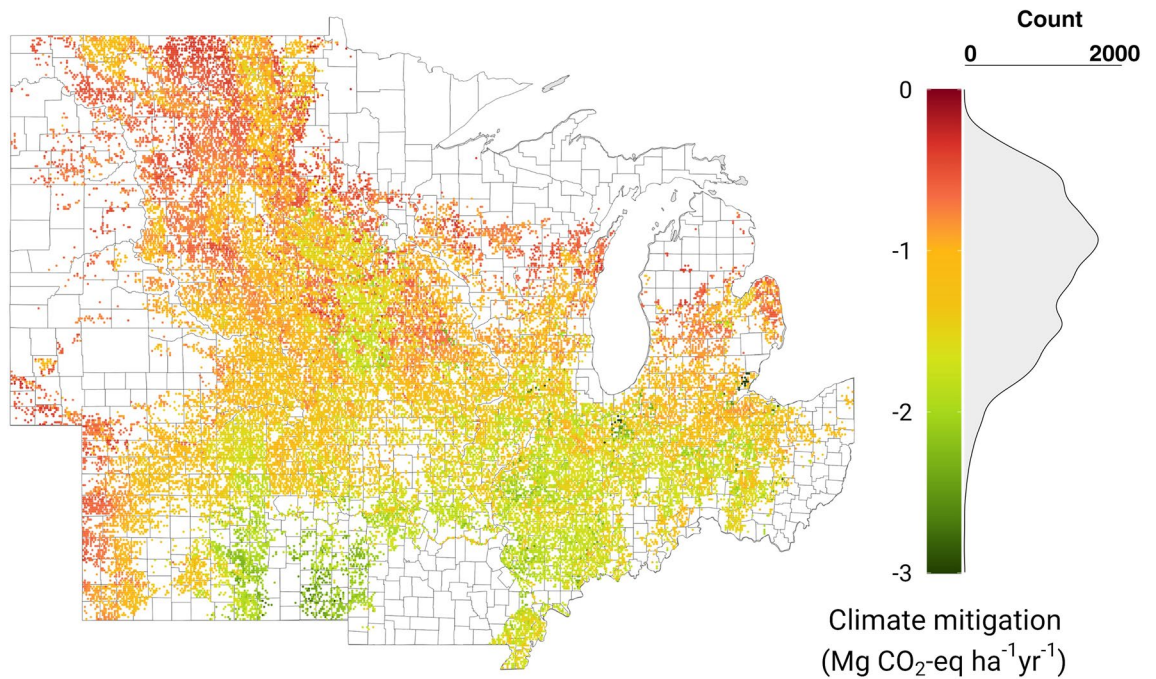


Fig. 5. Climate mitigation ($\text{Mg CO}_2\text{-eq ha}^{-1} \text{yr}^{-1}$) of Scenario 4 (NT FN CC, no-till with full nitrogen fertilization and cover crops) compared to Scenario 1 (CT FN, conventional till with full nitrogen fertilization). Climate mitigation accounts for N_2O emission and SOC change as reported in Eq. (3). Each UID included in the study ($n = 40,000$) is reported as a single-colored dot. The gray density histograms represent the distribution of UID values across the US Midwest cropland.

Overall performance ranking	Model ID	Savage score of the absolute prediction error rank	Weighted standard deviation of the absolute prediction error rank	Composite performance score
1	MME	39.3 (0.90)	2.07 (0.00)	0.948
2	3	41.5 (1.00)	2.17 (0.11)	0.946
3	7	37.3 (0.80)	2.23 (0.17)	0.813
4	6	34.9 (0.68)	2.18 (0.12)	0.781
5	2	32.9 (0.58)	2.31 (0.26)	0.663
6	5	28.2 (0.36)	2.80 (0.79)	0.285
7	1	27.5 (0.32)	2.80 (0.79)	0.268
8	8	25.4 (0.22)	3.00 (1.00)	0.108
9	4	20.9 (0.00)	2.81 (0.79)	0.106

Table 2. Comparison of the performance of individual CSMs and the MME across measured experimental sites and treatments. The Savage score was calculated by ranking the prediction errors between measured and simulated delta SOC changes within each site and treatment combination (Table S8). The weighted standard deviation reflects the variability in performance rankings across locations and treatments for each model with respect of the top rank position compared to the bottom ones. The composite score represents the overall performance, assigning greater weight to higher Savage scores and lower standard deviations, and is calculated using Eq. (6). Higher composite scores indicate better overall performance. Normalized values are presented within parenthesis for the Savage score and weighted standard deviation respectively.

offset N₂O emissions. Nevertheless, Scenario 4 still provides, on average, a modest negative balance (computed as in equation 3) of approximately 30 kg CO₂-eq ha⁻¹ retained annually.

A potential improvement to Scenario 4 involves reducing the nitrogen fertilizer application rate by 25%, as explored in Scenario 8 (Table S3). This comparison becomes extremely useful when evaluating whether potential yield reductions (due to decreased N supply) are justified by the corresponding reductions in N₂O emissions. In other words, the analysis helps clarify whether the expected reduction in N₂O emissions would sufficiently compensate for the likely decrease in SOC accrual due to reduced crop residues returned to the soil. Results show that the additional mitigation potential is minimal (~10 kg CO₂-eq ha⁻¹ yr⁻¹) when comparing Scenario 8 to Scenario 4. Furthermore, average maize and soybean yields under Scenario 8 are lower by approximately 150 kg ha⁻¹ (±390 kg ha⁻¹) compared to yields obtained under optimal fertilization (Scenario 4). Thus, identifying the best overall management practice requires an economic justification that balances yield response against fertilizer cost savings. A straightforward economic analysis could clarify whether cost savings from reduced nitrogen fertilizer inputs adequately compensate for potential yield reductions, at least until farmers receive direct economic incentives for additional CO₂ abatement efforts. Whether industry initiatives or emerging carbon markets will support such incentives remains to be determined and lies beyond the scope of this analysis.

In addition to the broad-scale assessment across the entire Midwest region, we believe further analysis conducted within homogeneous pedo-climatic zones (such as Major Land Resource Areas, MLRAs) would provide even greater value. Although beyond the immediate objectives of the present study, such localized analyses could significantly influence agricultural decision-making. Instead of solely pursuing maximum yield production, management practices could better reflect realistic land-use potential, optimizing the trade-offs between achievable crop yields and associated greenhouse gas emissions.

Multi-model ensemble performance

When tested against long-term experimental trials across the US Midwest, characterized by substantial differences in soil types and agricultural management, the MME approach provides higher accuracy, and lower uncertainty in estimating SOC change (Table 2). Its performance reduces the biases associated with short-term sampling and the use of an individual CSM. In comparison with individual CSMs, MME best simulates changes in SOC across the experimental sites and treatments, with the highest overall composite score of 0.948 (Table 2). The MME Savage score of 39.3, indicates that it frequently ranks among the top models in terms of accuracy (i.e., lowest absolute prediction error) and has the lowest weighted standard deviation of the absolute prediction error rank (2.07; Table 2). Uncertainty assessment using MSEP and its decomposition into squared bias and variance components, shows MME consistently exhibits the lowest uncertainty, evident when compared with the full CSM population (Fig. S15) and with individual CSMs (Fig. S16). Thus, an estimate from an individual CSMs that is ‘best’ at one field may not be so at another, often for reasons that cannot readily be determined; individual CSMs have varying constraints and needs and are better suited for specific contexts¹⁴.

When upscaled to the US Midwest, the coefficient of variation for the delta SOC change e-median decreases as the number of models increase, from 99% with one CSM to 36% with eight CSMs (Fig. S17). The notable disagreement among individual CSMs at large scale for the same scenarios (Fig. S18–S22), underscores the quantitative lottery of individual CSMs selection for a particular application. Thus, MME provides a more consistent and reliable output of long-term SOC change across multiple scales in agreement with prior studies^{13,20}.

Dynamic baselines and flexible scales

Inconsistent and contradictory outputs highlight the ongoing risk of using a single CSM to provide the robust accounting needed to support agricultural carbon market projects across broad regions. A large cause of this incongruity is the known and somewhat intractable challenge of determining additionality through the choice of a suitable BAU baseline: the counterfactual scenario that would have occurred without management intervention. This choice can have a significant influence on the degree of SOC change, and, by extension, on how many carbon credits are generated; if inaccurate, too many or too few may be issued, calling into question legitimacy³⁰. Indeed, only carbon that is additionally stored in soils or that is additional compared to a BAU scenario can be relevant for climate change mitigation⁶².

Usually, projected baselines are a 'best guess' manufactured scenario that considers steady-state historical conditions for a specific area (Fig. 1A) without accounting for within season climate variability and site-specific soil properties. Attributing SOC change to management using a static baseline is problematic because it is prone to errors if SOC changes in the counterfactual BAU scenario are not quantified. The use of process-based models avoids this issue by using dynamic baselines that are temporally and spatially adjusted to capture the impacts of fluctuating environmental variables. This allows quantification of the yearly SOC change difference between an adopted regenerative management practice and its dynamic counterfactual baseline to provide a more credible estimate of carbon gained or lost. Dynamic baseline scenarios were determined for every UID to allow for direct comparisons at the same location. This is essential for evaluating practice-level impacts on SOC accrual under the same pedo-climatic conditions, as a direct comparison with a business-as-usual (BAU) scenario is often unavailable or impractical on a field basis. For instance, when a farmer transitions from conventional tillage to no-till and measures SOC levels at the time of conversion, comparing SOC changes under no-till after 5 or 10 years against the initial SOC value (when conventional tillage ended) fails to account for what SOC levels would have been had conventional tillage continued on that field. Only through modeling counterfactual scenarios via dynamic baselines can realistic SOC differences between treatments be determined. These clarifications emphasize the importance of dynamic baselines and improve the overall clarity of our methodology and findings. MMEs can explore large pedo-climatic variability with a regionally consistent framework. Units of scale are flexible – they can be an agroecological zone with similar soils, weather, and agricultural practices, such as USDA Major Land Resource Areas⁶³ or a political jurisdiction such as county or state⁶⁴. Irrespective, estimates of SOC change can be quantified relative to a dynamic baseline for the scale chosen, and the results reported in readily accessible lookup tables, with the MMEs potential to generate and retrieve this information a valuable service, when reference values are needed for carbon credit payments or a regional policy document. As an example, the choice of county scale as the baseline jurisdictional unit allows their relatively uniform distribution across the US Midwest to capture soil and climate variability, and their administrative structure to provide a practical, credible framework for investment and implementation of regenerative practices for GHG abatement.

Opportunities and limitations for multi-model ensemble application

As part of its effort to reduce GHG emissions, the US has promoted programs to enhance soil carbon sinks⁶⁵. The expectation is that emissions reductions will occur through the adoption of regenerative practices on cropland⁶⁶, with market-based interventions best positioned to achieve this⁶⁷. Voluntary offset and inset initiatives predominate in the US agricultural sector; however, carbon credits generated by agriculture constitute only ~1% of the total volume transacted⁶⁸. Despite very high rates of producer carbon market awareness, participation rates are very low⁶⁹, in large part a reflection of the barriers that currently hinder engagement, primarily high enrollment costs driven by quantification, recordkeeping, and data analysis, scale and aggregation limitations, and verification, which are insufficiently compensated for by C change payments^{67,70}. Cultural and social factors can also deter participation, including confusion about the assortment of available market programs^{1,17,71}.

Use of an MME accounting tool could alleviate many producer participation barriers, ease onerous MMRV requirements, provide confidence in SOC stock change quantification at different scales, and in doing so address several strategic US policy priorities⁷². At the initial stage, producers with eligible fields would not be required to provide detailed agronomic data to participate, thereby minimizing record-keeping burdens, auditing requirements, and potential privacy concerns. However, when such data are available, they can be incorporated to enhance the accuracy and performance of the MME by improving the quality of model inputs.

In effect, an MME approach provides multi-scale, pre-run, practice-based, dynamic baselines for determining emission factors with low uncertainty levels. MMEs may be a practical, intermediate approach for payment for practice and payment for output mechanisms. As a decision support tool, MMEs can also allow producers or project developers to quickly screen lands targeted for regenerative practices to determine their economic and program viability⁶⁷.

Of the few registered agricultural land management (ALM) projects, most have generated low density (per hectare) and low volume (i.e., small area) credit numbers when compared to other sector project types^{68,73}. Unless aggregation (combining individual fields or small projects) to form large scale projects is allowed, producers, particularly those with limited resources or opportunity, find cost barriers insurmountable. While aggregation in varied design is allowed in current agricultural protocols and results in verification cost reduction, it does not reduce quantification costs, which scale with the area enrolled.

Similarly, approaches and accounting rules have been recommended to address issues of permanence, leakage, and additionality, inherent with small-scale, individual ALM projects⁷³. However, aggregated projects where management interventions are accounted for at larger scales reduce these concerns⁷⁴. With sufficiently large spatial and temporal scales, additionality can be demonstrated through reduced emissions below recent regional baseline trends. Larger project area coverage and longer project times can alleviate practice reversal risks, unintentional or otherwise, by pooling across broader scales and stakeholders. Large programs can also mitigate leakage risks by addressing underlying economic drivers of emissions and can reduce jurisdictional

risk by directly accounting for all emissions shifts inside the administrative unit⁶⁴. Measuring BAU baselines on a project-by-project basis is not economically feasible as emissions reductions are too small relative to costs. Technology-adjusted historical baselines designed in accordance with regional dynamics seem a superior option⁷⁵, and one that MMEs are well-suited to provide.

The present study comes with some limitations and possibilities for future work. Many process-based models simulate SOC dynamics across defined depth layers, yet the representation of vertical SOC redistribution—particularly under no-till systems—remains a challenge. Most models in this ensemble simulated the full soil profile, but one model was limited to 20 cm. While topsoil (0–30 cm), especially under no-till management, accounts for ~80% of the total management effects⁷⁶, it may miss subsoil contributions.

We also recognize that in many models, the effect of tillage on SOC is treated empirically, often modifying decomposition rates without explicitly simulating changes in soil physical properties. Nevertheless, especially in no-till, SOC stock stratification is mainly due to crop rootsatification, something that all the models fully represent. Bioturbation, which promotes SOM redistribution due to biotic activity, but most of the models do not include it yet.

Overall, the MME approach helps mitigate individual model biases, but its reliability still depends on the quality of input data. While large-scale projects can help harmonize datasets, local-scale agronomic details (e.g., tillage intensity, residue management, fertilization methods) remain crucial for accurate single CSM testing and calibration.

Despite the uncertainties, our study represents an important step forward, as it is the first to examine prospective responses on practice adoption at such scale and resolution.

Conclusions

MME results generate low uncertainty estimates of climate mitigation for common management practices that can be used in C market accounting at multiple scales. This removes the exposure of market actors to accusations of model selection and reduces the uncertainty of practice change outcomes. With MMEs, the fine granularity and consistent framework of input data across a large area allow for a flexible choice of scale for model output aggregation. This and the availability of practice-based dynamic baselines, along with decision support capabilities, can allow MMEs to reduce barriers to carbon market participation, and ameliorate long-standing concerns related to additionality, leakage, and permanence. In doing so, MME addresses several policy priorities and has a potential role in enhancing adoption of regenerative agriculture at scale.

Data availability

The data that support the findings of this study are openly available as supplemental material.

Received: 25 March 2025; Accepted: 20 June 2025

Published online: 11 July 2025

References

- Wongpiyabovorn, O., Plastina, A. & Crespi, J. M. Challenges to voluntary Ag carbon markets. *Appl. Econ. Perspect. Policy* **45**(2), 1154–1167. <https://doi.org/10.1002/aep.13254> (2023).
- Wiesmeier, M. et al. Soil organic carbon storage as a key function of soils—A review of drivers and indicators at various scales. *Geoderma* **333**, 149–162. <https://doi.org/10.1016/j.geoderma.2018.07.026> (2019).
- Jones, J. W. et al. Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. *Agric. Syst.* **155**, 269–288. <https://doi.org/10.1016/j.agsy.2016.09.021> (2017).
- Basso, B. et al. Soil organic carbon and nitrogen feedbacks on crop yields under climate change. *Agric. Environ. Lett.* **3**(1), 180026. <https://doi.org/10.1002/agj2.70018> (2018).
- Campbell, E. E. & Paustian, K. Current developments in soil organic matter modeling and the expansion of model applications: A review. *Environ. Res. Lett.* **10**(12), 123004. <https://doi.org/10.1088/1748-9326/10/12/123004> (2015).
- Grace, P. R., Colunga-Garcia, M., Gage, S. H., Robertson, G. P. & Safir, G. R. The potential impact of agricultural management and climate change on soil organic carbon of the north central region of the United States. *Ecosystems* **9**(5), 816–827. <https://doi.org/10.1007/s10021-004-0096-9> (2006).
- Del Grosso, S. J. et al. DAYCENT National-Scale Simulations of Nitrous Oxide Emissions from Cropped Soils in the United States. *J. Environ. Qual.* **35**(4), 1451–1460. <https://doi.org/10.2134/jeq2005.0160> (2006).
- Grace, P. & Robertson, G. P. Soil carbon sequestration potential and the identification of hotspots in the eastern Corn Belt of the United States. *Soil Sci. Soc. Am. J.* **85**(5), 1410–1424. <https://doi.org/10.1002/saj2.20273> (2021).
- Sulman, B. N. et al. Multiple models and experiments underscore large uncertainty in soil carbon dynamics. *Biogeochemistry* **141**(2), 109–123. <https://doi.org/10.1007/s10533-018-0509-z> (2018).
- Wallach, D. et al. The chaos in calibrating crop models: Lessons learned from a multi-model calibration exercise. *Environ. Model. Softw.* **145**, 105206. <https://doi.org/10.1016/j.envsoft.2021.105206> (2021).
- Rosenzweig, C. et al. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proc. Natl. Acad. Sci.* **111**(9), 3268–3273. <https://doi.org/10.1073/pnas.1222463110> (2014).
- Paustian, K. et al. Climate-smart soils. *Nature* **532**(7597), 49–57. <https://doi.org/10.1038/nature17174> (2016).
- Riggers, C. et al. Multi-model ensemble improved the prediction of trends in soil organic carbon stocks in German croplands. *Geoderma* **345**, 17–30. <https://doi.org/10.1016/j.geoderma.2019.03.014> (2019).
- Garsia, A., Moinet, A., Vazquez, C., Creamer, R. E. & Moinet, G. Y. K. The challenge of selecting an appropriate soil organic carbon simulation model: A comprehensive global review and validation assessment. *Glob. Change Biol.* **29**(20), 5760–5774. <https://doi.org/10.1111/gcb.16896> (2023).
- Jägermeyr, J. et al. Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nat. Food* **2**(11), 873–885. <https://doi.org/10.1038/s43016-021-00400-y> (2021).
- Martre, P. et al. Multimodel ensembles of wheat growth: Many models are better than one. *Glob. Change Biol.* **21**(2), 911–925. <https://doi.org/10.1111/gcb.12768> (2015).
- Oldfield, E. E. et al. Crediting agricultural soil carbon sequestration. *Science* **375**(6586), 1222–1225. <https://doi.org/10.1126/science.abl7991> (2022).

18. Sándor, R. et al. Ensemble modelling of carbon fluxes in grasslands and croplands. *Field Crops Res.* **252**, 107791. <https://doi.org/10.1016/j.fcr.2020.107791> (2020).
19. Fuchs, K. et al. Multimodel evaluation of nitrous oxide emissions from an intensively managed grassland. *J. Geophys. Res. Biogeosci.* **125**(1), e2019JG005261. <https://doi.org/10.1029/2019JG005261> (2020).
20. Farina, R. et al. Ensemble modelling, uncertainty and robust predictions of organic carbon in long-term bare-fallow soils. *Glob. Change Biol.* **27**(4), 904–928. <https://doi.org/10.1111/gcb.15441> (2021).
21. Lecocq, F. et al. Mitigation and development pathways in the near to mid-term. In *Climate Change 2022—Mitigation of Climate Change* 1st edn (ed. Lecocq, F.) 409–502 (Cambridge University Press, 2022). <https://doi.org/10.1017/9781009157926.006>.
22. Brinton, W. et al. An inter-laboratory comparison of soil organic carbon analysis on a farm with four agricultural management systems. *Agron. J.* <https://doi.org/10.1002/agj2.70018> (2025).
23. Even, R. J., Machmuller, M. B., Lavalley, J. M., Zelikova, T. J. & Cotrufo, M. F. Large errors in soil carbon measurements attributed to inconsistent sample processing. *SOIL* **11**(1), 17–34. <https://doi.org/10.5194/soil-11-17-2025> (2025).
24. Goidts, E., Van Wesemael, B. & Crucifix, M. Magnitude and sources of uncertainties in soil organic carbon (SOC) stock assessments at various scales. *Eur. J. Soil Sci.* **60**(5), 723–739. <https://doi.org/10.1111/j.1365-2389.2009.01157.x> (2009).
25. Stanley, P., Spertus, J., Chiartas, J., Stark, P. B. & Bowles, T. Valid inferences about soil carbon in heterogeneous landscapes. *Geoderma* **430**, 116323. <https://doi.org/10.1016/j.geoderma.2022.116323> (2023).
26. Bradford, M. A. et al. Testing the feasibility of quantifying change in agricultural soil carbon stocks through empirical sampling. *Geoderma* **440**, 116719. <https://doi.org/10.1016/j.geoderma.2023.116719> (2023).
27. Potash, E. et al. How to estimate soil organic carbon stocks of agricultural fields? Perspectives using ex-ante evaluation. *Geoderma* **411**, 115693. <https://doi.org/10.1016/j.geoderma.2021.115693> (2022).
28. Zhou, W. et al. How does uncertainty of soil organic carbon stock affect the calculation of carbon budgets and soil carbon credits for croplands in the U.S. Midwest?. *Geoderma* **429**, 116254. <https://doi.org/10.1016/j.geoderma.2022.116254> (2023).
29. Bossio, D. A. et al. The role of soil carbon in natural climate solutions. *Nat. Sustain.* **3**, 391–398. <https://doi.org/10.1038/s41893-020-0491-z> (2020).
30. Paustian, K. et al. Quantifying carbon for agricultural soil management: From the current status toward a global soil information system. *Carbon Manag.* **10**(6), 567–587. <https://doi.org/10.1080/17583004.2019.1633231> (2019).
31. Sainju, U. M. Net global warming potential and greenhouse gas intensity. *Soil Sci. Soc. Am. J.* **84**, 1393–1404. <https://doi.org/10.1002/saj2.20152> (2020).
32. Keating, B. A. et al. An overview of APSIM, a model designed for farming systems simulation. *Eur. J. Agron.* **18**(3–4), 267–288. [https://doi.org/10.1016/S1161-0301\(02\)00108-9](https://doi.org/10.1016/S1161-0301(02)00108-9) (2003).
33. Perego, A., Giussani, A., Sanna, M., Fumagalli, M., Carozzi, M., Brenna, S., & Acutis, M. *The ARMOSA simulation crop model: Overall features, calibration and validation results.* 16 (2013).
34. Stöckle, C. O., Donatelli, M. & Nelson, R. CropSyst, a cropping systems simulation model. *Eur. J. Agron.* **18**(3–4), 289–307. [https://doi.org/10.1016/S1161-0301\(02\)00109-0](https://doi.org/10.1016/S1161-0301(02)00109-0) (2003).
35. Parton, W. J., Hartman, M., Ojima, D. & Schimel, D. DAYCENT and its land surface submodel: Description and testing. *Glob. Planet. Change* **19**(1–4), 35–48. [https://doi.org/10.1016/S0921-8181\(98\)00040-X](https://doi.org/10.1016/S0921-8181(98)00040-X) (1998).
36. Jones, J. W. et al. The DSSAT cropping system model. *Eur. J. Agron.* **18**(3–4), 235–265. [https://doi.org/10.1016/S1161-0301\(02\)00107-7](https://doi.org/10.1016/S1161-0301(02)00107-7) (2003).
37. Izaurrealde, R. C., Williams, J. R., McGill, W. B., Rosenberg, N. J. & Jakas, M. C. Q. Simulating soil C dynamics with EPIC: Model description and testing against long-term data. *Ecol. Model.* **192**(3–4), 362–384. <https://doi.org/10.1016/j.ecolmodel.2005.07.010> (2006).
38. Basso, B. & Ritchie, J. T. Simulating crop growth and biogeochemical fluxes in response to land management using the SALUS model. In *The Ecology of Agricultural Landscapes: Long-Term Research on the Path to Sustainability* (eds Basso, B. & Ritchie, J. T.) 252–274 (Oxford University Press, 2015).
39. Beaudoin, N. et al. Stics soil crop model. *éditions Quae* <https://doi.org/10.35690/978-2-7592-3679-4> (2023).
40. Abatzoglou, J. T. Development of gridded surface meteorological data for ecological applications and modelling. *Int. J. Climatol.* **33**(1), 121–131. <https://doi.org/10.1002/joc.3413> (2013).
41. NASS, National Agricultural Statistics Service, & USDA, United States Department of Agriculture. *Cropland Data Layer: USDA NASS.* <https://croplandcros.scinet.usda.gov/>. (2024).
42. Soil Survey Staff. *Gridded Soil Survey Geographic (gSSURGO) Database for the Conterminous United States.* <https://gdg.sc.egov.usda.gov/>. (2020).
43. Ritchie, J. T., Gerakis, A. & Suleiman, A. Simple model to estimate field-measured soil water limits. *Trans. ASAE* **42**(6), 1609–1614. <https://doi.org/10.13031/2013.13326> (1999).
44. Suleiman, A. A. & Ritchie, J. T. Estimating saturated hydraulic conductivity from soil porosity. *Trans. ASAE* **44**(2), 235–339. <https://doi.org/10.13031/2013.4683> (2001).
45. Basso, B. et al. Procedures for initializing soil organic carbon pools in the DSSAT-CENTURY Model for agricultural systems. *Soil Sci. Soc. Am. J.* **75**:69–78 (2011).
46. Basso, B., Shuai, G., Zhang, J. & Robertson, G. P. Yield stability analysis reveals sources of large-scale nitrogen loss from the US Midwest. *Sci. Rep.* **9**(1), 5774. <https://doi.org/10.1038/s41598-019-42271-1> (2019).
47. Abendroth, L. J. et al. Lengthening of maize maturity time is not a widespread climate change adaptation strategy in the US Midwest. *Glob. Change Biol.* **27**(11), 2426–2440. <https://doi.org/10.1111/gcb.15565> (2021).
48. IPCC. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, In press. <https://doi.org/10.1017/9781009157896> (2021).
49. NASA, Langley Research Center (LaRC). *POWER Data.* <https://power.larc.nasa.gov>. (2024).
50. Bai, X. et al. Responses of soil carbon sequestration to climate-smart agriculture practices: A meta-analysis. *Glob. Change Biol.* **25**(8), 2591–2606. <https://doi.org/10.1111/gcb.14658> (2019).
51. Savage, I. R. Contributions to the theory of rank order statistics—the two-sample case. *Ann. Math. Stat.* **27**(3), 590–615. <https://doi.org/10.1214/aoms/1177728170> (1956).
52. Daren Harmel, R. & Smith, P. K. Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *J. Hydrol.* **337**(3–4), 326–336. <https://doi.org/10.1016/j.jhydrol.2007.01.043> (2007).
53. Maiorano, A. et al. Crop model improvement reduces the uncertainty of the response to temperature of multi-model ensembles. *Field Crops Res.* **202**, 5–20. <https://doi.org/10.1016/j.fcr.2016.05.001> (2017).
54. Schlesinger, W. H. Biogeochemical constraints on climate change mitigation through regenerative farming. *Biogeochemistry* **161**(1), 9–17. <https://doi.org/10.1007/s10533-022-00942-8> (2022).
55. Smith, P. et al. How to measure, report and verify soil carbon change to realize the potential of soil carbon sequestration for atmospheric greenhouse gas removal. *Glob. Change Biol.* **26**(1), 219–241. <https://doi.org/10.1111/gcb.14815> (2020).
56. Peng, Y. et al. Maximizing soil organic carbon stocks under cover cropping: Insights from long-term agricultural experiments in North America. *Agric. Ecosyst. Environ.* **356**, 108599. <https://doi.org/10.1016/j.agee.2023.108599> (2023).

57. McClelland, S. C., Paustian, K. & Schipanski, M. E. Management of cover crops in temperate climates influences soil organic carbon stocks: A meta-analysis. *Ecol. Appl.* **31**(3), e02278. <https://doi.org/10.1002/eap.2278> (2021).
58. Sperow, M. Updated potential soil carbon sequestration rates on U.S. agricultural land based on the 2019 IPCC guidelines. *Soil Till. Res.* **204**, 104–719. <https://doi.org/10.1016/j.still.2020.104719> (2020).
59. Uludere Aragon, N., Xie, Y., Bigelow, D., Lark, T. J. & Eagle, A. J. The realistic potential of soil carbon sequestration in U.S. croplands for climate mitigation. *Earth Future* **12**(6), e2023EF003866. <https://doi.org/10.1029/2023EF003866> (2024).
60. Basche, A. D., Miguez, F. E., Kaspar, T. C. & Castellano, M. J. Do cover crops increase or decrease nitrous oxide emissions? A meta-analysis. *J. Soil Water Conserv.* **69**(6), 471–482. <https://doi.org/10.2489/jswc.69.6.471> (2014).
61. Tian, H. et al. Global soil nitrous oxide emissions since the preindustrial era estimated by an ensemble of terrestrial biosphere models: Magnitude, attribution, and uncertainty. *Glob. Change Biol.* **25**(2), 640–659. <https://doi.org/10.1111/gcb.14514> (2019).
62. Don, A. et al. Carbon sequestration in soils and climate change mitigation—Definitions and pitfalls. *Glob. Change Biol.* **30**(1), e16983. <https://doi.org/10.1111/gcb.16983> (2024).
63. USDA, United States Department of Agriculture. Land resource regions and major land resource areas of the United States, the Caribbean, and the Pacific Basin. In *Agriculture Handbook* 296. U.S. Department of Agriculture. (2022).
64. Schwartzman, S. et al. Environmental integrity of emissions reductions depends on scale and systemic changes, not sector of origin. *Environ. Res. Lett.* **16**(9), 091001. <https://doi.org/10.1088/1748-9326/ac18e8> (2021).
65. UNFCCC. *Reducing Greenhouse Gases in the United States: A 2030 Emissions Target*. <https://unfccc.int/sites/default/files/NDC/2022-06/United%20States%20NDC%20April%2021%202021%20Final.pdf>. (2021).
66. USDA, United States Department of Agriculture. *Partnerships for Climate-Smart Commodities*. <https://www.usda.gov/climate-solutions/climate-smart-commodities>. (2025).
67. USDA, United States Department of Agriculture. *A General Assessment of the Role of Agriculture and Forestry in U.S. Carbon Markets*. <https://www.usda.gov/media/press-releases/2023/10/23/usda-releases-assessment-agriculture-and-forestry-carbon-markets>. (2023).
68. Haya, B. K., Abayo, A., So, I. S., & Micah, E. *Voluntary Registry Offsets Database* (Version v10). Berkeley Carbon Trading Project. <https://gspp.berkeley.edu/faculty-and-impact/centers/cepp/projects/berkeley-carbon-trading-project/offsets-database>. (2024).
69. Urban, C., & Cole, S. *Ready Or Not? Ag Carbon Markets and U.S. Farmers*. Trust In Food: A Farm Journal Initiative. <https://www.rustinfood.com/carboninsights/>. (2022).
70. Parkhurst, R., Moore, L.A., Wright, R., & Perez, M. *Agricultural Carbon Programs: From Chaos to Systems Change*. American Farmland Trust. <https://farmlandinfo.org/publications/ag-carbon-programs-chaos-to-systems-change>. (2023).
71. Brokish J., Burkhart D., Miller M., & Van Beck H. *An Overview of Voluntary Carbon Markets for Illinois Farmers*. Illinois Sustainable Ag Partnership. <https://ilsustainableag.org/ecomarkets>. (2023).
72. USDA, United States Department of Agriculture. *Federal Strategy to Advance Greenhouse Gas Emissions Measurement and Monitoring for the Agriculture and Forest Sectors*. <https://www.usda.gov/sites/default/files/documents/Draft-Federal-Ag-and-Forest-MMRV-Strategy.pdf>. (2023).
73. Smith, J. & Parkhurst, R. Opportunities for agricultural producers to participate in compliance and voluntary carbon markets. *OSF* <https://doi.org/10.31235/osf.io/yrfgz> (2018).
74. Santilli, M. et al. Tropical deforestation and the Kyoto protocol. *Clim. Change* **71**(3), 267–276. <https://doi.org/10.1007/s10584-005-8074-6> (2005).
75. Mendelsohn, R. O., Litan, R. E., & Fleming, J. *A framework to ensure that voluntary carbon markets will truly help combat climate change*. <https://www.brookings.edu/articles/a-framework-to-ensure-that-voluntary-carbon-markets-will-truly-help-combat-climate-change/>. (2021).
76. Skadell, L. E. et al. Twenty percent of agricultural management effects on organic carbon stocks occur in subsoils—Results of ten long-term experiments. *Agr. Ecosyst. Environ.* **356**, 108619. <https://doi.org/10.1016/j.agee.2023.108619> (2023).

Acknowledgements

We thank Roger L. Nelson for help with CropSyst; Isaiah L. Huber for help with APSIM, Marco Botta for help with ARMOSA, and Matt Wisniewski for help with figures.

Author contributions

BB conceived and designed the study. TT, FSM, SA, BB, LP, PS, CV, AF, MD, LD and YZ performed the study. TT, FSM, LP, CV and AF managed and analyzed the data. TT, NM, BB wrote the manuscript. GPR, KP, KRC, AR, MA, SA, LC, BD, PRG, GH, JWJ, AP and COS revised the manuscript. All authors read and revised the final manuscript.

Funding

Partial funding to Basso is provided by: Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Biological and Environmental Research Program under Award Number DE-SC0018409; the National Science Foundation Long-term Ecological Research Program (DEB 2224712) at the Kellogg Biological Station, USDA NIFA, Award no. 2020-67021-32799, Michigan State University, AgBioResearch, Climate TRACE, CERCA-FFAR project, Walton Family Foundation, United Soybean Board, Morgan Stanley Sustainable Solution Collaborative, Michigan Department of Agriculture and Rural Development. M. Delandmeter was granted a Research Fellow (number 44221) fellowship by the F.R.S.-FNRS (Belgian Fund for Scientific Research).

Declarations

Competing interests

Bruno Basso is a cofounder of CIBO Technologies. Keith Paustian and Yao Zhang has financial interest in Indigo Ag. The other coauthors declare no conflict of interest.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-08419-y>.

Correspondence and requests for materials should be addressed to B.B. or T.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025