nature sustainability

Supplementary information https://doi.org/10.1038/s41893-024-01495-4

Diversified cropping systems with limited carbon accrual but increased nitrogen supply

In the format provided by the authors and unedited

Supplementary Methods

SOC concentration

Soil organic C (SOC) was determined by subtracting carbonate C from total soil C. The $\delta^{13}C$ value of SOC was calculated by a two-source mixing model[1]. Briefly, we measured total C and N and their $\delta^{13}C$ and $\delta^{15}N$ values with an elemental analyzer (Vario Cube, Elementar, Hanau, Germany) coupled with an IsoPrime 100 mass spectrometer (Isoprime Ltd, Cheadle, UK) at the Center for Stable Isotope Biogeochemistry at the University of California, Berkeley, CA, USA. The inorganic C (carbonate C) and its $\delta^{13}C$ value were measured using a method modified from Amundson et al.[2] and described by Huang and Hall[3], in which we measured CO_2 concentration produced from soil samples after 3 M HCl treatment and its $\delta^{13}C$ value by tunable diode laser absorption spectrometer (TGA200A; Campbell Scientific, Logan, UT, USA), as described in detail by Hall et al.[4]. The two-source mixing model is as follows:

$$\delta^{13}C_{\mathrm{SOC}} = \frac{\delta^{13}C_{\mathrm{total}} \times C_{\mathrm{total}} - \delta^{13}C_{\mathrm{carbonate}} \times C_{\mathrm{carbonate}}}{C_{\mathrm{total}} - C_{\mathrm{carbonate}}}$$

Here, $\delta^{13}C_{SOC}$, $\delta^{13}C_{total}$, and $\delta^{13}C_{carbonate}$ are the $\delta^{13}C$ values of SOC, total C, and carbonate C, respectively; C_{total} and $C_{carbonate}$ are the concentrations of soil total C and carbonate C, respectively.

Model description in Agro-IBIS soil C model

In the Agro-IBIS soil C subroutine, litter C input consists of three fractions, including metabolic, structural, and lignified materials. The metabolic and structural litter C can be decomposed by microbial biomass, while the lignified fraction directly enters soil organic C pools. Litter C pools are separated into above- (leaf and wood) and below-ground (root) litter C pools. As soil cores were collected without surface litter, we only had the belowground root biomass as litter C input in the model. The root litter C was separated into metabolic, structural, and lignified root C pools. Root litter C decomposition for each pool (i) is described by first-order kinetics:

$$\frac{dC_i}{dt} = -\,k_i \times C_i$$

where k_i is the root turnover rate constant in day^{-1}, and C_i is C pool size in kg C m^{-2}. As we assumed there was no C fractionation during C transformation (see discussion below), the $\delta^{13}C$ values of root litter C pools were constant during the incubation.

Soil C pools are considered to consist of microbial biomass C, active organic C (non-protected plus protected fractions), and old inactive (passive) C components. Soil C transformations between different pools are assumed to be first-order reactions:

$$\frac{dC_i}{dt} = I(t)_i - E_i\ \times k_i \times C_i$$

where k_i is the first-order decomposition rate constant (day^{-1}), $I(t)_i$ is the input from other C pools (kg C m^{-2}), C_i is C pool size (kg C m^{-2}), and E_i is the C use efficiency factor with the value between 0 to 1. The remainder ($1-E_i$) represents C leaving the system as CO_2. The change in $\delta^{13}C$ values of each soil C pool was on the assumption of no C fractionation during C transformation:

$$\frac{d\delta^{13}C_{C_i}}{dt} = \frac{1}{\sum(I(t)_i - E_i\ \times k_i \times C_i)} \times \sum(I(t)_i \times (\delta^{13}C_{I_i} - \ \delta^{13}C_{C_i}))$$

Where $\delta^{13}C_{C_i}$ and $\delta^{13}C_{I_i}$ are $\delta^{13}C$ values of C pool and input C, respectively.

Microbial respiration (CO_2 production) is the sum of CO_2 lost during C transformations from root litter C, active organic C (non-protected plus protected fractions), and passive C pools to microbial biomass. The $\delta^{13}C$ values of CO_2 were calculated based on mass balance with the assumption of no C isotope fractionation during microbial heterotrophic respiration:

$$\delta^{13}C_{CO_2} = \frac{1}{\sum(\,(1-E_i) \times k_i \times C_i)} \times \sum((1-E_i) \times k_i \times C_i \times \delta^{13}C_{C_i})$$

Many studies indicated that isotopic discrimination might occur during heterotrophic respiration, although it remains controversial, and the role of intrinsic discrimination vs. differences in isotope composition among C substrates is difficult to resolve[5]. Here, we tested the influence of discrimination by adding a parameter ε as the discrimination factor which was independent of incubation time and decomposition rate. A positive value of ε indicates that microbes preferentially metabolize ^{12}C over ^{13}C during microbial heterotrophic respiration[5]. Conversely, a negative value indicates that microbes selectively utilize litter carbohydrates, which have a relatively more positive $\delta^{13}C$ value than bulk litter C, during C decomposition[6]. The changes in $\delta^{13}C$ values of microbial biomass and respired CO_2 are as follows (see Supplementary Table 8 for more details):

$$\delta^{13}C_{mic} = \frac{1}{\sum(I(t)_i - E_i \times k_i \times C_i)} \times \sum(E_i \times k_i \times C_i \times ((\delta^{13}C_{I_i} + \mathcal{E} - \delta^{13}C_{C_i})$$

$$\delta^{13}C_{CO_2} = \frac{1}{\sum((1 - E_i) \times k_i \times C_i)} \times \sum((1 - E_i) \times k_i \times C_i \times (\delta^{13}C_{C_i} - E_i \times (1 - E_i) \times \mathcal{E}))$$

We applied the optimized three influential paraments in the Agro-IBIS model for each plot (described below and in the Methods) and calibrated the discrimination factor (ε) across all plots, assuming the value did not vary among crop rotation systems or crop production phases. After data assimilation, the calibrated ε value turned out to be negligible (-0.3‰) and in the opposite direction as assumed by studies that postulated the importance of kinetic isotope fractionation during heterotrophic respiration[5]. Thus, we did not account for C isotope fractionation in any model in this study.

Sobol sensitivity analysis

A parameter sensitivity analysis was conducted to assess the importance of model parameters, which is crucial step for implementing formal Bayesian techniques to uncertainty analysis. Global sensitivity analysis apportions the uncertainty in model outcomes to individual inputs and their interactions[7]. Among various methods, the Sobol method is often considered one of the most comprehensive global sensitivity analysis due to its sampling strategy, which effectively explores the parameter space[8].

The decomposition of the variance of the output is as follows:

$$\mathrm{V} = \sum_{i=1}^{n} V_i + \sum_{1 \le i \le j \le n} V_{ij} + \cdots + V_{1,2\ldots n}$$

where n denotes the total number of variables, and Vi represents the portion of the output variance explained solely by the i^{th} variable, and the terms of $V_{ij} + \cdots + V_{1,2\ldots n}$ correspond to the portion of the output variance attributed to interactions among variables. The computational details can be found in Sobol et al.[9].

The Sobol sensitivity indices are defined as follows:

$$S_{i1\ldots is} = \frac{V_{i1\ldots is}}{\mathrm{V}}$$

where $V_{i1\ldots is}$ denotes the variance associated with $(i_1 \ldots i_s)$.

The first-order Sobol sensitivity indices (main effects) for each variable are:

$$S_i = \frac{V_i}{V}$$

Total order indices of Sobol sensitivity (total effects):

$$S_{Ti} = S_i + \sum_{1 \le i \le j \le n} S_{ij} + \cdots + S_{1,2\ldots n}$$

In this study, we conducted a Sobol sensitivity analysis using 26 parameters in the Agro-IBIS and 27 parameters in the CN-SIM model. In Sobol sensitive analysis simulation, the prior parameter distribution was assumed to be a uniform, and the parameter ranges informed by relevant literature as detailed in Supplementary Tables 9 and 11.

Data assimilation in Agro-IBIS soil C model

The three important parameters identified in Sobol analysis were estimated by data assimilation using Bayesian probabilistic inversion and a Markov chain Monte Carlo (MCMC) technique[10]. The Bayesian probabilistic inversion was conducted using the Metropolis-Hastings (M-H) algorithm to generate the posterior probability density function (PDF) of parameters. Bayesian probabilistic inversion approach based on Bayes' theorem states that the posterior probability density function (PDF, $P(\theta|Z)$) of model parameters can be obtained from the prior probability density function ($P(\theta)$), an assumption range of parameters based on empirical knowledge, and the likelihood function ($P(Z|\theta)$), a conditional probability density of parameter under observation: $P(\theta|Z) \propto P(Z|\theta)P(\theta)$. The prior probability density function ($P(\theta)$) was assumed as a uniform distribution as it is the best prior to choose without any other knowledge regarding parameter distribution[10]. The likelihood function ($P(Z|\theta)$) was calculated with the assumption that errors are independent and follow Gaussian distributions: $P(Z|\theta) \propto \exp\left\{-\sum_{i=1}^{k} \frac{[z_i(t)-x_i(t)]}{2\sigma_i^2(t)}\right\}$, where k is the number of the observations, z_i and x_i are modeled and observed values, respectively, and σ_i is the standard deviation of the observations.

The Metropolis-Hastings (M-H) algorithm, an MCMC technique[11,12], was used to generate the posterior probability density function (PDF) of parameters by repeating two steps: a proposing step and a moving step. In the proposing step, a new proposed parameter value (θ^{new}) is generated based on the parameter value accepted in the previous step (θ^{old}) with a proposal distribution: $\theta^{new} = \theta^{old} + \frac{d \times (\theta_{max} - \theta_{min})}{D}$, where D is a coefficient to control the proposing step size, θ_{max} and θ_{min} are the upper and lower limits of parameter, respectively, and d is a uniformly distributed random value between -0.5 and 0.5 to help find a randomly new parameter in the moving step. In each moving step, the new point is tested against the Metropolis criterion to determine if it should be accepted or rejected[10].

Data assimilation in CN-SIM soil C model

Markov-Chain Monte-Carlo (MCMC) techniques have been applied to quantify the uncertainty in crop and ecological models[13,14]. Using a Bayesian formalism, the posterior distribution of the model parameters ($P(\boldsymbol{\theta}|Y)$) can be obtained by conditioning the spatial-temporal behavior of the model on measurements (Y) as reflected in the modeled response ($\hat{Y}$).

$$P(\theta|Y) = \frac{P(\theta|Y) * P(\theta)}{P(Y)}$$

where $P(\theta|Y)$ represents the likelihood of the data, and $P(\theta)$ and $P(\theta|Y)$ denote the prior and posterior parameter distributions, respectively. In practice, $P(Y)$ is not required for estimating the posterior, as $P(Y)$ serves as a normalization constant (scalar) that ensures the posterior distribution integrates to unity over the parameter space. If the prior distribution, $P(\theta)$, is well defined, then the main challenge often lies in specifying the likelihood function. The likelihood $P(\theta|Y)$ is determined from the probability distribution of the residuals between observed (Y) and modeled responses ($\hat{Y}$). These residuals are generally assumed to be uncorrelated, independent, and normally distributed[15], and then the likelihood of the n-vector of error residuals can be expressed as follows:

$$L(\theta|Y) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\Pi\sigma_i^2}} \exp\left[-\frac{1}{2}\left(\frac{Y_i - \hat{Y}_i(\theta)}{\sigma_i}\right)^2\right]$$

where $\sigma = \{\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_n\}$ is an n-vector with standard deviations of the measurement error in the observations. Assuming homoscedastic model residuals, this formulation can be converted into a log-likelihood expression:

$$L(\theta|Y) = -\frac{1}{2}\log(2\Pi) - \sum_{i=1}^{n}(\log(\sigma_i) - \frac{1}{2}\sum_{i=1}^{n}(\frac{Y_i - \hat{Y}_i(\theta)}{\sigma_i})^2$$

In this study, we used the DREAM algorithm[16] to estimate the posterior parameter distribution of three important parameters (i.e., kre, kpoc and kmoc) for the CN-SIM[17] soil C decomposition model.

MEND model parameterization and uncertainty quantification

We implemented multi-objective calibration of the MEND model[18]. Each objective evaluated the goodness-of-fit of a specific observed variable (e.g., CO_2 flux and its $\delta^{13}C$ value). The parameter optimization was to minimize the total objective function (J) that was computed as the weighted average of multiple single objectives[19]

$$J = \sum_{i=1}^{m} w_i \cdot J_i \qquad \text{(S40a)}$$

$$\sum_{i=1}^{m} w_i = 1 \text{ with } w_i \in [0,1] \qquad \text{(S40b)}$$

where m denotes the number of objectives, and w_i is the weighting factor for the i^{th} ($i = 1,2,\ldots,m$) objective (J_i).

In this study, the individual objective function J_i (i.e., for CO_2 flux and its $\delta^{13}C$ value) was calculated as ($1 - R^2$), where R^2 denoted the coefficient of determination[20]:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}[Y_{sim}(i) - Y_{obs}(i)]^2}{\sum_{i=1}^{n}[Y_{obs}(i) - \bar{Y}_{obs}]^2} \qquad \text{(S41)}$$

where n is the number of data; Y_{obs} and Y_{sim} are observed and simulated values, respectively; and $\bar{Y}_{obs}$ is the mean value for Y_{obs}.

We used the SCE algorithm[20,21] to determine model parameters. SCE was a stochastic optimization method that includes the competitive evolution of a 'complex' of points spanning the parameter space and the shuffling of complexes[21].

Initial values for C pools and their C isotope values in three soil C models

We estimated the initial plant root C inputs mainly by using previous studies at the Marsden Farm[22, 23]. For initial root biomass C input, we considered the field-averaged root biomass as the baseline: 0.17 kg C m^{-2} for corn and 0.06 kg C m^{-2} for soybean in the 2-year rotation, 0.18 kg C m^{-2} for corn, 0.07 kg C m^{-2} for soybean, and 0.11 for oat/red clover in the 3-year rotation, 0.18 kg C m^{-2} for corn, 0.07 kg C m^{-2} for soybean, 0.11 kg C m^{-2} for oat/alfalfa, and 0.20 kg C m^{-2} for the alfalfa in the 4-year rotation. However, root biomass varies systematically between row and inter-row locations. We also explored the scenarios of increased and decreased root C inputs by applying in-row root biomass values (0.27 kg C m^{-2} for corn and 0.10 kg C m^{-2} for soybean in the 2-year rotation, 0.27 kg C m^{-2} for corn, 0.11 kg C m^{-2} for soybean, and 0.17 for oat/red clover in the 3-year rotation, 0.28 kg C m^{-2} for corn, 0.12 kg C m^{-2} for soybean, 0.17 kg C m^{-2} for oat/alfalfa, and 0.30 kg C m^{-2} for the alfalfa in the 4-year rotation) and between-row root biomass values (0.14 kg C m^{-2} for corn and 0.04 kg C m^{-2} for soybean in the 2-year rotation, 0.14 kg C m^{-2} for corn, 0.04 kg C m^{-2} for soybean, and 0.09 for oat/red clover in the 3-year rotation, 0.15 kg C m^{-2} for corn, 0.04 kg C m^{-2} for soybean, 0.09 kg C m^{-2} for oat/alfalfa, and 0.16 kg C m^{-2} for the alfalfa in the 4-year rotation), respectively. Specifically, we first estimated the in-row root biomass using the average annual dry matter yield and its proportion allocated to belowground C measured in the 2-year and 4-year rotations at the same experimental site[22]. The in-row root biomass in the 3-year rotation was calculated by the mean values of the corresponding phases of the 2-year and 4-year rotations, except for the phase of oat/alfalfa, which was equal to the value from the same phase in the 4-year rotation. The in-row values were then used to estimate the field-averaged and between-row root biomass according to the measurements by Nichols et al.[23] and the equation developed by Ordóñez et al.[24]. The root C/N ratio was estimated by the values in the 0–10 cm depth measured by Lazicki et al.[25] at the same site (21, 19, and 19 on average for the 2-, 3- and 4-year rotations, respectively).

We initialized soil C pools in the three models by multiplying the total SOC measured in each plot and the corresponding estimated fractions of the C pools to the total SOC. Microbial biomass C accounted for 0.78%, 0.84%, and 0.89% of the initial total SOC for the 2-year, 3-year, and 4-year rotations, respectively. The fraction of non-protected active organic C to total SOC in the Agro-IBIS, functionally similar to the particulate organic C (POC) in the CN-SIM and MEND models, was estimated by the measurements from Poffenbarger et al.[22], with 3.6%, 3.7% and 3.9% for the 2-, 3- and 4-year rotations, respectively. We further investigated scenarios involving the increase or decrease of the fraction of non-protected active organic C or POC to SOC, which was achieved by utilizing the maximum (4.5%, 5.0%, and 5.5% for the 2-, 3- and 4-year rotations, respectively) and minimum (2.5%, 2.5% and 2.5% for the 2-, 3- and 4-year rotations, respectively) values obtained from measurements[22]. The fraction of passive C to total SOC in the Agro-IBIS model was assumed to be 0.55[26]. In the MEND model, the dissolved organic C (DOC) was estimated as 0.4% of SOC[6], the active part of MOC (QOC pool) was estimated as 1% of the MOC pool[20], and the initial enzyme concentrations were assumed to approximately 0.1% of microbial biomass[20].

The initial $\delta^{13}C$ values of root C inputs varied among different crops, whereas $\delta^{13}C$ values of soil C pools varied with plots. The initial $\delta^{13}C$ values of plant roots were assumed to be -12‰ for the C_4 plant (i.e., corn) and -26‰ for C_3 plants (i.e., soybean, oat, and alfalfa). For simplicity, we did not assume there was a difference in the $\delta^{13}C$ values between metabolic, structural, and lignified root C pools. We used the $\delta^{13}C$ values of SOC stock in 0–30 cm measured in this study as initial $\delta^{13}C$ values of SOC for each plot as the topsoil layers dominate CO_2 production. The $\delta^{13}C$ values of CO_2 measured on the first day of incubation were used as initial $\delta^{13}C$ values of microbial biomass C, and the values measured on the last day of incubation were regarded as the initial $\delta^{13}C$ values of non-protected active C in the Agro-IBIS and POC in the CN-SIM and MEND models. We used the measured $\delta^{13}C$ value of SOC from a nearby prairie (-17.0‰) as the $\delta^{13}C$ value of passive C in the Agro-IBIS model, as the passive C pool most likely shows a C isotope signature in soils without any modern agricultural practices. We then calculated the initial $\delta^{13}C$ values of protected active organic C or MAOC by mass balance. In the MEND model, the initial $\delta^{13}C$ values of DOC and enzyme were assumed to be equivalent to those of microbial biomass C, and the initial $\delta^{13}C$ values of QOC were set to match those of MOC.

Supplementary Results

Comparison between isotope-calibrated and no-isotope-calibrated models

To test the importance of incorporating isotopic features ($\delta^{13}C$ of C pools and fluxes) in soil C models, we compared two parameterization strategies in the Agro-IBIS, CN-SIM, and MEND models, whereby we calibrated the models individually with data from each experimental plot for CO_2 flux only (no-isotope-calibrated model) or with CO_2 flux and its $\delta^{13}C$ value (isotope-calibrated model). Overall, the isotope-calibrated models well-simulated CO_2 from all 36 field plots (mean determination R^2 = 0.92, 0.83, 0.91 for Agro-IBIS, CN-SIM, and MEND, respectively; Supplementary Table 3), with slightly lower performance than the no-isotope-calibrated models (mean determination R^2 = 0.93, 0.84, 0.94 for Agro-IBIS, CN-SIM, and MEND, respectively). The performance of isotope-calibrated models in simulating the $\delta^{13}C$ value of CO_2, however, was significantly improved over the no-isotope-calibrated models (mean RMSE for no-isotope-calibrated vs. C-isotope-calibrated model: 1.40‰ vs. 1.33‰ for Agro-IBIS at $P < 0.05$, 1.51‰ vs. 1.49‰ for CN-SIM at $P < 0.05$, and 1.40‰ vs. 1.30‰ for MEND at $P < 0.05$, as detailed in Supplementary Table 3). These results indicate the isotope-calibrated model improved tracking C losses from the correct ratio from C_3 and C_4 sources. Consequently, we focused on the isotope-calibrated models.

Model performance of isotope-calibrated models with varying initial input values

As some of the values used to initialize the models (i.e., root C mass and soil C content among different pools) were derived from empirical measurements in previous field studies, we further assessed the influence of varying initial values on model parameter estimation by implementing models with best-estimate and plausible low/high initial values[27]. We applied several scenarios by varying initial values of root biomass C inputs and particulate organic C pool size (POC) to test the impact of uncertainties in initial values on model parameterization. We found that the isotope-calibrated model performance was more sensitive to changes in initial values in the Agro-IBIS model than the MEND model. In the Agro-IBIS and CN-SIM models, varying initial values did not lead to better model performance in CO_2 production simulation; instead, increasing root biomass C inputs in the isotope-calibrated model resulted in its even poorer model performance (R^2= 0.91 for Agro-IBIS and R^2= 0.90 for MEND on average, Supplementary Table 3). The isotope-calibrated models with decreasing root biomass C inputs consistently performed better in simulating the $\delta^{13}C$ value of CO_2 than the models with the best-estimated values ($P < 0.01$ for the MEND model, and $P < 0.05$ for the Agro-IBIS and CN-SIM models, Supplementary Table 3), while the other scenarios did not result in consistently better model performance across all three models. Thus, for clarity of explanation, we presented the results from the isotope-calibrated model simulations with best-estimated (or field-averaged) and lower-bound (or between-row) root biomass C inputs. After removing the plots with unsatisfactory parameter estimates (the plots with determination $R^2 < 0.6$ for CO_2 and RMSE > 2‰ for $\delta^{13}C$ values of CO_2), we synthesized the results from the three models by calculating their response ratios of the key parameters and the respective CO_2 production or C flux (i.e., in the MEND model) from soil C pools to test the differences between the longer rotation (3- and 4-year rotations) vs. the 2-year rotation, and the results from the comparison of only corn plots among different rotations were further detailed in the subsequent sections.

Response of model parameters and CO_2 fluxes with low-bound root C initial values

Model parameters and C fluxes calibrated by low-bound root C input had similar differences among cropping systems as the best-estimate model initial values (Supplementary Fig. 6). In the

Agro-IBIS model, the longer rotations significantly (*P* < 0.05) increased the decomposition of non-protected organic C (knb). In the CN-SIM model, the longer rotations significantly (*P* < 0.05) increased the decomposition of particulate organic C (kpoc). In the MEND model, the longer rotations significantly decreased intrinsic microbial C use efficiency (Yg; *P* < 0.05), and the 4-year rotations significantly (*P* < 0.05) decreased the parameter for enzyme production to decompose MAOC (fpEM).

Corresponding to the parameter changes, the Agro-IBIS and CN-SIM model-estimated cumulative CO_2 production from the C pools with the residence time of months to years were significantly (*P* < 0.05) higher in longer rotations than in the 2-year rotation (Supplementary Fig. 8). In the MEND model, higher cumulative CO_2 production in the longer rotations was mainly from microbial C pool with months to years residence time (*P* < 0.08 for the 3-year rotation and P=0.06 for 4-year rotations, Supplementary Fig. 8).

Model parameters during the corn production phase in different rotations

During the corn production phase, we found that the longer rotations tended to stimulate the decomposition of the C pools with residence time of months–years (Supplementary Fig. 9) under different model input scenarios (best-estimated and lower-bound root C inputs), which was relatively consistent with our analysis that included all phases of each rotation. Regarding the key parameters that controlled C decomposition in the Agro-IBIS model, longer rotations (3- and 4-year) increased knb (decomposition rate of non-protected organic C), relative to the 2-year rotation. In the CN-SIM model, the 4-year rotation increased kpoc (decomposition rate of particulate organic C) in the C isotope-calibrated model. In the MEND model, the 3- and 4-year rotations tended to increase Vm (maximum specific uptake rate of *D* for maintenance) under both best-estimated and lower-bound root C inputs (Supplementary Fig. 7). However, the 4-year rotation also tended to decrease fpEM (related to enzyme production of MOC) relative to the 2-year rotation.

C fluxes from C pools during the corn production phase in different rotations

During the corn production phase across different rotations, the longer rotations notably affected the cumulative CO_2 production from certain C pools (Supplementary Fig. 9). Compared to the 2-year rotation in the Agro-IBIS model, longer rotations resulted in a higher cumulative CO_2 production from the root C pool and the non-protected active C pool under different model input scenarios (best-estimated and lower-bound root C inputs, Supplementary Fig. 9). In the CN-SIM model, during the corn production phase, the 4-year rotations had increased CO_2 production from particulate organic C pool**.** In the MEND model, the 3-year rotations increased CO_2 production from particulate organic C1 and mineral-associated C pools, and the 4-year rotations increased CO_2 production from microbial C pool with months-years resident time (Supplementary Fig. 9).

Partial N balance in different rotations

We estimated the partial nitrogen (N) balance for each rotational system (Supplementary Table 4). Following a common approach in the literature (e.g., Blesh and Drinkwater 2013), we considered N inputs from synthetic N fertilizer, manure N, and N fixation from legumes, and N outputs in harvested crops. The net partial N balance was calculated as the difference between the summed N inputs and N outputs. Certain N inputs and outputs were estimated using field-measured crop yield and biomass data, along with regression models and coefficients from literature. Specifically, synthetic fertilizer and manure N inputs were applied only before or during the corn planting phase, and data sourced from historical management records. The average application rate in 2005-2020 were used as the N input for corn in our calculation. Biological N fixation (BNF) from legumes was estimated as follows:1) N fixation by soybean was calculated using a regression model based on soybean yield data reported in Córdova et al.[28]; 2) N fixation by red clover and alfalfa was estimated using biomass or yield measurement, plant N content, and the proportion of N fixation reported by Blesh & Drinkwater[29]. N removal in harvested crops was estimated as follows: 1) N removal in corn and soybean grain was calculated by multiplying historical grain yield data (average in 2005-2020) by grain N content reported in the literatures[30,31]; 2) N removal in oat grain/straw and alfalfa hay was calculated in a similar manner. However, we note that many of these estimated values were highly uncertain, particularly the legume N fixation, and so we provide three different scenarios. Details about construction of the N balance and the individual terms are provided in Supplementary Table 4.

Overall, the estimated partial N balance in the 2-year rotation system ranged from –13 to 15 kg N ha^{-1} yr^{-1}, and it ranged from –30 to –1 and –39 to 22 kg N ha^{-1} yr^{-1} in the 3-year and 4-year systems, respectively (Supplementary Table 4). The key take-home point of the partial N balance is that all rotations were approximately in balance with respect to N inputs and outputs in agricultural products when optimistic rates of N fixation were considered, acknowledging high uncertainty in many of the individual terms. Importantly, the inputs of N from manure and legume fixation appeared sufficient to replace the decreased synthetic fertilizer inputs.

Supplementary Tables 1 to 14

Supplementary Table 1 Carbon pool settings and microbial process representation across the three tested models.

	Agro-IBIS[26]	CN-SIM[17]	MEND[18]
Litter C pools	Metabolic C; Structure C; Lignified C	Easily decomposable C pool; Poorly decomposable C pool	Root biomass C
Soil C pools	Protected active organic C (POC); Non-protected active organic C (NOC); Passive C	Particulate organic C (POC); Mineral-associated organic C (MAOC)	Particulate organic C1 (POC1); Particulate organic C2 (POC2); Mineral-associated organic C (MOC); Adsorbed phase of DOC/active part of MOC (QOC); Dissolved organic C
Microbial C pools	Unprotected biomass C; Protected microbial biomass C	Microbial biomass (K-strategists); Microbial biomass (r-strategists)	Active microbial biomass (MBA); Dormant microbial biomass (MBD)
Microbial processes representation	A first-order kinetic; CO_2 fluxes are represented as losses during C transfers between litter/soil pools and microbial pools.	Michaelis–Menten kinetics; Decomposed C from litter and soil is partly released as CO_2 after microbial assimilation by two functional microbial groups (r- and K-strategists) with distinct C use strategies, while the rest is allocated to microbial biomass or other SOC pools.	Michaelis–Menten kinetics and ECA (Equilibrium Chemistry Approximation); soil C fluxes are regulated by distinct microbial and enzyme functional groups.

Model references for Agro-IBIS soil C model: Verberne et al.[26]; CN-SIM: Yi et al.[17]; MEND: Wang et al.[18].

Supplementary Table 2 Different sources of carbon (C) input in different cropping systems at the Marsden Farm field site.

System	Crop	Rotation (length in years)	Plant C input (kg m^{-2}) ǂ1	Manure C (kg m^{-2}) ǂ2	Manure $\delta^{13}C$ (‰)	Manure C_4-C (kg m^{-2})	Manure C_3-C (kg m^{-2})	Total C_4-4 (kg m^{-2})	Total C_3-C (kg m^{-2})
2-year	Corn	2	0.72	0				0.72	
	Soybean	2	0.29	0					0.29
3-year	Corn	3	0.82	0.31	-21.71	0.09	0.21	0.92	0.21
	Soybean	3	0.31	0					0.31
	Oat/red clover	3	0.22	0					0.22
4-year	Corn	4	0.93	0.31	-21.71	0.09	0.21	1.02	0.21
	Soybean	4	0.34	0					0.34
	Oat/alfalfa	4	0.22	0					0.22
	Alfalfa	4	0.40	0					0.40

ǂ1 Data cited from Poffenbarger et al. (2020)[22].
ǂ2 Based on manure application in 2020.

Supplementary Table 3 Model performance for CO_2 production and its C isotope ratio ($\delta^{13}C$) with different model strategies and scenarios (different input data). The two strategies involve model calibration with CO_2 flux and its $\delta^{13}C$ values (C isotope-calibrated) and model calibration with CO_2 flux only (no-isotope). The different scenarios include increasing and decreasing initial root C inputs or particulate organic carbon (POC) in the C isotope-calibrated model. * $P < 0.05$, ** $P < 0.01$. Values are represented in the format of determination R^2 (standard error) for CO_2 and RMSE (standard error) for $\delta^{13}C$ of CO_2, and the text "nan" indicates simulations where < 2 replicates achieved satisfactory model performance.

	CO_2						$\delta^{13}C$ of CO_2					
	Isotope enabled ǂ[1]	No-isotope ǂ[2]	Decreased root ǂ[3]	Increased root	Decreased POC	Increased POC	Isotope enabled	No-isotope	Decreased root	Increased root	Decreased POC	Increased POC
Agro-IBIS model												
C2	0.86 (0.03)	0.90 (0.03)	0.89 (0.02)	0.86 (0.01)	0.90 (0.02)	0.87 (0.03)	1.32 (0.02)	1.59 (0.20)	1.37 (0.08)	1.28 (0.04)	1.37 (0.05)	1.34 (0.02)
S2	0.97 (nan)	0.97 (nan)	0.97 (nan)	0.96 (nan)	0.89 (0.08)	0.97 (nan)	1.00 (nan)	0.96 (nan)	0.95 (nan)	1.11 (nan)	1.46 (0.40)	0.94 (nan)
C3	0.90 (0.03)	0.90 (0.02)	0.91 (0.03)	0.88 (0.06)	0.92 (0.03)	0.88 (0.01)	1.31 (0.17)	1.34 (0.18)	1.31 (0.14)	1.22 (0.01)	1.37 (0.23)	1.32 (0.21)
S3	0.91 (0.03)	0.93 (0.02)	0.94 (0.01)	0.93 (0.02)	0.93 (0.01)	0.91 (0.04)	1.46 (0.11)	1.50 (0.06)	1.38 (0.12)	1.58 (0.08)	1.45 (0.14)	1.49 (0.07)
O3	0.92 (0.02)	0.93 (0.03)	0.94 (0.02)	0.88 (0.03)	0.93 (0.01)	0.91 (0.03)	1.34 (0.12)	1.40 (0.13)	1.36 (0.11)	1.36 (0.15)	1.33 (0.12)	1.20 (0.06)
C4	0.96 (0.01)	0.95 (0.00)	0.97 (0.00)	0.94 (nan)	0.93 (0.01)	0.93 (0.03)	1.39 (0.11)	1.49 (0.16)	1.31 (0.11)	1.59 (nan)	1.54 (0.11)	1.47 (0.06)
S4	0.94 (0.01)	0.94 (0.02)	0.94 (0.01)	0.92 (0.02)	0.93 (0.01)	0.92 (nan)	1.37 (0.27)	1.43 (0.20)	1.34 (0.18)	1.40 (0.19)	1.50 (0.16)	1.15 (nan)
O4	0.89 (0.01)	0.90 (0.04)	0.91 (0.02)	0.88 (0.01)	0.93 (0.02)	0.85 (0.02)	1.65 (0.32)	1.58 (0.16)	1.48 (0.13)	1.54 (0.09)	1.33 (0.06)	1.47 (0.11)
A4	0.93 (0.03)	0.94 (0.02)	0.94 (0.03)	0.91 (0.03)	0.94 (0.02)	0.92 (0.03)	1.15 (0.11)	1.25 (0.11)	1.18 (0.09)	1.13 (0.14)	1.21 (0.06)	1.13 (0.14)
Mean	0.92 (0.01)	0.93 (0.01)	0.93 (0.01)	0.91** (0.01)	0.92 (0.01)	0.91** (0.01)	1.33* (0.06)	1.40 (0.07)	1.30** (0.05)	1.36 (0.06)	1.40 (0.03)	1.28* (0.06)
CN-SIM model												
C2	0.88 (0.01)	0.88 (0.02)	0.86 (0.01)	0.92 (0.01)	0.90 (0.02)	0.85 (0.02)	1.57 (0.17)	1.58 (0.17)	1.51 (0.14)	1.43 (0.12)	1.57 (0.18)	1.58 (0.17)
S2	0.68 (nan)	0.67 (nan)	nan (nan)	0.84 (nan)	nan (nan)	0.76 (nan)	1.11 (nan)	1.11 (nan)	nan (nan)	1.37 (nan)	nan (nan)	1.69 (0.58)
C3	0.83 (0.08)	0.82 (0.08)	0.81 (0.06)	0.87 (0.04)	0.85 (0.00)	0.80 (0.13)	1.38 (0.07)	1.36 (0.06)	1.48 (0.21)	1.76 (0.15)	1.39 (0.12)	1.64 (0.28)
S3	0.75 (0.11)	0.79 (0.08)	0.84 (nan)	0.84 (0.05)	0.75 (nan)	0.73 (0.03)	1.67 (0.04)	1.72 (0.09)	1.55 (nan)	1.92 (0.06)	1.73 (nan)	1.72 (0.09)
O3	0.89 (0.02)	0.91 (0.01)	0.87 (0.02)	0.92 (0.02)	0.84 (0.04)	0.86 (0.02)	1.44 (0.18)	1.45 (0.18)	1.39 (0.16)	1.45 (0.17)	1.46 (0.15)	1.45 (0.18)
C4	0.93 (nan)	0.92 (nan)	0.92 (nan)	nan (nan)	0.92 (nan)	0.90 (nan)	1.77 (nan)	1.87 (nan)	1.55 (nan)	nan (nan)	1.93 (nan)	2.13 (0.26)
S4	0.79 (0.01)	0.82 (0.01)	0.71 (0.01)	0.90 (0.00)	0.65 (0.01)	0.89 (0.03)	1.46 (0.22)	1.47 (0.20)	1.47 (0.21)	1.46 (0.00)	1.45 (0.14)	1.67 (0.24)
O4	0.87 (0.04)	0.88 (0.03)	0.85 (0.04)	0.92 (0.02)	0.87 (0.02)	0.83 (0.02)	1.67 (0.09)	1.68 (0.09)	1.59 (0.09)	1.83 (0.12)	1.60 (0.11)	1.68 (0.09)
A4	0.87 (0.03)	0.88 (0.02)	0.85 (0.04)	0.89 (0.03)	0.85 (0.05)	0.86 (0.01)	1.33 (0.22)	1.33 (0.20)	1.33 (0.19)	1.31 (0.16)	1.36 (0.12)	1.33 (0.20)
Mean	0.83 (0.03)	0.84 (0.03)	0.84 (0.03)	0.89 (0.03)	0.83 (0.02)	0.83 (0.02)	1.49* (0.07)	1.51 (0.08)	1.48* (0.10)	1.57 (0.08)	1.56 (0.09)	1.65 (0.07)
MEND model												
C2	0.91 (0.04)	0.92 (0.04)	0.91 (0.04)	0.91 (0.05)	0.95 (0.00)	0.93 (0.04)	1.33 (0.09)	1.45 (0.06)	1.31 (0.09)	1.38 (0.10)	1.35 (0.11)	1.34 (0.11)
S2	0.94 (0.03)	0.94 (0.03)	0.94 (0.03)	0.93 (0.03)	0.93 (0.03)	0.94 (0.03)	1.13 (0.24)	1.14 (0.24)	1.13 (0.23)	1.16 (0.25)	1.13 (0.24)	1.14 (0.24)
C3	0.93 (0.01)	0.93 (0.01)	0.94 (0.00)	0.89 (0.04)	0.91 (0.01)	0.91 (0.05)	1.32 (0.06)	1.54 (0.03)	1.16 (0.06)	1.34 (0.08)	1.34 (0.08)	1.33 (0.06)
S3	0.90 (0.00)	0.92 (0.02)	0.89 (0.01)	0.90 (0.00)	0.89 (0.01)	0.92 (0.01)	1.27 (0.13)	1.30 (0.10)	1.28 (0.13)	1.25 (0.13)	1.26 (0.13)	1.28 (0.13)
O3	0.89 (0.02)	0.94 (0.00)	0.89 (0.02)	0.90 (0.01)	0.89 (0.02)	0.90 (0.02)	1.22 (0.11)	1.36 (0.11)	1.24 (0.12)	1.19 (0.10)	1.22 (0.12)	1.23 (0.11)
C4	0.89 (0.04)	0.97 (nan)	0.90 (0.05)	0.86 (nan)	0.82 (nan)	0.89 (0.06)	1.84 (0.14)	1.87 (nan)	1.70 (0.11)	1.96 (nan)	1.79 (nan)	1.70 (0.09)
S4	0.93 (0.01)	0.95 (0.01)	0.94 (0.01)	0.92 (0.01)	0.92 (0.01)	0.95 (0.01)	1.18 (0.15)	1.26 (0.13)	1.20 (0.14)	1.17 (0.16)	1.19 (0.14)	1.18 (0.15)
O4	0.89 (0.01)	0.93 (0.02)	0.90 (0.01)	0.89 (0.01)	0.88 (0.01)	0.91 (0.01)	1.15 (0.04)	1.26 (0.06)	1.18 (0.04)	1.14 (0.06)	1.14 (0.06)	1.17 (0.03)
A4	0.90 (0.02)	0.96 (0.01)	0.89 (0.02)	0.91 (0.01)	0.89 (0.02)	0.90 (0.02)	1.23 (0.12)	1.47 (0.03)	1.26 (0.12)	1.16 (0.13)	1.23 (0.12)	1.23 (0.12)
Mean	0.91** (0.01)	0.94 (0.01)	0.91* (0.01)	0.90** (0.01)	0.90* (0.01)	0.91* (0.01)	1.30** (0.07)	1.40 (0.07)	1.27** (0.06)	1.31* (0.09)	1.30** (0.07)	1.29** (0.06)

ǂ[1] Isotope-calibrated model represents the simulation with the best-estimate initial C pools and using CO_2 production + C isotope ratio ($\delta^{13}C$) in model calibration.
ǂ[2] No-isotope model represents the simulation with the best-estimate initial C pools and using only CO_2 production in model calibration.
ǂ[3] Decreased/increased root and decreased/increased POC models represent the simulation with plausible lower/higher-bound values for initial root and POC pools and using CO_2 production + C isotope ratio ($\delta^{13}C$) in model calibration.

Supplementary Table 4 Partial N balance in different cropping systems from 2005 to 2020 at the Marsden Farm field site. Total output values for oat/red clover and oat/alfalfa represent the combined values from both crops.

	N application ($kg N ha^{-1}$)		N fixation ($kg N ha^{-1}$) [#3]			Total N input ($kg N ha^{-1}$)			Total N Output ($kg N ha^{-1}$) [#4]	Partial N balance ($kg N ha^{-1} yr^{-1}$)		
Treatment	Synthetic N Fertilizer [#1]	Manure N [#2]	low	mid	high	low	mid	high	Harvest removal	low	mid	high
Corn	171	0	0	0	0	171	171	171	139	-	-	-
Soybean	0	0	123	159	179	123	159	179	181	-	-	-
2-year system	171	0	123	159	179	294	330	350	320	**-13**	**5**	**15**
Corn	60	105	0	0	0	165	165	165	145	-	-	-
Soybean	0	0	130	165	189	130	165	189	205	-	-	-
Oat/red clover	0	0	141	153	169	141	153	169	177	-	-	-
3-year system	60	105	271	318	358	436	483	523	527	**-30**	**-15**	**-1**
Corn	52	105	0	0	0	157	157	157	146	-	-	-
Soybean	0	0	133	169	194	133	169	194	217	-	-	-
Oat/alfalfa	0	0	81	107	124	81	107	124	122	-	-	-
Alfalfa	0	0	221	303	358	221	303	358	261	-	-	-
4-year system	52	105	435	579	676	592	736	833	746	**-39**	**-3**	**22**

[#1]: Synthetic N fertilization from 2005-2020.

[#2]: Manure N application from 2005-2020.

[#3]: N fixation from soybean derived from Córdova et al.[28] and assuming that 24% of total plant N is in the roots (assumed by Ciampitti and Salvagiotti[30], cited from Rochester et al.[32]). N fixation from alfalfa was derived from Blesh and Drinkwater[29], considering three scenarios: low (45%), medium (70%), and high (87%) percent N derived from the atmosphere, with 60% allocated to roots. N fixation from red clover also followed Blesh and Drinkwater[29], with three scenarios: low (70%), medium (76%), and high (84%) percent N derived from the atmosphere. The biomass and N content (3.38%) of residual alfalfa and red clover biomass (4.74 $Mg ha^{-1}$) measured at the end of the season (data provided by M. Liebman) was summed with the total harvested yield of alfalfa to estimate the N fixation rate, in conjunction with the values provided above.

[#4]: Total output represents the N harvested by crops. N harvested from corn was estimated using the mean N content of corn grain from Tenorio et al.[31] and historical field yield records (2005-2020). N harvested from soybean was calculated using the mean N content of soybean grain from Ciampitti and Salvagiotti[30] and historical field yield records (2005-2020). N content in oats was estimated using the mean N content of oat grain from FAO 2011 Food Balance Sheets, and the mean N content of oat straw from Hergoualc'h et al.[33], along with historical grain and straw biomass data. N harvested from alfalfa was based on historical harvested alfalfa biomass and N content measurements.

Supplementary Table 5 Synthetic nitrogen fertilizer application timing, average rates, and soil organic carbon (values in 2021 from 0-1 m soil depth) to nitrogen fertilizer ratio in different cropping systems from 2005 to 2021 at the Marsden Farm field site.

System	Crop	Timing	Synthetic N fertilizer (kg N ha^{-1})	Composted manure N (kg N ha^{-1})	Greenhouse gas emission (kg CO_2-e ha^{-1} yr^{-1})[ǂ4]
2-year	Corn[ǂ1]	Pre-plant	106		570
		Side-dress	65		
	Soybean[ǂ1]				
3-year	Corn[ǂ1]	Side-dress	60		228
	Soybean[ǂ1]				
	Oat/red clover[ǂ1]	Post-Season[ǂ2]		105	
4-year	Corn[ǂ1]	Side-dress	52		152
	Soybean[ǂ1]				
	Oat/alfalfa[ǂ1]				
	Alfalfa[ǂ3]	Post-Season[ǂ2]		105	

ǂ1 25 kg N ha^{-1} synthetic N fertilizer were applied at pre-season during 2018-2020.
ǂ2 Post-season fertilizer values were calculated using data from 2005 to 2021.
ǂ3 12 kg N ha^{-1} synthetic N fertilizer were applied during the pre-season in 2018 and 2019.
ǂ4 Using the IPCC (2019)[33] emission factors of 1.6% for N fertilizer and 0.6% for manure N emitted as N_2O and 265 times global warming potential of CO_2 over a 100-year timescale.

Supplementary Table 6 Mean soil inorganic carbon (carbonate) contents in different crop production phases, rotations, and depths at the Marsden Farm field site. Units: mg C g^{-1}.

System	Crop	Rotation (length in years)	Soil depth			
			0-15cm	15-30cm	30-60cm	60-100cm
2-year	Corn	2	0.16	0.64	2.27	9.08
	Soybean	2	0.42	0.89	0.33	3.97
3-year	Corn	3	1.82	2.5	2.44	7.16
	Soybean	3	4.05	4.54	3.9	3.84
	Oat/red clover	3	1.51	2.67	3.92	10.25
4-year	Corn	4	6.08	6.24	6.42	6.55
	Soybean	4	0.02	0.03	0.01	1.47
	Oat/alfalfa	4	0.44	0.37	0.25	5.25
	Alfalfa	4	0.16	0.21	0.45	7.46

Supplementary Table 7 Soil organic carbon concentration and bulk density in different crop production phases, rotations, and depths at the Marsden Farm field site. Values are represented in the format of the mean value (standard error).

Crop	Rotation (length in years)	Soil depth (cm)	SOC (mg g^{-1})	Bulk density (g cm^{-3})
Corn	2	0-15	30.51 (0.71)	1.51 (0.05)
	2	15-30	22.75 (2.30)	1.43 (0.04)
	2	30-60	10.66 (2.74)	1.45 (0.06)
	2	60-100	5.59 (1.51)	1.53 (0.04)
Soybean	2	0-15	32.24 (4.79)	1.43 (0.08)
	2	15-30	24.32 (2.95)	1.42 (0.08)
	2	30-60	11.73 (6.41)	1.45 (0.13)
	2	60-100	5.02 (1.49)	1.54 (0.09)
Corn	3	0-15	28.94 (1.07)	1.42 (0.04)
	3	15-30	24.05 (3.49)	1.35 (0.10)
	3	30-60	10.35 (4.39)	1.36 (0.06)
	3	60-100	5.75 (2.99)	1.53 (0.06)
Soybean	3	0-15	27.96 (3.88)	1.43 (0.16)
	3	15-30	25.69 (4.69)	1.36 (0.07)
	3	30-60	11.74 (2.83)	1.45 (0.06)
	3	60-100	4.76 (1.34)	1.49 (0.03)
Oat/red clover	3	0-15	34.71 (1.27)	1.36 (0.03)
	3	15-30	25.46 (3.41)	1.29 (0.14)
	3	30-60	12.38 (5.56)	1.39 (0.06)
	3	60-100	5.48 (0.91)	1.53 (0.07)
Corn	4	0-15	31.08 (4.99)	1.38 (0.07)
	4	15-30	24.68 (6.31)	1.40 (0.19)
	4	30-60	10.3 (1.04)	1.50 (0.06)
	4	60-100	3.82 (1.48)	1.55 (0.06)
Soybean	4	0-15	29.11 (3.78)	1.47 (0.09)
	4	15-30	24.49 (2.62)	1.37 (0.05)
	4	30-60	11.36 (3.18)	1.43 (0.10)
	4	60-100	4.33 (1.35)	1.51 (0.06)
Oat/alfalfa	4	0-15	33.77 (5.72)	1.37 (0.09)
	4	15-30	27.89 (6.74)	1.31 (0.04)
	4	30-60	12.1 (1.07)	1.40 (0.06)
	4	60-100	4.79 (0.83)	1.53 (0.03)
Alfalfa	4	0-15	28.53 (6.12)	1.50 (0.17)
	4	15-30	24.04 (3.65)	1.42 (0.11)
	4	30-60	11.67 (1.58)	1.44 (0.06)
	4	60-100	5.05 (0.62)	1.54 (0.07)

Supplementary Table 8 Detailed balance equations for each C pool, soil respiration (CO_2 production), and their C isotope ratio ($\delta^{13}C$) in the Agro-IBIS soil C isotope-calibrated model. The detailed definition and description of parameters can be found in Supplementary Table 9.

Pool and flux respiration	Equation
Balance equations of each C pool size	
Root metabolic C (C_{rm})	$\frac{dC_{rm}}{dt} = -krm \times C_{rm}$
Root structural C (C_{rs})	$\frac{dC_{rs}}{dt} = -krs \times C_{rs}$
Root resistant (or lignin) C (C_{rl})	$\frac{dC_{3rl}}{dt} = -krl \times C_{rl}$
Non-protected active C (C_{NOC})	$\frac{dC_{NOC}}{dt} = lig\_frac \times (krl \times C_{NOC} \times yrl) + (1 - fbpom) \times (kbn \times C_{MICN} \times effac \times ybn + kbp \times C_{MICP} \times effac \times ybp) - knb \times C_{NOC} - kns \times C_{NOC}$
Protected active C (C_{POC})	$\frac{dC_{POC}}{dt} = (1 - lig_{frac}) \times (krl \times C_{NOC} \times yrl) + fbpom \times (kbn \times C_{MICN} \times effac \times ybn + kbp \times C_{MICP} \times effac \times ybp) - kpb \times C_{POC} - kps \times C_{POC}$
Passive (slow) C (C_S)	$\frac{dC_S}{dt} = kns \times C_{NOC} \times yns + kps \times C_{POC} \times yps - ksb \times C_S$
Microbial biomass C (CMIC)	$\frac{dC_{MIC}}{dt} = krm \times C_{rm} \times yrm + krs \times C_{rs} \times yrs + knb \times C_{NOC} \times ynb + kpb \times C_{POC} \times ypb + ksb \times C_S \times ysb - kbn \times C_{MICN} \times effac - kbp \times C_{MICP} \times effac$ If (C_{MIC} > fbsom× C_{SOC}) {C_{MICP}= fbsom× C_{SOC}} else {C_{MICP}= C_{MICP}} C_{MICN}= C_{MICP}−C_{MICP}
Soil respiration (C_R)	$\frac{dC_R}{dt} = krm \times C_{rm} \times (1 - yrm) + krs \times C_{rs} \times (1 - yrs) + knb \times C_{NOC} \times (1 - ynb) + kpb \times C_{POC} \times (1 - ypb) + ksb \times C_S \times (1 - ysb)$
Balance equations of $\delta^{13}C$ value in each C pool without microbial discrimination	
$\delta^{13}C$ of metabolic C ($\delta^{13}C_{rm}$), structural C ($\delta^{13}C_{rm}$) and resistant C (($\delta^{13}C_{rl}$)) in root	$\frac{d\delta^{13}C_{rm}}{dt} = \frac{d\delta^{13}C_{rs}}{dt} = \frac{d\delta^{13}C_{rl}}{dt} = 0$
$\delta^{13}C$ of non-protected active C ($\delta^{13}C_{NOC}$)	$\frac{d\delta^{13}C_{NOC}}{dt} = \frac{1}{(C_{NOC} + \frac{dC_{NOC}}{dt})}(lig\_frac \times (krl \times C_{NOC} \times yrl \times (\delta^{13}C_{rl} - \delta^{13}C_{NOC})) + (1 - fbpom) \times (kbn \times C_{MICN} \times effac \times ybn \times (\delta^{13}C_{MICN} - \delta^{13}C_{NOC}) + kbp \times C_{MICP} \times effac \times ybp \times (\delta^{13}C_{MICP} - \delta^{13}C_{NOC})))$ $\delta^{13}C_{MICN} = \delta^{13}C_{MICP} = \delta^{13}C_{MIC}$
$\delta^{13}C$ of protected active C ($\delta^{13}C_{POC}$)	$\frac{d\delta^{13}C_{POC}}{dt} = \frac{1}{(C_{POC} + \frac{dC_{POC}}{dt})}((1 - lig_{frac}) \times (krl \times C_{NOC} \times yrl \times (\delta^{13}C_{rl} - \delta^{13}C_{POC})) + fbpom \times (kbn \times C_{MICN} \times effac \times ybn \times (\delta^{13}C_{MICN} - \delta^{13}C_{POC}) + kbp \times C_{MICP} \times effac \times ybp \times (\delta^{13}C_{MICP} - \delta^{13}C_{POC})))$

Supplementary Table 8 Continued

δ^{13}C of passive (slow) C ($\delta^{13}C_S$)	$\frac{d\delta^{13}C_S}{dt} = \frac{1}{(C_S + \frac{dC_S}{dt})}(kns \times C_{NOC} \times yns \times (\delta^{13}C_{NOC} - \delta^{13}C_S) + kps \times C_{POC} \times yps \times (\delta^{13}C_{POC} - \delta^{13}C_S))$
δ^{13}C of microbial biomass C ($\delta^{13}C_{MIC}$)	$\frac{d\delta^{13}C_{MIC}}{dt} = \frac{1}{(C_{MIC} + \frac{dC_{MIC}}{dt})} \times (krm \times C_{rm} \times yrm \times (\delta^{13}C_{rm} - \delta^{13}C_{MIC}) + krs \times C_{rs} \times yrs \times (\delta^{13}C_{rs} - \delta^{13}C_{MIC}) + knb \times C_{NOC} \times ynb \times (\delta^{13}C_{NOC} - \delta^{13}C_{MIC}) + kpb \times C_{POC} \times ypb \times (\delta^{13}C_{POC} - \delta^{13}C_{MIC}) + ksb \times C_S \times ysb \times (\delta^{13}C_S - \delta^{13}C_{MIC}))$
δ^{13}C of soil respiration ($\delta^{13}C_R$)	$\frac{d\delta^{13}C_R}{dt} = \frac{1}{C_R} \times (krm \times C_{rm} \times (1 - yrm) \times \delta^{13}C_{rm} + krs \times C_{rs} \times (1 - yrs) \times \delta^{13}C_{rs} + knb \times C_{NOC} \times (1 - ynb) \times \delta^{13}C_{NOC} + kpb \times C_{POC} \times (1 - ypb) \times \delta^{13}C_{NOP} + ksb \times C_S \times (1 - ysb) \times \delta^{13}C_S)$
Balance equations of δ^{13}C value in each C pool with microbial discrimination	
δ^{13}C of microbial biomass C ($\delta^{13}C_{MIC}$)	$\frac{d\delta^{13}C_{MIC}}{dt} = \frac{1}{(C_{MIC} + \frac{dC_{MIC}}{dt})} \times (krm \times C_{rm} \times yrm \times (\delta^{13}C_{rm} + \varepsilon - \delta^{13}C_{MIC}) + krs \times C_{rs} \times yrs \times (\delta^{13}C_{rs} + \varepsilon - \delta^{13}C_{MIC}) + knb \times C_{NOC} \times ynb \times (\delta^{13}C_{NOC} + \varepsilon - \delta^{13}C_{MIC}) + kpb \times C_{POC} \times ypb \times (\delta^{13}C_{POC} + \varepsilon - \delta^{13}C_{MIC}) + ksb \times C_S \times ysb \times (\delta^{13}C_S + \varepsilon - \delta^{13}C_{MIC}))$
δ^{13}C of soil respiration ($\delta^{13}C_R$)	$\frac{d\delta^{13}C_R}{dt} = \frac{1}{C_R} \times (krm \times C_{rm} \times (1 - yrm) \times (\delta^{13}C_{rm} - \varepsilon \times \frac{yrm}{1 - yrm}) + krs \times C_{rs} \times (1 - yrs) \times (\delta^{13}C_{rs} - \varepsilon \times \frac{yrs}{1 - yrs}) + knb \times C_{NOC} \times (1 - ynb) \times (\delta^{13}C_{NOC} - \varepsilon \times \frac{ynb}{1 - ynb}) + kpb \times C_{POC} \times (1 - ypb) \times (\delta^{13}C_{NOP} - \varepsilon \times \frac{ypb}{1 - ypb}) + ksb \times C_S \times (1 - ysb) \times (\delta^{13}C_S - \varepsilon \times \frac{ysb}{1 - ysb}))$
The equations for the δ^{13}C values of the remaining C pools are the same as those in the C model with the assumption of no microbial discrimination.	

Supplementary Table 9 Parameters in the Agro-IBIS soil C isotope-calibrated model. The star symbol (*) denotes the parameter estimated by data assimilation in this study (calibrated values shown in Supplementary Fig. 16). The parameters without units are ratios. The efficiency represents the ratio with which organic C is used when being transformed in soil. In other words, the efficiency represents the ratio of C retained in biomass during decomposition when C is transferred from litter or soil conceptual C pools to the microbial C pool, in which case the value is less than 1. However, the efficiency is equal to 1 when C is transferred between litter and/or soil C pools (not including the microbial C pool).

Parameters	Description	Values	Units
krm	Decomposition rate of root metabolic C	0.1	day^{-1}
krs	Decomposition rate of root structural C	*	Day^{-}
krl	Decomposition rate of root resistant C	0.005	day^{-1}
knb	Decomposition rate of non-protected active C to microbial biomass C	*	day^{-1}
kpb	Decomposition rate of protected active C to microbial biomass C	*	day^{-1}
ksb	Decomposition rate of passive (slow) C to microbial biomass C	0.0000008	day^{-1}
kns	Transformation rate of non-protected active C to passive (slow) C	0.000001	day^{-1}
kps	Transformation rate of protected active C to passive (slow) C	0.000001	day^{-1}
kbp	Transformation rate of microbial biomass C to protected active C	0.005	day^{-1}
kbn	Transformation rate of microbial biomass C to non-protected active C	0.045	day^{-1}
yrm	Root metabolic C decomposition efficiency	0.4	—
yrs	Root structural C decomposition efficiency	0.3	—
ynb	Decomposition efficiency of non-protected active C to microbial biomass C	0.25	—
ypb	Decomposition efficiency of protected active C to microbial biomass C	0.2	—
ysb	Decomposition efficiency of passive (slow) C to microbial biomass C	0.2	—
yrl	Root resistant C decomposition efficiency	1	—
yns	Transformation efficiency of non-protected active C to passive (slow) C	1	—
yps	Transformation efficiency of protected active C to passive (slow) C	1	—
ybp	Transformation efficiency of microbial biomass C to protected active C	1	—
ybn	Transformation efficiency of microbial biomass C to non-protected active C	1	—
effac	Efficiency of microbial biomass C reincorporated into biomass pool	0.4	—
fbpom	Proportion of decomposing dead microbial biomass that is transferred to a protected pool vs. a non-protected pool	0.5	—
fbsom	Fraction of protected biomass to total soil organic carbon	0.017	—
lig_frac	Fraction of lignified litter material between protected vs. non-protected active C pools	0.5	—
ε	Carbon isotope discrimination during decomposition	*	‰

Supplementary Table 10 Detailed balance equations for each C pool, soil respiration (CO_2 production), and their C isotope ratio ($\delta^{13}C$) in the CN-SIM soil C model. The detailed definition and description of parameters can be found in Supplementary Table 11.

Pool and flux respiration	Equations
Balance equations of each C pool size	
Easily decomposable root C	$\frac{dC_{re}}{d_t} = [MIC] \times \frac{Vmax_{re} \times C_{re}}{K_m + C_{re}}$ $Vmax_{rm} = Kmax\_rm * f(T) * f(W) * f(N)$ K_m is the half − saturation constant: $K_m = e^{K_{coeff} * T + K_{int}} * ak * Kmod_{MBi}$ f(T) is the temperature scalar: $f(T) = 4.89 \times e^{-3.432+0.1\times T\times(1-0.5\times T/36.9)}$ f(W) is the soil moisture scalar: $f(W) = \begin{cases} \frac{1-e^{-\theta/\theta_{sat}}}{1-e^{-\theta_{fc}/\theta_{sat}}} & \theta \le \theta_{fc} \\ 1.0044 - \frac{0.0044}{e^{-5\frac{\theta/\theta_{sat}-\theta_{fc}/\theta_{sat}}{1-\theta_{fc}/\theta_{sat}}}} & \theta > \theta_{fc} \end{cases}$ f(N) is the nitrogen scalar: *If the decomposition is accompanied by nitrogen immobilization:* $f(N) = \frac{avn}{n_{imm}}$ *If decomposition is accompanied by nitrogen mineralization:* $f(N) = \begin{cases} 1 - \frac{avn - avn_{opt}}{avn_{opt}} & avn > avn_{opt} \\ 1 & avn_{opt}/2 \le avn \le avn_{opt} \\ 1 + \frac{0.5avn_{opt} - avn}{avn_{opt}} & avn \le avn_{opt}/2 \end{cases}$
Poorly decomposable root C	$\frac{dC_{rms}}{d_t} = [MIC] \times \frac{Vmax_{rm} \times C_{rm}}{K_m + C_{rm}}$ $Vmax_{rs} = Kmax\_rs * f(T) * f(W) * f(N)$
Particulate Organic C	$\frac{dC_{poc}}{d_t} = [MIC] \times \frac{Vmax_{poc} \times C_{poc}}{K_m + C_{poc}}$ $Vmax_{poc} = Kmax_{poc} * f(T) * f(W) * f(N)$

Supplementary Table 10 Continued

Mineral-associated Organic C	$$\frac{dC_{rm}}{d_t} = [MIC] \times \frac{Vmax_{moc} \times C_{moc}}{K_m + C_{moc}}$$ $$Vmax_{moc} = Kmax_{moc} * f(T) * f(W) * f(N) * f(clay)$$ f(clay) is the soil texture scalar: $$f(clay) = 1 - 0.75P_{clay}/100$$

$\delta^{13}C$ of the C pools are calculated based on the $^{13}C/^{12}C$ ratio in each C pool:

C pool $\delta^{13}C$ values	$$\delta^{13}C = \left[\frac{[^{13C}/_{12C}]Pool}{[^{13C}/_{12C}]standard} - 1\right] * 1000$$

Supplementary Table 11 Parameters in CN-SIM soil C model. The star symbol (*) denotes parameters estimated by data assimilation in this study (calibrated values shown in Supplementary Fig. 16).

	Parameter	Description	Values	Units
1	***Kmax_re*** * ***(kre)***	The maximum reaction velocity	[0.01, 1]	$year^{-1}$
2	*Kmax_rh*	The maximum reaction velocity	[1,10]	$year^{-1}$
3	***Kmax_poc**** ***(kpoc)***	The maximum reaction velocity	[0.1, 10]	$year^{-1}$
4	***Kmax_moc*** * ***(kmoc)***	The maximum reaction velocity	[10, 300]	$year^{-1}$
5	K_{coeff}	Regression coefficient	0.007	$\ln(mg\ C\ cm^{-3})\ ^{o}C^{-1}$
6	K_{int}	Regression intercept	3.19	$\ln(mg\ C\ cm^{-3})$
7	$Kmod_{MB\_r}$	Modifies Km for each substrate pool entering MB_r *(P_{clay}: clay %)*	0.125, 0.5, $1/(2.5 \times e^{-3\times fmet})$, $1/(1.4 + 0.2 \times P_{clay})$ ǂ1	-
8	$Kmod_{MB\_K}$	Modifies K_m for each C substrate pool entering MB_K *(P_{clay}: clay %)*	0.5, 0.25, $1/(2.5 \times e^{-3\times fmet})$, $1/(1.4 + 0.2 \times P_{clay})$ ǂ2	-
9	*aK*	Tuning coefficient	10	-
10	f_{met}	Partitioning of root inputs to easily decomposed root C pool *(C_{root}: root C, N: root N, CN_{re}, CN_{rh}: C: N ratio for easily/poorly decomposable C pool)*	$\frac{C_{root} - CN_{re} \times N_{root}}{CN_{rh} - CN_{re}} * \frac{CN_{rh}}{C_{root}}$	-
11	*MGE*	Microbial growth efficiency for substrate pools	0.6, 0.3, 0.6, 0.3 ǂ3	$mg\ mg^{-1}$
12	τ	Microbial biomass turnover rate	$5.25 \times e^{0.9\times fmet}$ (MB_r), 2.6 (MB_K)	$year^{-1}$
13	fc	Fraction of τ partitioned to MOC *(MOC: mineral-associated organic C)*	$0.2 \times e^{-2\times fmet}$ (MB_r) $0.4 \times e^{-3\times fmet}$ (MB_K)	-

ǂ1: For easily decomposable, poorly composable, particulate organic carbon, mineral-associated organic carbon fluxes entering MB_r, respectively.
ǂ2: For easily decomposable, poorly composable, particulate organic carbon, mineral-associated organic carbon fluxes entering MB_K, respectively.
ǂ3: For easily decomposable, poorly composable, particulate organic carbon, mineral-associated organic carbon pool, respectively.

Supplementary Table 12 Governing equations of each soil carbon pool in the MEND model. The transformation fluxes are elucidated by Eqs. S13–S27 in Supplementary Table 13.

Carbon pool	Governing Equation	Eq#
Particulate organic C (POC) pool 1 (P1) decomposed by oxidative enzymes	$\frac{dP_1}{dt} = I_{P1} + (1 - g_D) \cdot F_{12} - F_1$	(S1)
POC pool 2 (P2) decomposed by hydrolytic enzymes	$\frac{dP_2}{dt} = I_{P2} - F_2$	(S2)
Mineral-associated organic C (MOC, M)	$\frac{dM}{dt} = (1 - f_D) \cdot (F_1 + F_2) - F_3$	(S3)
Active MOC interacting with dissolved organic C (QOC, Q)	$\frac{dQ}{dt} = F_4 - F_5$	(S4)
Dissolved organic C (DOC, D)	$\frac{dD}{dt} = I_D + f_D \cdot (F_1 + F_2) + g_D \cdot F_{12} + F_3 + (F_{14,EP1} + F_{14,EP2} + F_{14,EM}) - F_6 - (F_4 - F_5)$	(S5)
Active microbial biomass (MBA, BA)	$\frac{dBA}{dt} = F_6 - (F_7 - F_8) - (F_9 + F_{10} + F_{15}) - F_{12} - (F_{13,EP1} + F_{13,EP2} + F_{13,EM})$	(S6)
Dormant microbial biomass (MBC, BD)	$\frac{dBD}{dt} = (F_7 - F_8) - F_{11}$	(S7)
Oxidative enzymes decomposing P1	$\frac{dEP_1}{dt} = F_{13,EP1} - F_{14,EP1}$	(S8)
Hydrolytic enzymes decomposing P2	$\frac{dEP_2}{dt} = F_{13,EP2} - F_{14,EP2}$	(S9)
Enzymes decomposing MOC	$\frac{dEM}{dt} = F_{13,EM} - F_{14,EM}$	(S10)
Carbon dioxide (CO_2)	$\frac{dCO_2}{dt} = (F_9 + F_{10}) + F_{11} - F_{15} + F_{16}$	(S11)
Carbon balance	$\frac{d}{dt}(P_1 + P_2 + M + Q + D + BA + BD + EP_1 + EP_2 + EM) = (I_{P1} + I_{P2} + I_D) - (F_9 + F_{10} + F_{11} + F_{15})$	(S12)

Supplementary Table 13 Component fluxes in the MEND model. The parameters are described in Supplementary Table 14.

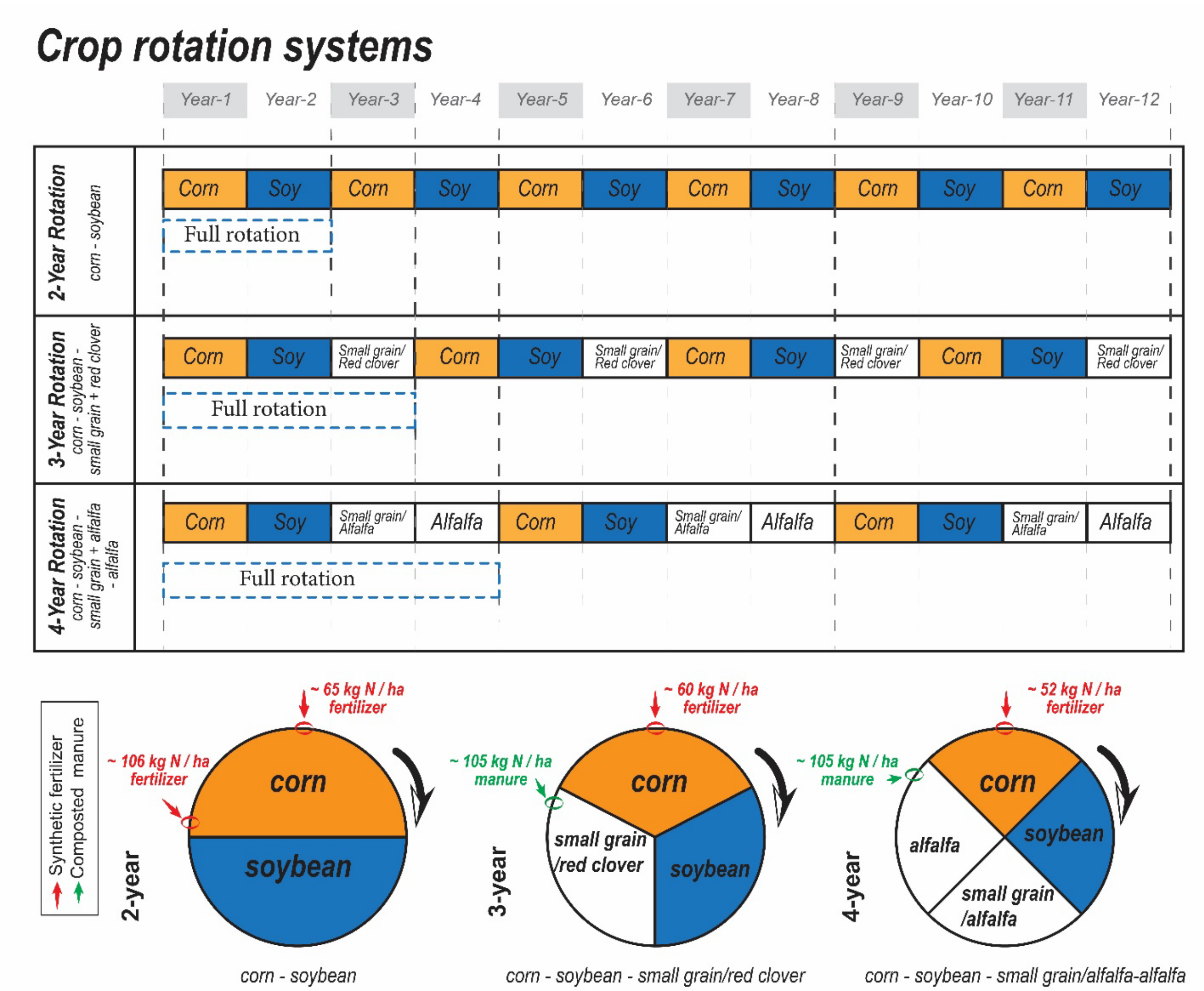
Flux description	Equation	Eq#
Particulate organic carbon (POC) pool 1 (P_1) decomposition (F_1)	$F_1 = \frac{Vd_{P1} \cdot EP_1 \cdot P_1}{K_{P1} + P_1}$	(S13)
POC pool 2 (P_2) decomposition	$F_2 = \frac{Vd_{P2} \cdot EP_2 \cdot P_2}{K_{P2} + P_2}$	(S14)
Mineral-associated organic carbon (MOC, *M*) decomposition	$F_3 = \frac{Vd_M \cdot EM \cdot M}{K_M + M}$	(S15)
Adsorption (F_4) and desorption (F_5) between dissolved organic carbon (DOC, *D*) and adsorbed DOC (QOC, *Q*)	$F_4 = k_{ads} \cdot (1 - Q/Q_{\max}) \cdot D$ $F_5 = k_{des} \cdot (Q/Q_{\max})$	(S16) (S17)
DOC (*D*) uptake by microbes	$F_6 = \frac{1}{Y_g}(V_g + V_m)\frac{D \cdot BA}{K_D + D}$	(S18)
Dormancy (F_7) and reactivation (F_8) between active (MBA) and dormant (MBD) microbial biomass (*BA* and *BD*)	$F_7 = [1 - D/(K_D + D)] \cdot V_{AD} \cdot BA$ $F_8 = [D/(K_D + D)] \cdot V_{DA} \cdot BD$	(S19) (S20)
MBA (*BA*) growth respiration (F_9) and maintenance respiration (F_{10})	$F_9 = \left(\frac{1}{Y_g} - 1\right)\frac{V_g \cdot BA \cdot D}{K_D + D}$ $F_{10} = \left(\frac{1}{Y_g} - 1\right)\frac{V_m \cdot BA \cdot D}{K_D + D}$	(S21) (S22)
MBD (*BD*) maintenance respiration	$F_{11} = \beta \cdot V_m \cdot BD$	(S23)
MBA (*BA*) mortality	$F_{12} = \gamma \cdot V_m \cdot BA$	(S24)
Synthesis of enzymes for P_1 (EP_1, $F_{13,\mathrm{EP1}}$), enzymes for P_2 (EP_2, $F_{13,\mathrm{EP2}}$), and enzymes for *M* (*EM*, $F_{13,\mathrm{EM}}$)	$F_{13,EP1} = P_1/(P_1 + P_2) \cdot p_{EP} \cdot V_m \cdot BA$ $F_{13,EP2} = P_2/(P_1 + P_2) \cdot p_{EP} \cdot V_m \cdot BA$ $F_{13,EM} = p_{EM} \cdot V_m \cdot BA$	(S25)
Turnover of enzymes (EP_1, EP_2, *EM*)	$F_{14,EP1} = r_E \cdot EP_1$ $F_{14,EP2} = r_E \cdot EP_2$ $F_{14,EM} = r_E \cdot EM$	(S26)

Notes: Italic symbols like F_i represent component fluxes in equations. Italic symbols P_1, P_2, *M*, *Q*, *D*, *BA*, *BD*, EP_1, EP_2, and *EM* are state variables (soil carbon pools) in equations.

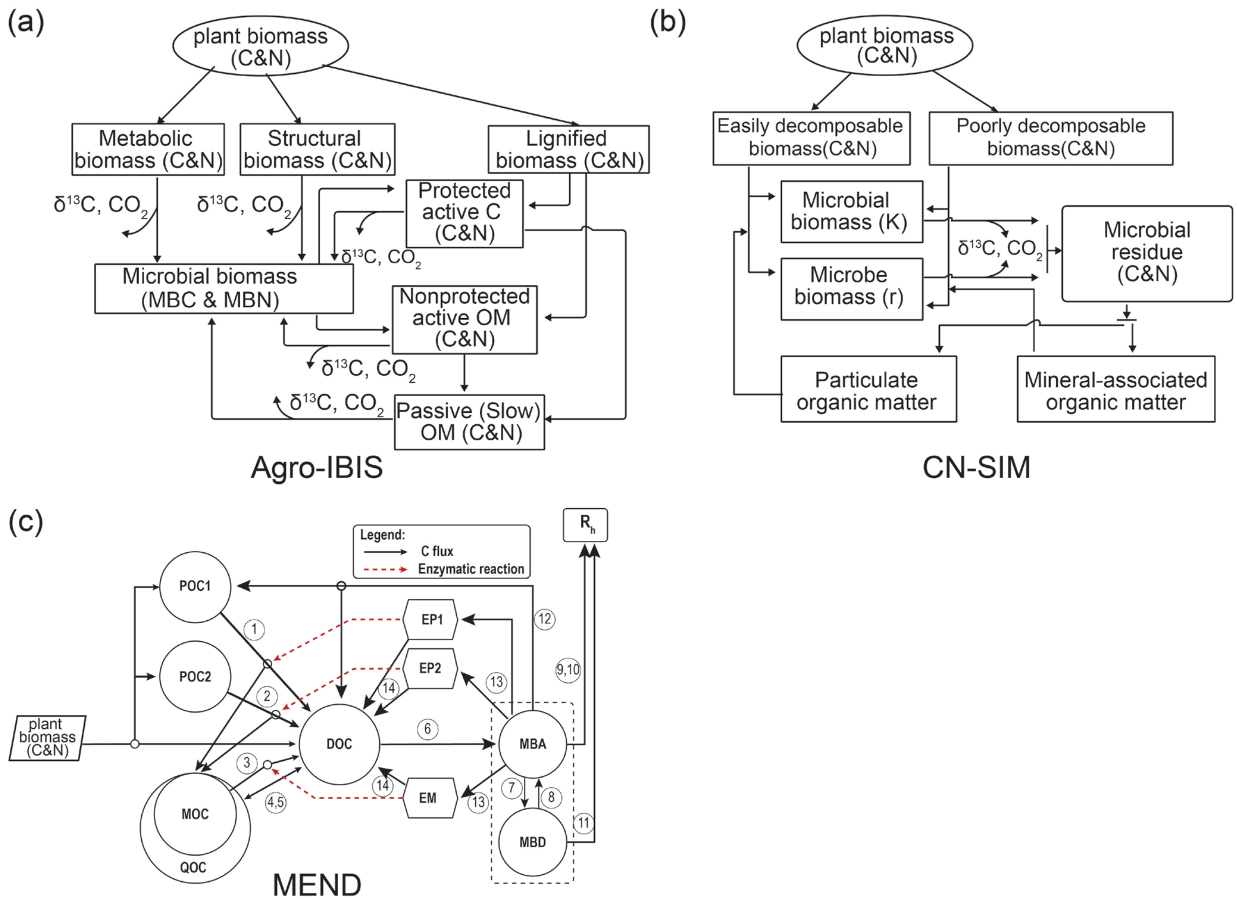
Supplementary Table 14 The MEND model parameters. The star symbol (*) denotes the parameter estimated by data assimilation in this study (calibrated values shown in Supplementary Fig. 16).

ID	Parameter	Description	Units	Prior range	Value
1	LF_0	Initial fraction of P_1, $LF_0 = P_1/(P_1+P_2)$	—	(0.1, 1)	0.1
2	r_0	Initial active fraction of microbes, $r_0 = BA/(BA+BD)$	—	(0.01, 1)	*
3	Vd_{P1}	Maximum specific decomposition rate for P_1	mg C mg^{-1} C h^{-1}	(0.1, 100)	20
4	Vd_{P2}	Maximum specific decomposition rate for P_2	mg C mg^{-1} C h^{-1}	(0.1, 100)	2
5	Vd_M	Maximum specific decomposition rate for M	mg C mg^{-1} C h^{-1}	(0.1, 100)	2
6	K_{P1}	Half-saturation constant for P_1 decomposition	mg C g^{-1} soil	(40, 100)	51
7	K_{P2}	Half-saturation constant for P_2 decomposition	mg C g^{-1} soil	(1, 40)	6
8	K_M	Half-saturation constant for M decomposition	mg C g^{-1} soil	(100, 1000)	503
9	$Q_{\max}$	Maximum sorption capacity	mg C g^{-1} soil	(0.5, 5)	3.7
10	K_{ba}	Binding affinity, Sorption rate $k_{ads} = k_{des} \times K_{ba}$	(mg C g^{-1} soil)$^{-1}$	(1, 16)	6
11	k_{des}	Desorption rate	mg C g^{-1} soil h^{-1}	(0.0001, 0.01)	0.005
12	r_E	Turnover rate of EP_1, EP_2, and EM	mg C mg^{-1} C h^{-1}	(0.0001, 0.01)	0.003
13	p_{EP}	$[V_m \times p_{EP}]$ is the production rate of EP ($EP_1 + EP_2$), V_m is the specific maintenance rate for BA	—	(0.0001, 0.05)	0.05
14	fp_{EM}	$fp_{EM} = p_{EM}/p_{EP}$, $[V_{mt} \times p_{EM}]$ is the production rate of EM	—	(0.5, 3.0)	*
15	f_D	Fraction of decomposed P_1 and P_2 allocated to D	—	(0.05, 1)	0.5
16	g_D	Fraction of dead BA allocated to D	—	(0.01, 1)	0.5
17	V_g	Maximum specific uptake rate of D for growth	mg C mg^{-1} C h^{-1}	(0.001, 0.1)	*
18	V_m	Maximum specific uptake rate of D for maintenance	mg C mg^{-1} C h^{-1}	(0.00001, 0.1)	*
19	K_D	Half-saturation constant for microbial uptake of D	mg C g^{-1} soil	(0.01, 0.5)	0.02
20	Y_g	True growth yield at reference temperature (T_{ref})	—	(0.1, 0.6)	*
21	k_{Yg}	Slope for Y_g dependence of temperature	1/°C	(0.001, 0.016)	0.0073
22	γ	Max microbial mortality rate = $V_m \times \gamma$	—	(0.1, 20)	1
23	β	Ratio of dormant maintenance rate to V_m	—	(0.0005, 0.05)	0.001
24	ψ_{A2D}	Soil water potential (SWP) threshold for microbial dormancy; both ψ_{A2D} & $\psi_{D2A} < 0$	–MPa	(0.01, 0.6)	0.07
25	τ	$\psi_{D2A} = \psi_{A2D} \times \tau$, ψ_{D2A} is the SWP threshold for microbial resuscitation	—	(0.1, 0.95)	0.25
26	ω	Exponential in SWP function for microbial dormancy or resuscitation	—	(0.5, 6)	4

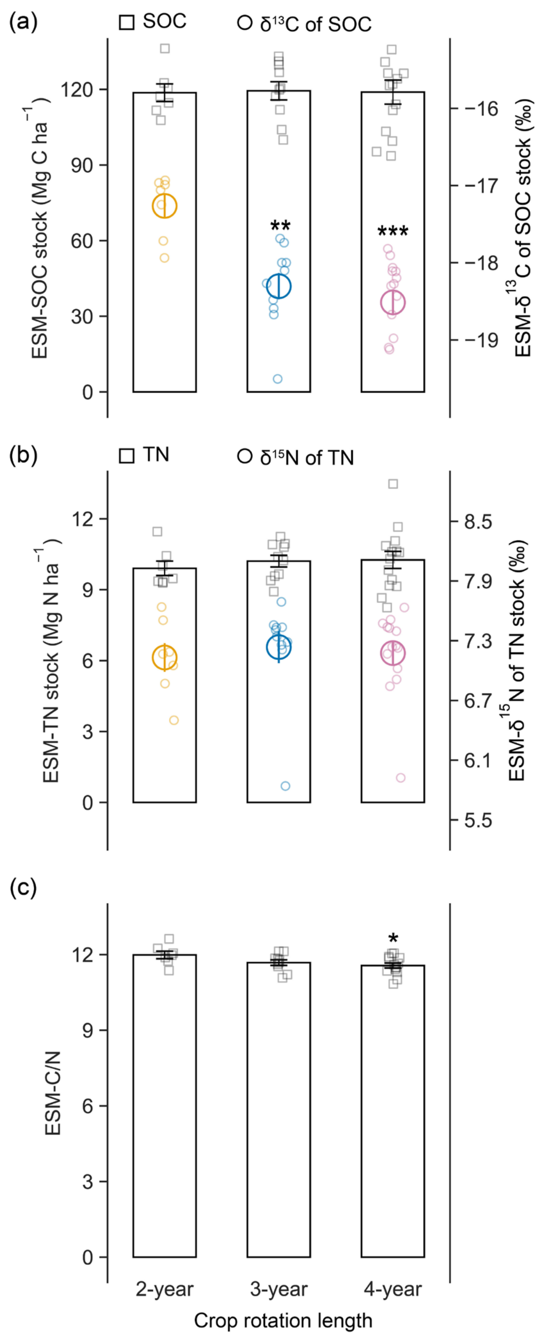
Supplementary Figs. 1 to 19



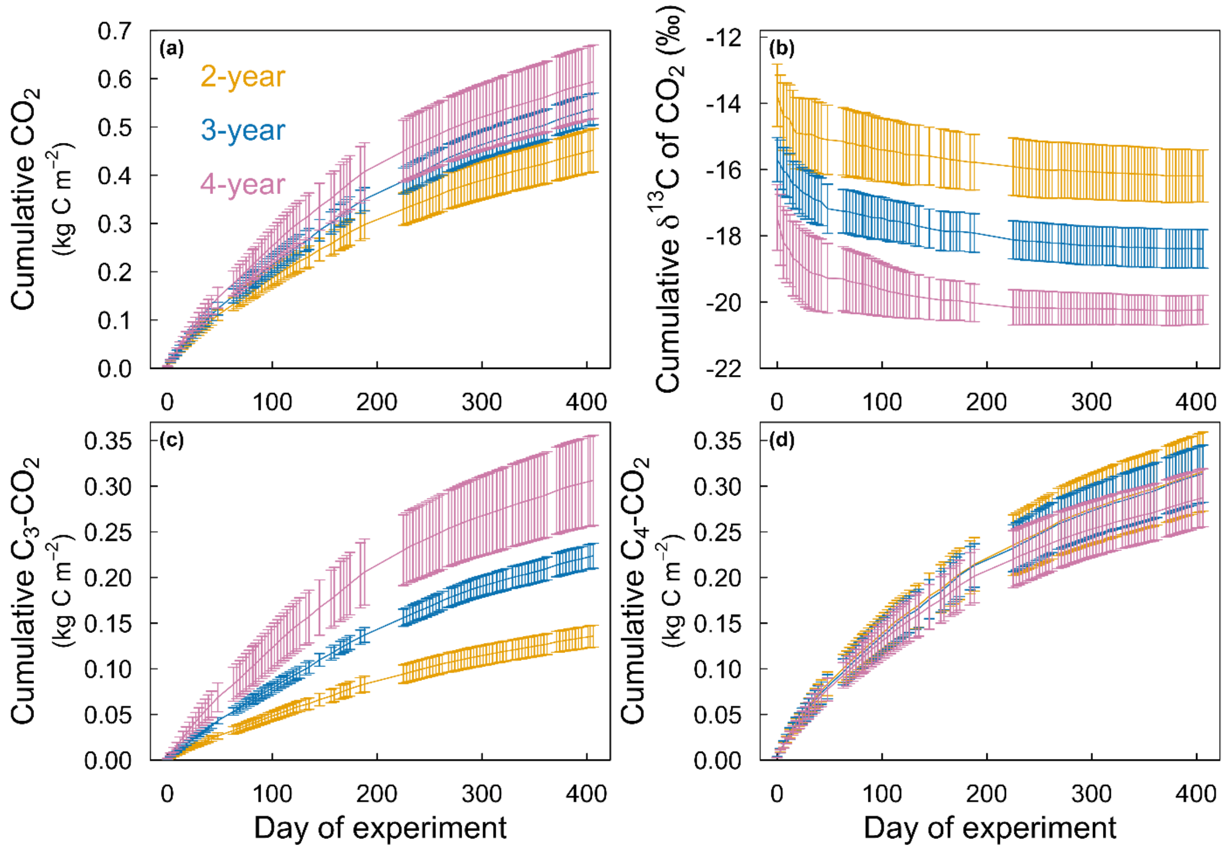
Supplementary Fig. 1 Three different rotation systems in the Marsden Farm experiment, including a 2-year corn-soybean rotation, a 3-year corn-soybean-small grain/red clover rotation, and a 4-year corn-soybean-small grain/alfalfa-alfalfa rotation



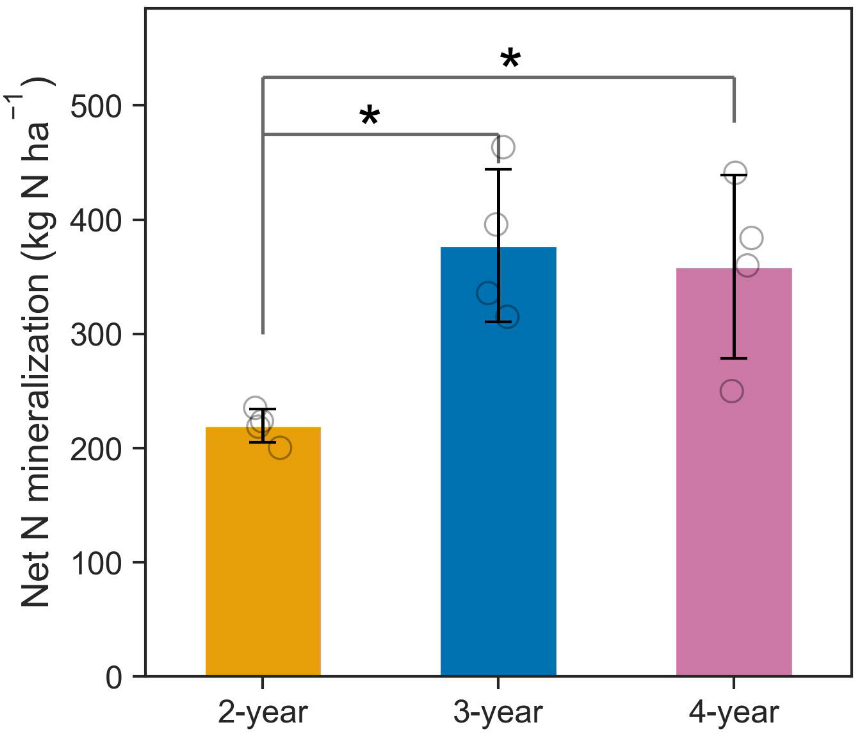
Supplementary Fig. 2 Flowchart of soil carbon (C) in the Agro-IBIS, CN-SIM, and MEND models. (a) Agro-IBIS model redrawn from Verberne et al.[26]; (b) CN-SIM Model, POC: particulate organic C; MOC: mineral-associated organic C. (c) Microbial-ENzyme Decomposition (MEND) model. POC, particulate organic C; MOC, mineral-associated organic C; DOC, dissolved organic C; QOC, adsorbed phase of DOC or active part of MOC; MBA, active microbial biomass; MBD, dormant microbial biomass; EP1, EP2, and EM, enzymes for decomposing POC1, POC2, and MOC. Transformation fluxes include: (1, 2) POC1 and POC2 decomposition, (3) MOC decomposition, (4, 5) adsorption and desorption between DOC and QOC, (6) DOC uptake by MBA, (7, 8) microbial dormancy and reactivation between MBA and MBD, (9, 10) MBA growth and maintenance respiration, (11) MBD maintenance respiration, (12) MBA mortality (13) Synthesis of EP1 and EP2, and EM, and (14) Mortality (turnover) of enzymes.



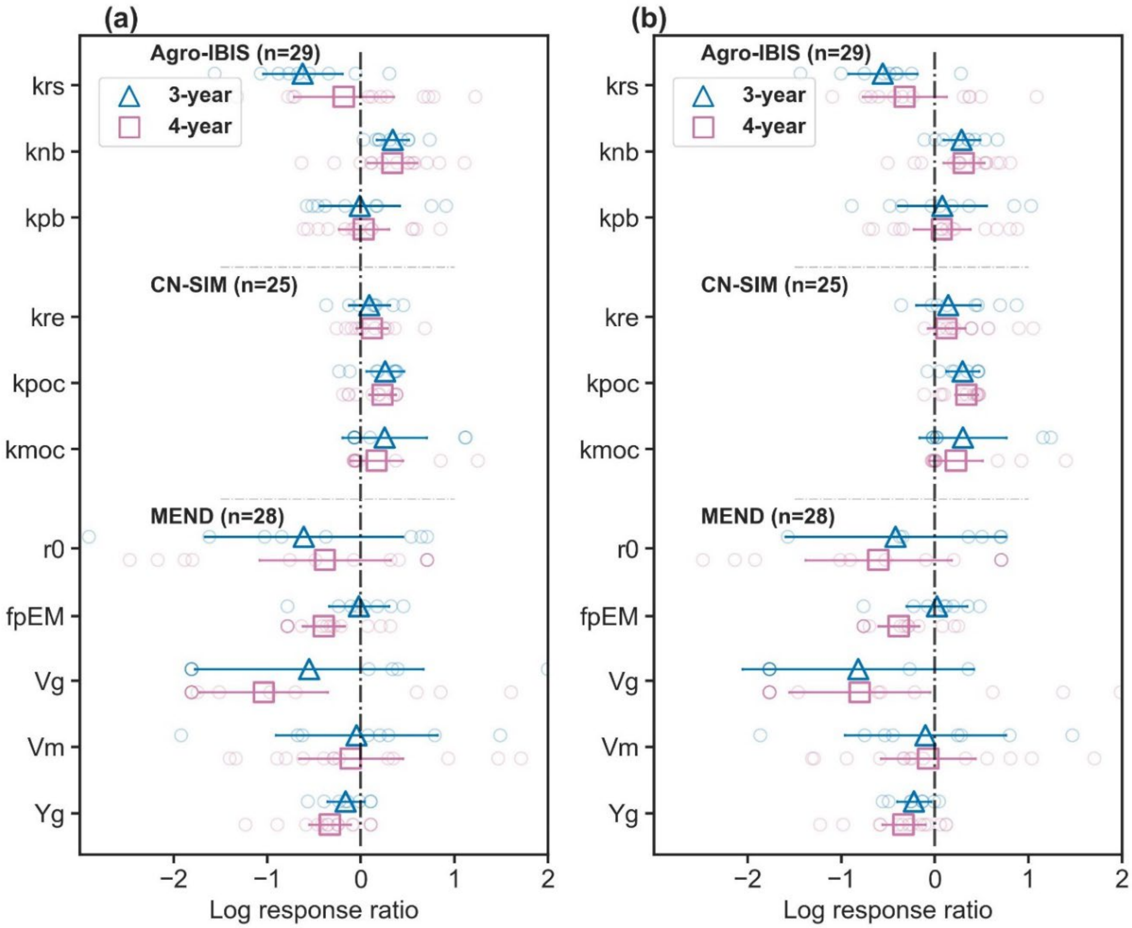
Supplementary Fig. 3 ESM-based soil organic carbon (SOC), total nitrogen (TN), their mass-weighted C and N isotope values (circles) and SOC/TN ratios from soils collected to 30 cm depth under a 2-year corn-soybean rotation (2-year), a corn-soybean-small grain/red clover rotation (3-year), and a 4-year corn-soybean-small grain/alfalfa-alfalfa rotation (4-year). The values were calculated by equivalent soil mass (ESM) approach, which used 0–30 cm soil mass at 2-year rotation as a reference. The error bars indicate standard errors (n = 8, 12, and 16 for 2-, 3- and 4-year rotations, respectively), with the center indicating the mean value of the replicates. ** $P = 0.001$, *** $P = 8.1\times10^{-5}$ in one-way ANOVA followed by Dunnett's test. The circle within the bar plot represents isotopic values: yellow for 2-year, blue for 3-year, and pink for 4-year rotations.



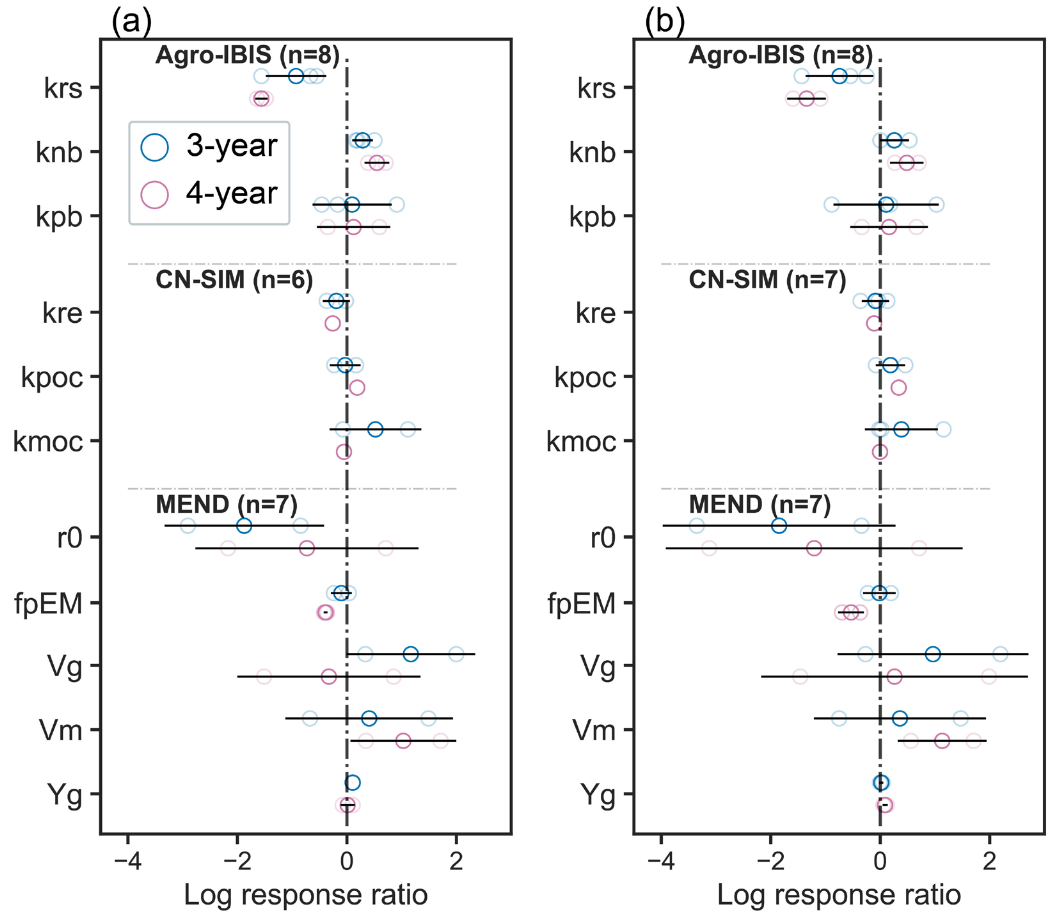
Supplementary Fig. 4 Lab observations of carbon (C) decomposition from soil cores (1 m depth), only including data from the corn production phase from a 2-year corn-soybean rotation (2-year), a corn-soybean-small grain/red clover rotation (3-year), and a 4-year corn-soybean-small grain/alfalfa-alfalfa rotation (4-year). (a) Cumulative CO_2 production; (b) Cumulative C isotope ratio ($\delta^{13}C$) of CO_2 production; (c, d) Cumulative decomposition of C_3-derived C (c) and C_4-derived C (d) respired as CO_2. The error bars indicate standard errors (n = 8, 12, and 16 for 2-, 3- and 4-year rotations, respectively), with the center indicating the mean value of the replicates.



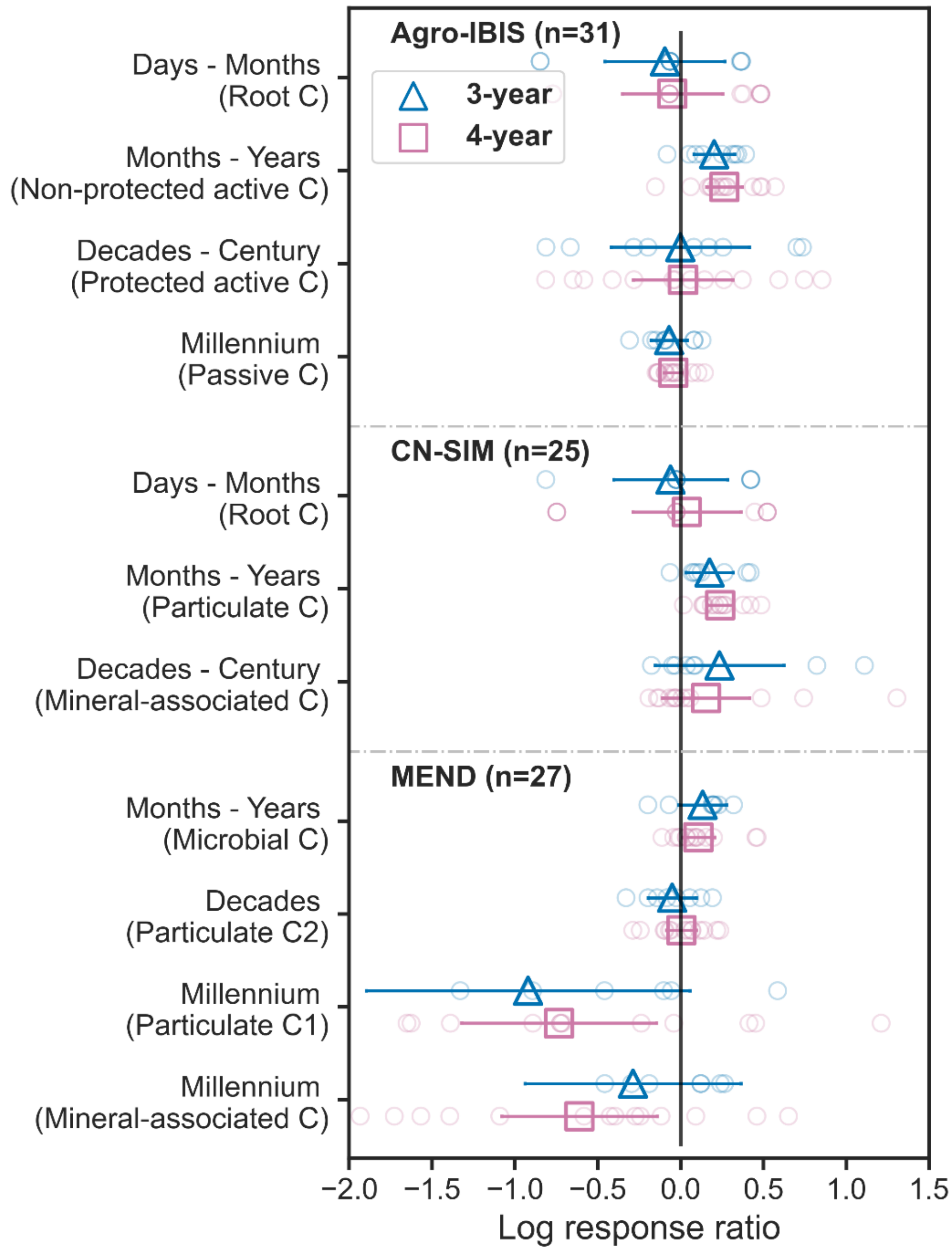
Supplementary Fig. 5 Lab observations of net N mineralization from soil cores (1 m depth), only including data from the corn production phase from a 2-year corn-soybean rotation (2-year), a corn-soybean-small grain/red clover rotation (3-year), and a 4-year corn-soybean-small grain/alfalfa-alfalfa rotation (4-year). The error bars indicate standard errors (n = 4, 4, and 4 for 2-, 3- and 4-year rotations, respectively), with the center denoting the mean value of the replicates. * $P = 0.02$ and 0.04 in one-way ANOVA followed by Dunnett's test for 2-year vs. 3-year and 2-year vs. 4-year comparisons, respectively.



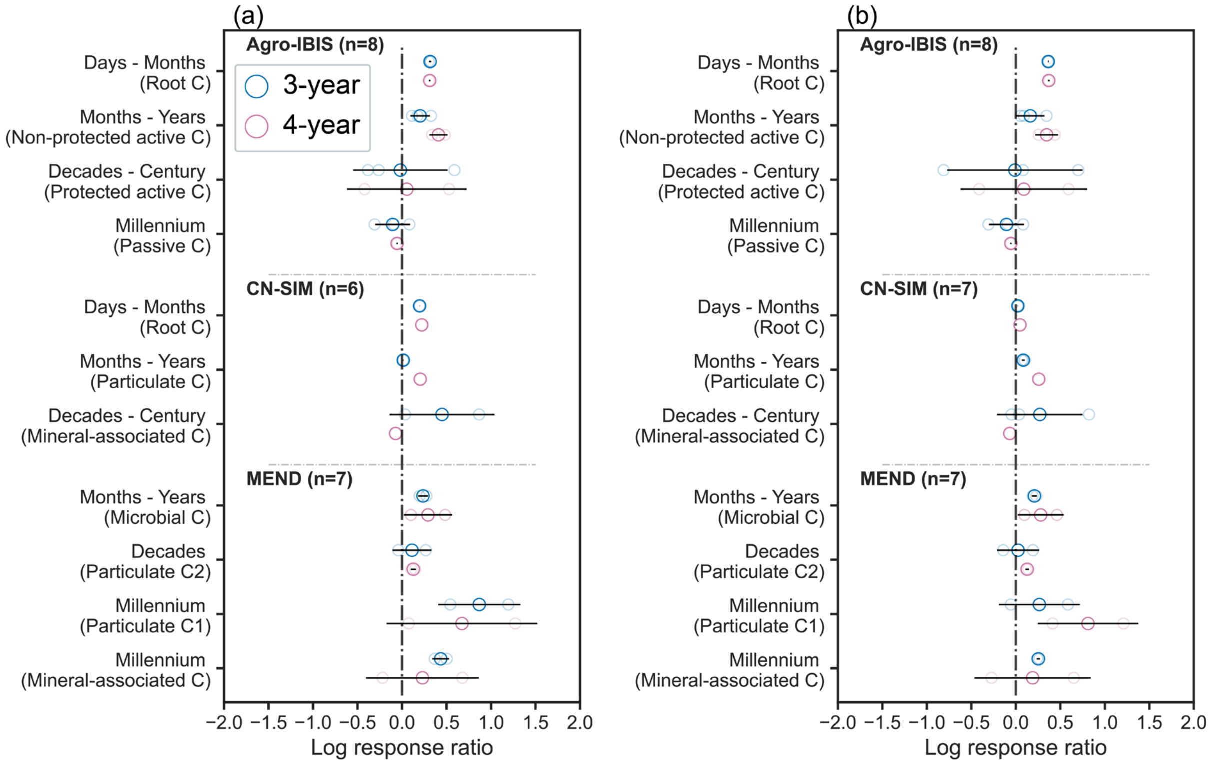
Supplementary Fig. 6 Effects of diversified cropping systems (3-year and 4-year rotations) on the influential parameters in three models relative to a conventional corn-soybean rotation (2-year rotation). The log response ratio is the natural logarithm of the value in the 3- or 4-year rotations divided by the value in the 2-year rotation. Panel (a) represents the scenario with the best-estimated values for initial C pools based on previous measurements, and panel (b) represents the scenario with plausible lower-bound values for initial C pools. The error bars represent 95% confidence intervals, with the center denoting the mean value of the replicates. Dark symbols are the mean log response ratios, and partially transparent points show the individual value for each replicate plot. The n values (n = 29, 25, and 28 for Agro-IBIS, CN-SIM, and MEND, respectively) listed after the name of the model in each panel represent the numbers of plots that achieved satisfactory model performance (determination $R^2 > 0.6$ for CO_2 and RMSE < 2 ‰ for the $\delta^{13}C$ values) in model simulation. The model parameters listed on the y-axis are defined as follows: krs, decomposition rate of root structural C; knb, decomposition rate of non-protected active C transferred to microbial biomass C; kpb, decomposition rate of protected active C transferred to microbial biomass C; kre, the decomposition rate of easily decomposable root C pool; kpoc, decomposition rate of particulate organic C pool; kmoc: decomposition rate of mineral-associated organic C pool; r0, initial active fraction of microbes; fpEM, related to enzyme production of MOC (details shown in Supplementary Table 14); Vg, maximum specific uptake rate of *D* for growth; Vm, maximum specific uptake rate of *D* for maintenance; Yg, true growth yield at reference temperature.



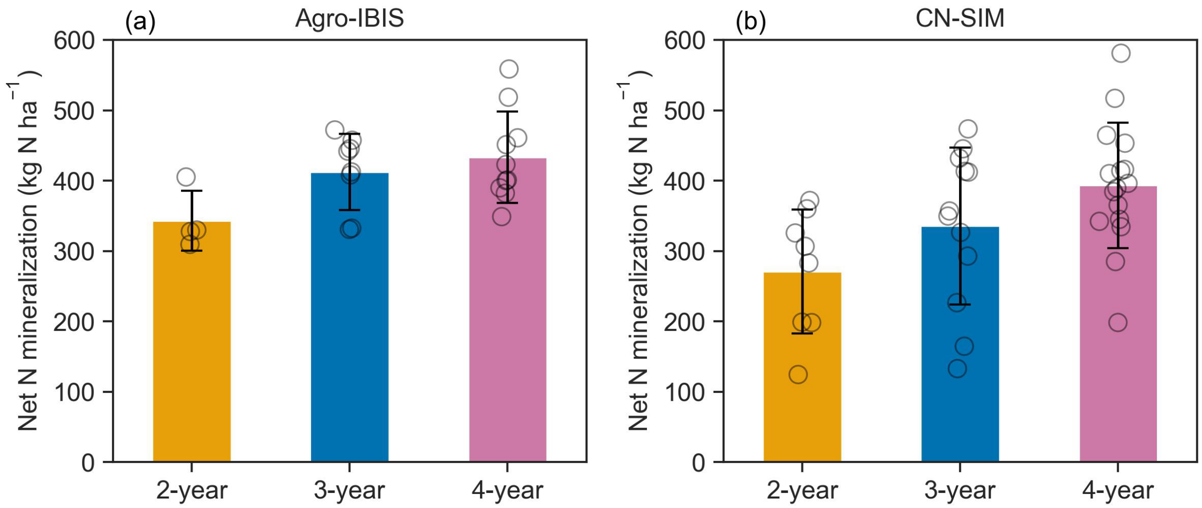
Supplementary Fig. 7 Comparison of the influential parameters in three models during only the corn production phase of the rotations. The log response ratio is the natural logarithm of the value in the corn plots of the 3- or 4-year rotations divided by the value in the corn plots of the 2-year rotation. Panel (a) represents the scenario with the best-estimated values for initial C pools based on previous measurements, and panel (b) represents the scenario with plausible lower-bound values for initial C pools. The error bars represent 95% confidence intervals, with the center denoting the mean value of the replicates. The n values (n = 8, 6, and 7 for Agro-IBIS, CN-SIM, and MEND for the best-estimate initial C pools scenario, and n = 8, 7, and 7 for the lower-bound initial C pools scenario, respectively) listed after the name of the model in each panel represent the numbers of plots that achieved satisfactory performance (determination $R^2 > 0.6$ for CO_2 and RMSE < 2 ‰ for the $\delta^{13}C$ value) in model simulation. Dark symbols are the mean log response ratios, and partially transparent points show the individual value for each replicate plot. The model parameters listed on the y-axis are defined as follows: krs, decomposition rate of root structural C; knb, decomposition rate of non-protected active C transferred to microbial biomass C; kpb, decomposition rate of protected active C transferred to microbial biomass C; kre, the decomposition rate of easily decomposable root C pool; kpoc, decomposition rate of particulate organic C pool; kmoc: decomposition rate of mineral-associated organic C pool; r0, initial active fraction of microbes; fpEM, related to enzyme production of MOC (details shown in Supplementary Table 14); Vg, maximum specific uptake rate of *D* for growth; Vm, maximum specific uptake rate of *D* for maintenance; Yg, true growth yield at reference temperature.



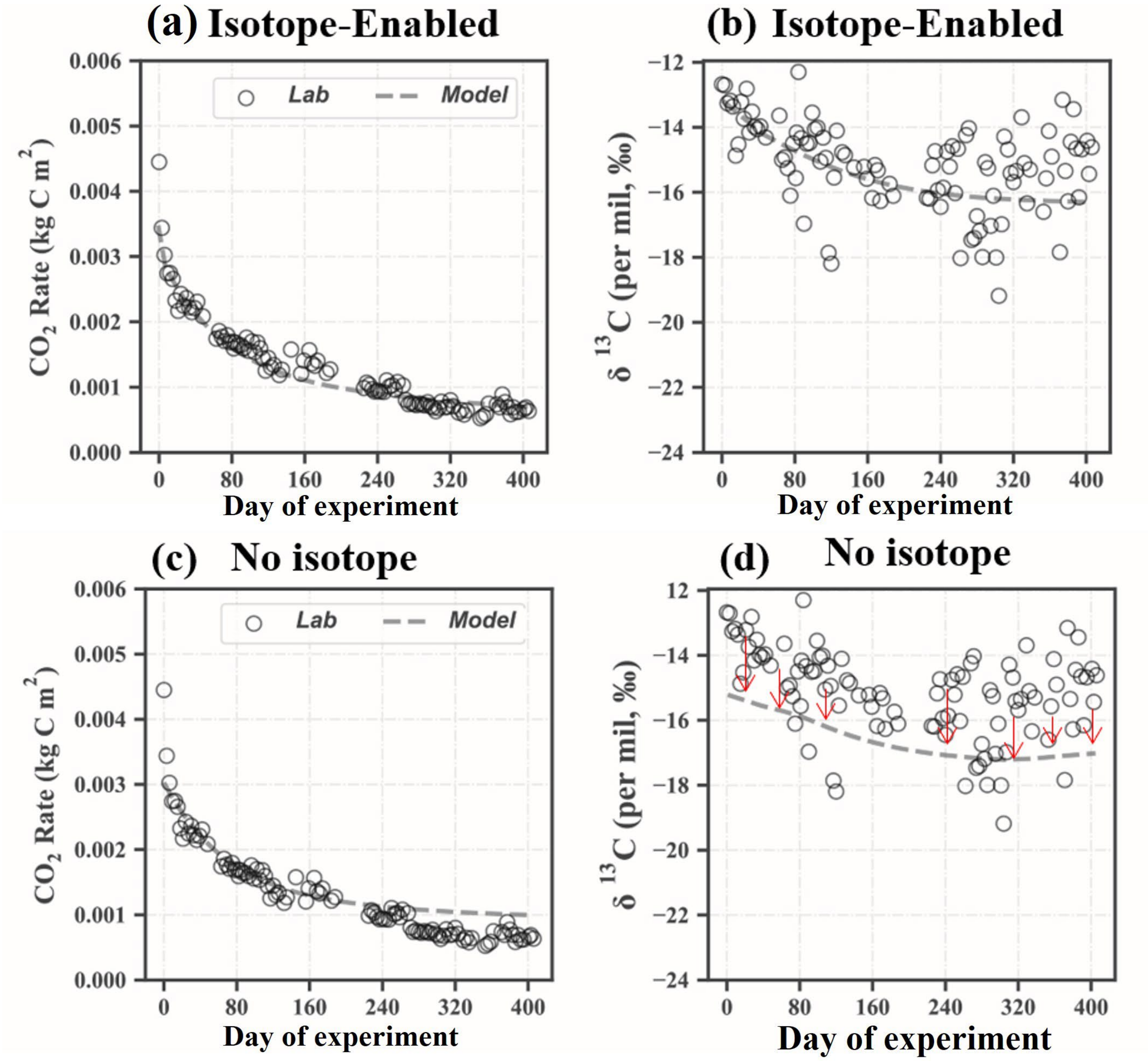
Supplementary Fig. 8 Effects of the diversified cropping systems (3-year and 4-year crop rotations) on decomposition of different C pools relative to a conventional corn-soybean rotation (2-year rotation), as simulated using three mechanistic models driven by the plausible lower-bound root C input and calibrated with laboratory CO_2 fluxes and $\delta^{13}C$ of CO_2. The log response ratio is the natural logarithm of the value in the 3- or 4-year rotations divided by the value in the 2-year rotation. The error bars represent 95% confidence intervals, with the center denoting the mean value of the replicates. Dark symbols are the mean log response ratios, and partially transparent points show the individual values for each replicate plot. The n (n = 31, 25, and 27 for Agro-IBIS, CN-SIM, and MEND, respectively) values listed after the name of the model in each panel represent the numbers of plots that achieved satisfactory model performance (determination $R^2 > 0.6$ for CO_2 and RMSE < 2 ‰ for the $\delta^{13}C$ values) in model simulation.



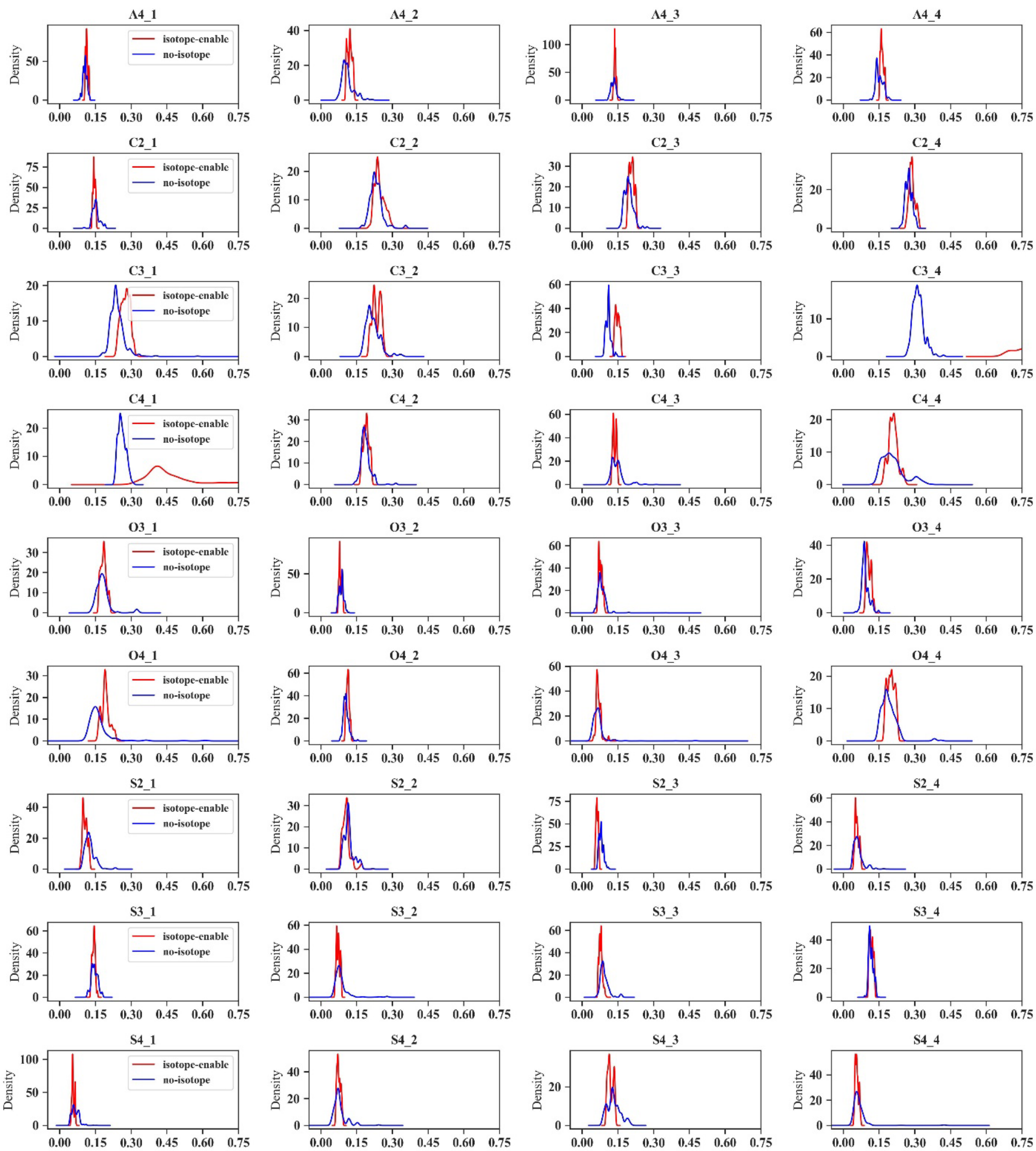
Supplementary Fig. 9 Comparison of decomposition of different carbon pools during the corn production phase as simulated using three mechanistic models calibrated with laboratory CO_2 fluxes and $\delta^{13}C$ of CO_2. The log response ratio is the natural logarithm of the value in the corn plots of the 3- or 4-year rotations divided by the value in the corn plots of the 2-year rotation. Panel (a) represents the simulation results using the best-estimate values for initial C pools based on previous measurements, and panel (b) represents the scenario with plausible lower-bound root C values for initial C pools. The error bars represent 95% confidence intervals, with the center denoting the mean value of the replicates. The n values (n = 8, 9, and 7 for Agro-IBIS, CN-SIM, and MEND for the best-estimate initial C pools scenario, and n = 8, 9, and 7 for the lower-bound initial C pools scenario, respectively) listed after the name of the model in each panel represent the numbers of plots that achieved satisfactory performance (determination $R^2 > 0.6$ for CO_2 and RMSE < 2 ‰ for the $\delta^{13}C$ value) in model simulation.



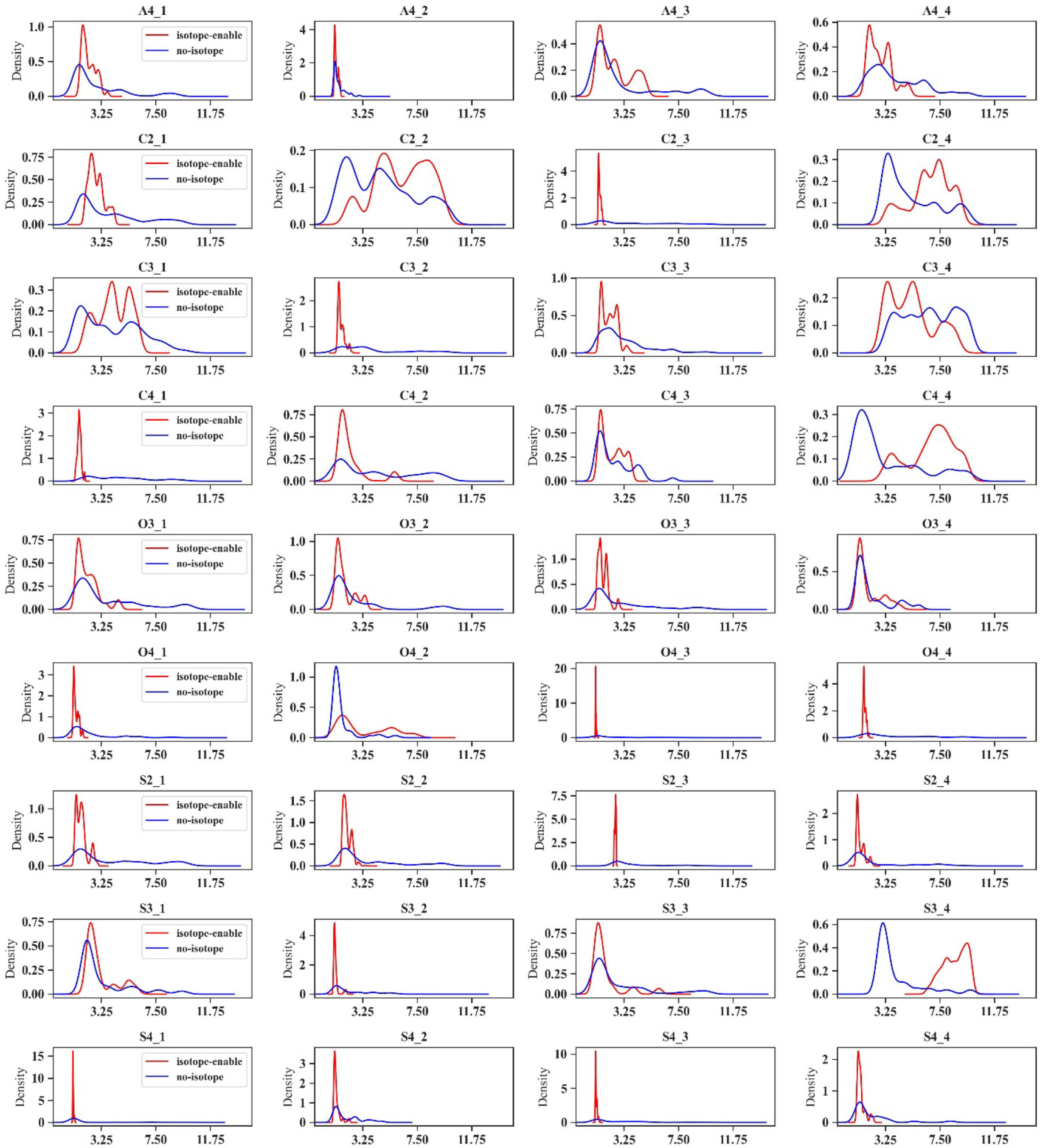
Supplementary Fig. 10 Simulated net N mineralization by the isotope-calibrated models after 13.5 months of soil core incubation (1 m depth) under best-estimate input data scenarios. The error bars indicate standard errors, with the center denoting the mean value of the replicates. The n values are 4, 8, and 10 for the 2-, 3-, and 4-year rotations in Agro-IBIS, and 8, 12, and 16 for the 2-, 3-, and 4-year rotations in CN-SIM, respectively, representing the number of plots that achieved satisfactory performance ($R^2 > 0.6$ for CO_2 and RMSE < 2 ‰ for $\delta^{13}C$) in model simulation.



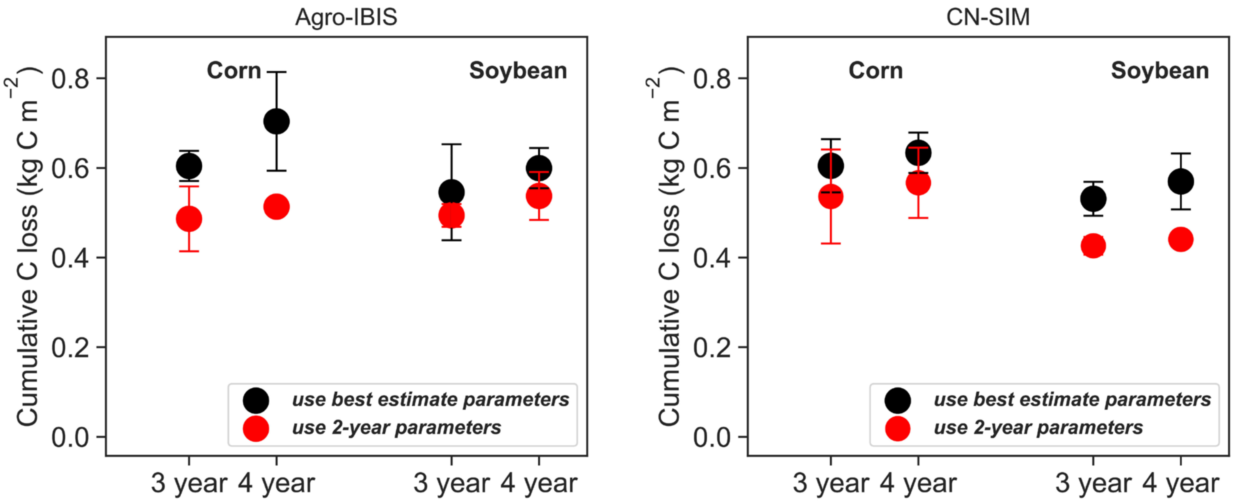
Supplementary Fig. 11 Comparison between isotope-calibrated CN-SIM model (a, b) and no-isotope-calibrated (c, d) CN-SIM model simulation results during corn production phase in the 2-year corn-soybean rotation system. The no-isotope-calibrated model overestimated total soil C loss but underestimated $\delta^{13}C$, indicating a lower estimated proportion of CO_2 loss from corn roots.



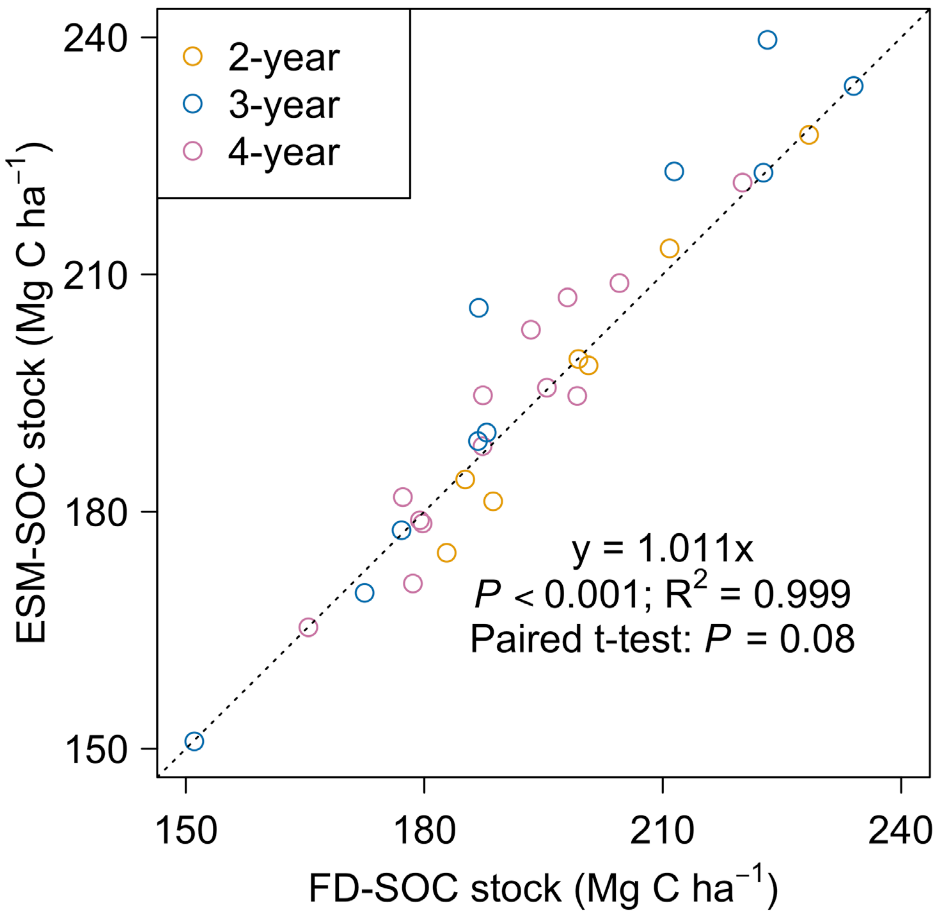
Supplementary Fig. 12(a) Comparison of calibrated model parameters for the decay rate of root carbon pool between the isotope-calibrated (red) and the no-isotope-calibrated models (blue) in the CN-SIM model. A convergent posterior parameter distribution indicates relatively low model uncertainty due to parameter equifinality. The X-axis represent the parameter prior range. The y-axis represents the probability density of the calibrated parameters with and without using $\delta^{13}C$ values of CO_2.



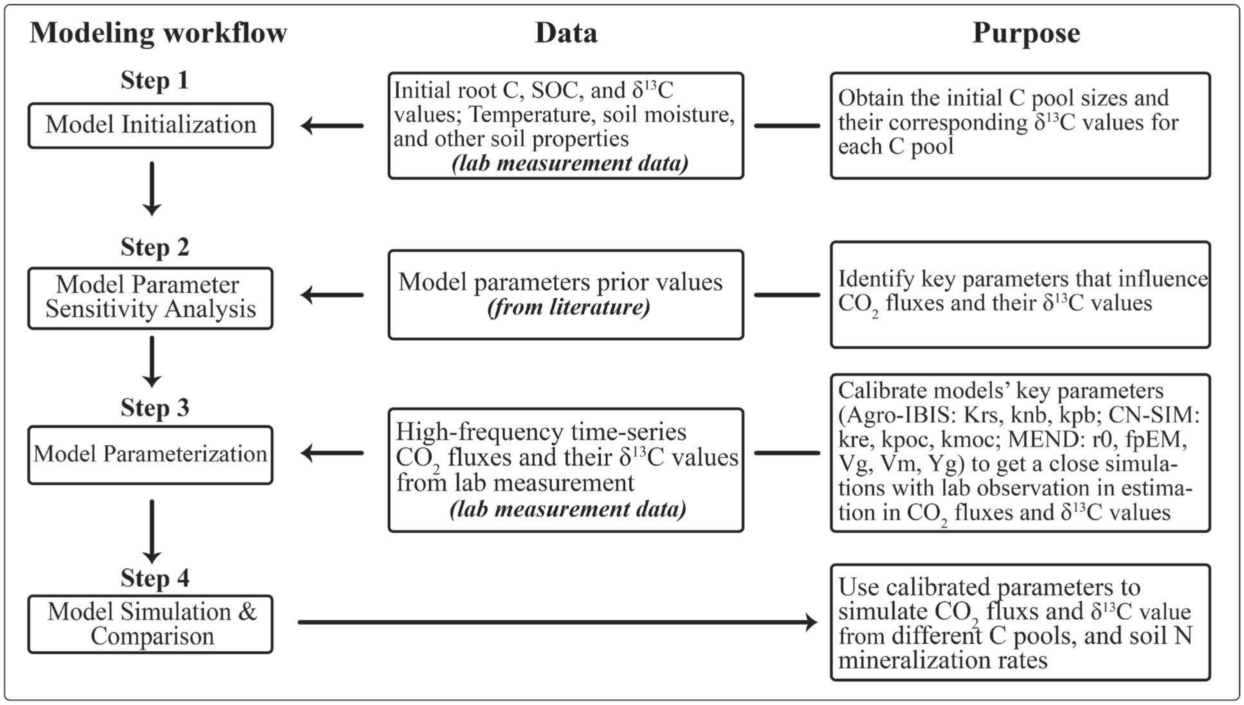
Supplementary Fig. 12(b) Comparison of calibrated model parameters for the decay rate of particulate soil C pool between the isotope-calibrated (red) and the no-isotope-calibrated models (blue) in the CN-SIM model. A convergent posterior parameter distribution indicates relatively low model uncertainty due to parameter equifinality. The y-axis represents the probability density of the calibrated parameters with and without using $\delta^{13}C$ values of CO_2.



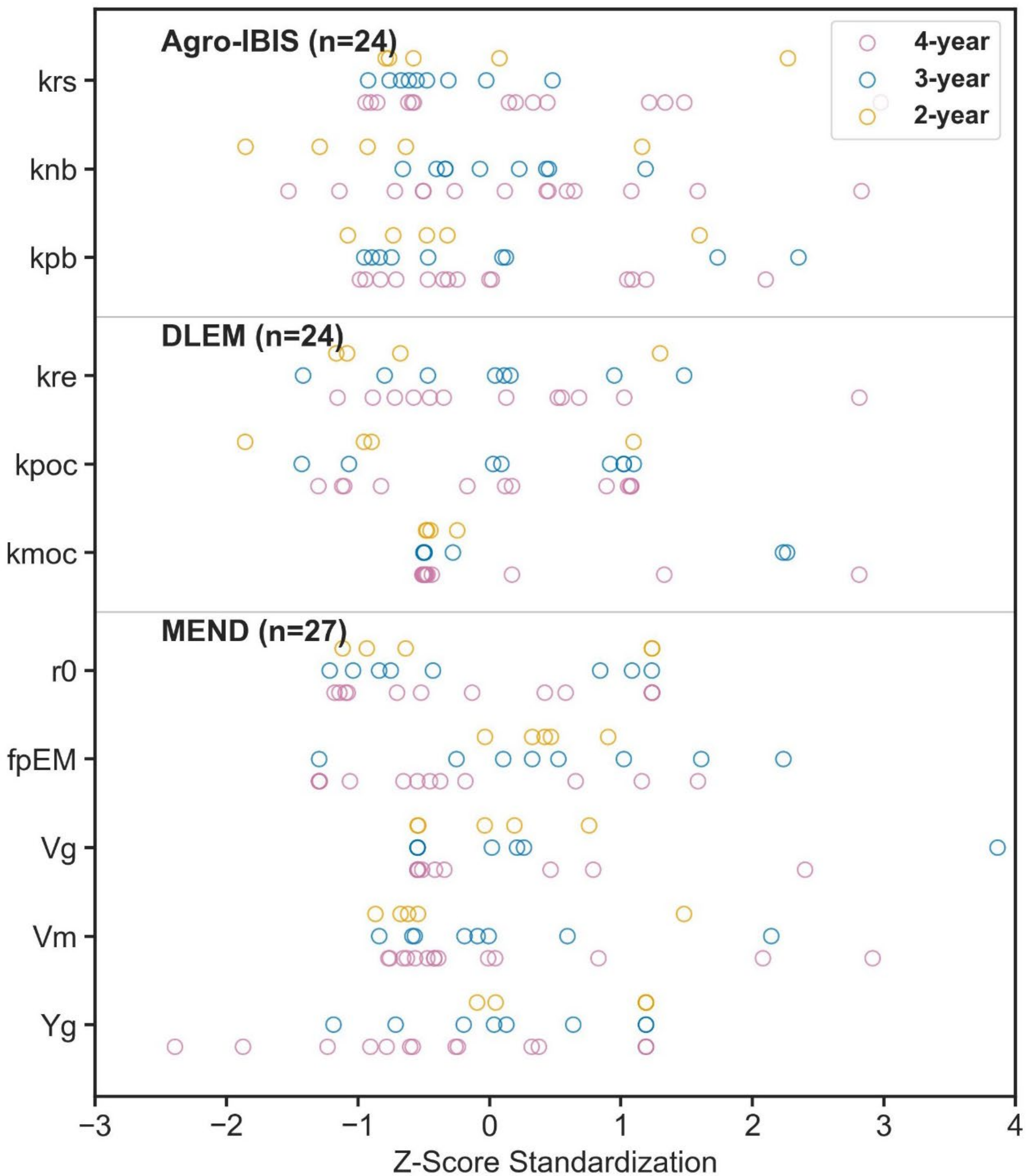
Supplementary Fig. 13 Cumulative C loss (CO_2 production) simulated by the Agro-IBIS (left) and CN-SIM (right) models during the corn and soybean production phases of the diversified cropping systems (3- and 4-year rotations) calculated by utilizing best estimate parameters (black) or by applying the calibrated parameters from the 2-year rotation (red). The error bars indicate standard errors (n = 12, and 16 for 3- and 4-year rotations, respectively), with the center denoting the mean value of the replicates.



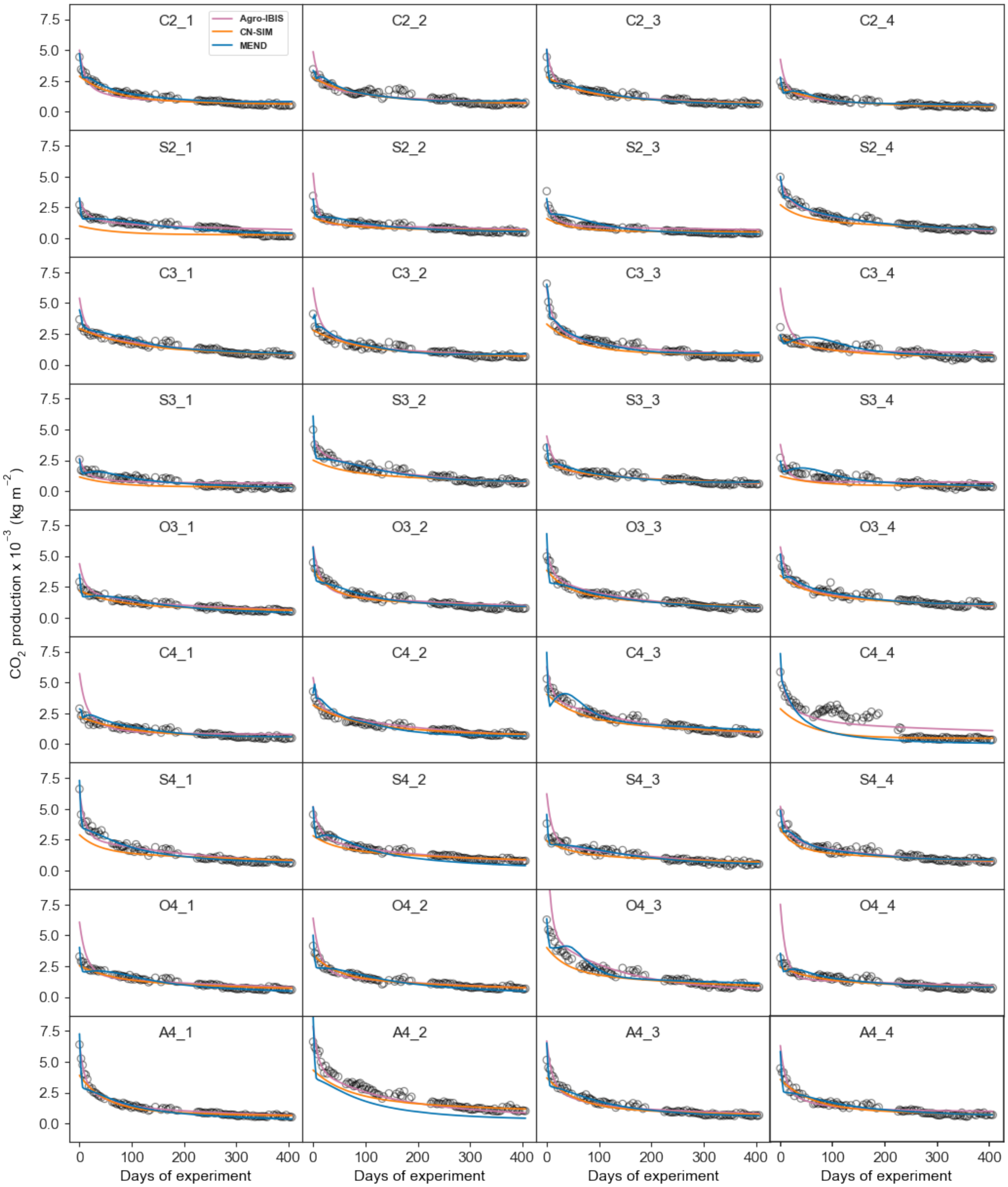
Supplementary Fig. 14 Soil organic carbon (SOC) stock to 1 m depth calculated by fixed depth (FD) and equivalent soil mass (ESM) methods. Each circle represents the SOC stock value to 1 m soil depth from each plot (n = 31 after excluding the outliers). The dashed line represents y = x, and the solid line is the fitted regression line with the intercept equal to 0 ($y = 1.011x$; $P < 2.2\times10^{-16}$; $R^2 = 0.999$). A two-sided paired t-test is used to compare the difference between ESM-SOC and FD-SOC stocks across all the rotations ($P = 0.08$).



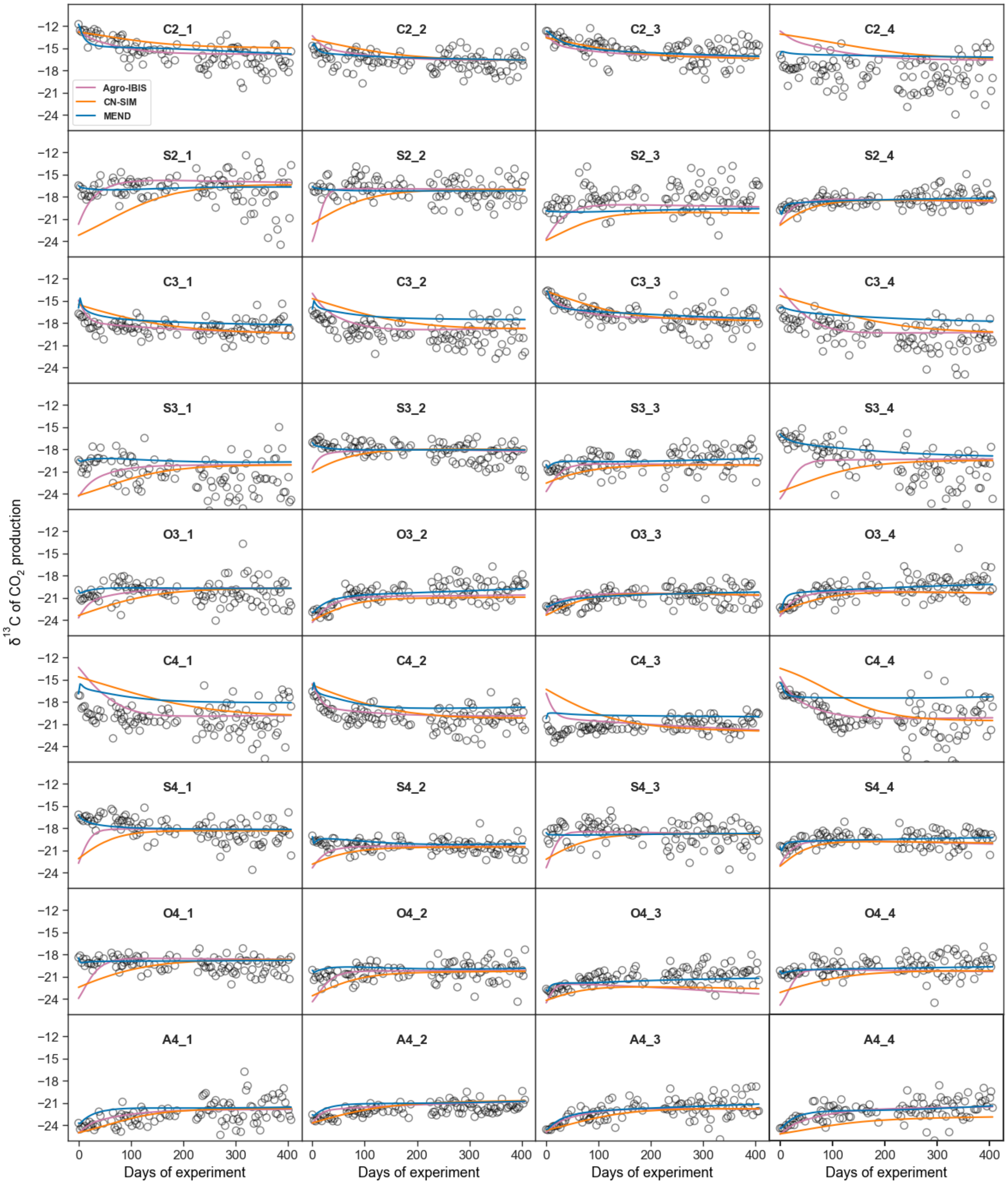
Supplementary Fig. 15 Modeling workflow in this study.



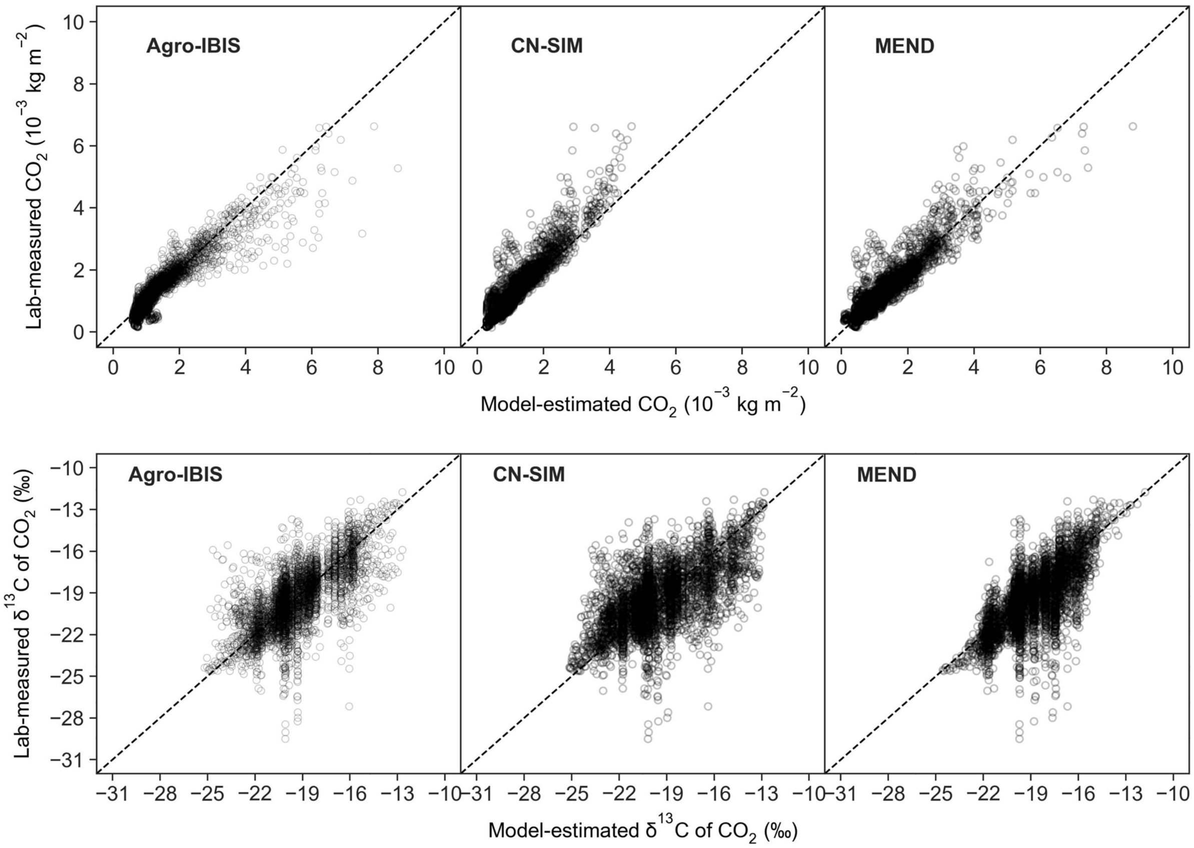
Supplementary Fig. 16 Optimal parameter values after data assimilation in the three soil C models under a 2-year corn-soybean rotation (2-year), a corn-soybean-small grain/red clover rotation (3-year), and a 4-year corn-soybean-small grain/alfalfa-alfalfa rotation (4-year). Values in the figure were calibrated based on the scenario simulation with the best-estimate values as initial C pools and both CO_2 fluxes and $\delta^{13}C$ of CO_2 as model output. The Z-score standardization values are calculated using the equation (Z-Score (Z) = (x - Mean) / Standard Deviation). The mean and standard deviation values for each parameter are 0 and 1, respectively.



Supplementary Fig. 17 Model simulation of CO_2 production from a 2-year corn-soybean rotation, a 3-year corn-soybean-small grain/red clover rotation, and a 4-year corn-soybean-small grain/alfalfa-alfalfa rotation by C isotope-calibrated models. The black points represent the observations. The colored lines represent three different model simulations defined in the legend. The treatment name is defined by using the first letter abbreviation of the currently planted crop in 2021 (C: corn; S: soybean; O: oats; A: alfalfa) and the rotation length in years. The name for each plot is defined by concatenating the treatment name and the replicate number with an underscore.



Supplementary Fig. 18 Model simulation of $\delta^{13}C$ values of CO_2 from a 2-year corn-soybean rotation, a 3-year corn-soybean-small grain/red clover rotation, and a 4-year corn-soybean-small grain/alfalfa-alfalfa rotation by C isotope-calibrated models. The black points represent the observations. The colored lines represent three different model simulations defined in the legend. The treatment name is defined by using the first letter abbreviation of the currently planted crop in 2021 (C: corn; S: soybean; O: oats; A: alfalfa) and the rotation length in years. The name for each plot is defined by concatenating the treatment name and the replicate number with an underscore.



Supplementary Fig. 19 Comparison in CO_2 production and its $\delta^{13}C$ values between lab observation and C isotope-calibrated model simulation from all incubated soil samples in the 2-year, 3-year and 4-year rotations. The dashed black lines represent the 1: 1 line. n = 3816 (36 soil cores ×108 measurements for each soil core).

Supplementary References

1. Huang, W. *et al.* High carbon losses from oxygen-limited soils challenge biogeochemical theory and model assumptions. *Global Change Biology* **27**, 6166–6180 (2021).
2. Amundson, R. G., Trask, J. & Pendall, E. A rapid method of soil carbonate analysis using gas chromatography. *Soil Science Society of America Journal* **52**, 880–883 (1988).
3. Huang, W. & Hall, S. J. Large impacts of small methane fluxes on carbon isotope values of soil respiration. *Soil Biology and Biochemistry* **124**, 126–133 (2018).
4. Hall, S. J., Huang, W. & Hammel, K. E. An optical method for carbon dioxide isotopes and mole fractions in small gas samples: Tracing microbial respiration from soil, litter, and lignin. *Rapid Communications in Mass Spectrometry* **31**, 1938–1946 (2017).
5. Ehleringer, J. R., Nina, B. & Lawrence B, F. Carbon isotope ratios in belowground carbon cycle processes. *Ecological Applications* **10**, 412–422 (2000).
6. Huang, W., Ye, C., Hockaday, W. C. & Hall, S. J. Trade-offs in soil carbon protection mechanisms under aerobic and anaerobic conditions. *Global Change Biology* **26**, 3726–3737 (2020).
7. Saltelli, A., Ratto, M., Tarantola, S. & Campolongo, F. Sensitivity analysis practices: Strategies for model-based inference. *Reliability Engineering & System Safety* **91**, 1109–1125 (2006).
8. Mousivand, A., Menenti, M., Gorte, B. & Verhoef, W. Global sensitivity analysis of the spectral radiance of a soil–vegetation system. *Remote Sensing of Environment* **145**, 131–144 (2014).
9. Sobol′, I. M. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation* **55**, 271–280 (2001).

10. Xu, T., White, L., Hui, D. & Luo, Y. Probabilistic inversion of a terrestrial ecosystem model: Analysis of uncertainty in parameter estimation and model prediction. *Global Biogeochem. Cycles* **20**, GB2007 (2006).
11. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).
12. Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109 (1970).
13. Ceglar, A., Črepinšek, Z., Kajfež-Bogataj, L. & Pogačar, T. The simulation of phenological development in dynamic crop model: The Bayesian comparison of different methods. *Agricultural and Forest Meteorology* **151**, 101–115 (2011).
14. Hararuk, O. & Luo, Y. Improvement of global litter turnover rate predictions using a Bayesian MCMC approach. *Ecosphere* **5**, 1–13 (2014).
15. Box, G. E. P. & Cox, D. R. A theoretical model of litter decay and microbial interaction. *Journal of the Royal Statistical Society: Series B (Methodological)* **26**, 211–243 (1964).
16. Vrugt, J. A. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software* **75**, 273–316 (2016).
17. Yi, B. *et al.* Resolving the influence of lignin on soil organic matter decomposition with mechanistic models and continental-scale data. *Global change biology* (2023) doi:10.1111/gcb.16875.
18. Wang, G. *et al.* Soil moisture drives microbial controls on carbon decomposition in two subtropical forests. *Soil Biology and Biochemistry* **130**, 185–194 (2019).

19. Wang, G. & Chen, S. A review on parameterization and uncertainty in modeling greenhouse gas emissions from soil. *Geoderma* **170**, 206–216 (2012).

20. Wang, G. *et al.* Microbial dormancy improves development and experimental validation of ecosystem model. *The ISME Journal* **9**, 226–237 (2015).

21. Duan, Q., Sorooshian, S. & Gupta, V. Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research* **28**, 1015–1031 (1992).

22. Poffenbarger, H. J. *et al.* Whole-profile soil organic matter content, composition, and stability under cropping systems that differ in belowground inputs. *Agriculture, Ecosystems & Environment* **291**, 106810 (2020).

23. Nichols, V. A. *et al.* Maize root distributions strongly associated with water tables in Iowa, USA. *Plant Soil* **444**, 225–238 (2019).

24. Ordóñez, R. A. *et al.* A solution for sampling position errors in maize and soybean root mass and length estimates. *European Journal of Agronomy* **96**, 156–162 (2018).

25. Lazicki, P. A., Liebman, M. & Wander, M. M. Root parameters show how management alters resource distribution and soil quality in conventional and low-input cropping systems in central Iowa. *PLoS ONE* **11**, e0164209 (2016).

26. Verberne, E. L. J., Hassink, J., Willigen, P. de, Groot, J. J. R. & Veen, J. A. van. Modelling organic matter dynamics in different soils. *NJAS* **38**, 221–238 (1990).

27. Yi, B. *et al.* Carbon decomposition and nitrogen mineralization in Marsden agroecosystem diversification experiment, Iowa, 2021. Environmental Data Initiative https://doi.org/10.6073/pasta/7f1dac6bf350221c54d1eb349ca89bbf (2024).

28. Córdova, S. C. *et al.* Soybean nitrogen fixation dynamics in Iowa, USA. *Field Crops Research* **236**, 165–176 (2019).

29. Blesh, J. & Drinkwater, L. E. The impact of nitrogen source and crop rotation on nitrogen mass balances in the Mississippi River Basin. *Ecological Applications* **23**, 1017–1035 (2013).

30. Ciampitti, I. A. & Salvagiotti, F. New insights into soybean biological nitrogen fixation. *Agronomy Journal* **110**, 1185–1196 (2018).

31. Tenorio, F. A. M. *et al.* Assessing variation in maize grain nitrogen concentration and its implications for estimating nitrogen balance in the US North Central region. *Field Crops Research* **240**, 185–193 (2019).

32. Rochester, I. J., Peoples, M. B., Constable, G. A. & Gault, R. R. Faba beans and other legumes add nitrogen to irrigated cotton cropping systems. *Aust. J. Exp. Agric.* **38**, 253–260 (1998).

33. Hergoualc'h, K. *et al.* "N_2O emissions from managed soils, and CO_2 emissions from lime and urea application" in *2019 Refinement to the 2006 IPCC Guidelines for National Greenhouse Gas Inventories*, E. Calvo Buendia et al., Eds. (Intergovernmental Panel on Climate Change, Geneva, 2019), pp. 11.1–11.48.